Universitätsklinikum Hamburg-Eppendorf

Faculty of Medicine
Institute for Structural- and Systemsbiology

Prof. Dr. Thomas Marlovits

Towards deciphering the Type III secretion signal

Dissertation
zur Erlangung des Grades eines Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von:
Maurice Pantel
aus Berlin

Hamburg 2024

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 15.04.2025

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Thomas C. Marlovits

Prüfungsausschuss, zweite/r Gutachter/in: Prof. Dr. Martin Aepfelbacher

Contents

Glossary **x**

I. Abstract **1**

1. Introduction **4**
1.1. Gram-negative pathogens . . . 5
1.2. Secretion systems . . . 6
1.3. The type III secretion system . . . 7
1.3.1. T3SS structure . . . 7
1.3.1.1. The needle complex . . . 9
1.3.1.2. The export apparatus . . . 9
1.3.1.3. The sorting platform . . . 10
1.3.2. T3SS assembly and substrate secretion . . . 10
1.4. Type III effector proteins . . . 12
1.4.1. Role of the chaperone-binding domain . . . 12
1.4.2. Secretion signals . . . 14
1.5. Salmonella . . . 17
1.6. Aim of this work . . . 17

2. Materials & Methods **19**
2.1. Materials . . . 20
2.1.1. Buffers & solutions . . . 23
2.1.2. Plasmids & primers . . . 24
2.1.2.1. Plasmids . . . 24
2.1.2.2. Primers & oligos . . . 27
2.1.3. Bacterial strains . . . 30
2.1.3.1. Strain selection . . . 30
2.2. Experimental procedures . . . 32
2.2.1. Cell culture . . . 32
2.2.2. Cloning . . . 34
2.2.3. Sample preparation . . . 36
2.2.4. Imaging . . . 36
2.2.5. Computational procedures . . . 37

3. Results **39**
3.1. T3SS effector quantification . . . 40
3.1.1. Qualitative detection of type 3 secretion system (T3SS)-secreted proteins . . . 40
3.1.1.1. T3SS reporter system design . . . 40
3.1.2. Quantitative detection of T3SS-secreted proteins . . . 46
3.1.2.1. Mass spectrometric quantification of T3SE secretion . . . 46
3.1.2.2. Luminescence-based quantification of T3SE secretion . . . 47
3.1.2.3. Species-independent T3SS secretion . . . 53
3.1.3. Optimization of assay conditions . . . 55
3.1.3.1. Culture volume & sample number . . . 55
3.1.3.2. Cultivation time . . . 57
3.1.3.3. Reporter expression control . . . 59
3.1.3.4. Absorbance calibration . . . 61
3.1.3.5. Device settings . . . 62
3.1.3.6. Time & temperature effects . . . 66
3.1.3.7. Optimized parameters . . . 67
3.1.4. Quantification of secretion signal effects . . . 68
3.1.4.1. Generation of plasmid variants . . . 68
3.1.4.2. Dual reporter secretion quantification . . . 70
3.1.4.3. Single reporter secretion quantification . . . 72
3.1.4.4. Reproducibility . . . 73
3.1.4.5. Calculation of secretion efficiency scores . . . 77
3.2. T3SS effector database . . . 79
3.2.1. Semi-automated T3SE dataset assembly . . . 79
3.2.2. T3SE database assembly . . . 80
3.2.2.1. Database classes . . . 82
3.2.3. Manually curated T3SE dataset assembly . . . 91
3.2.4. Cytoplasmic proteins dataset assembly . . . 91
3.2.5. LSTM prediction . . . 92

4. Discussion **95**
4.1. Assay design . . . 97
4.1.1. Reporter choice . . . 98
4.1.2. Reporter quantification . . . 98
4.1.3. Expression control . . . 100
4.1.4. Effect of the secretion signal on secretion quantity . . . 102
4.1.4.1. Secretion quantities . . . 102
4.1.4.2. Secretion efficiencies . . . 103
4.2. T3SE database . . . 105
4.2.1. Dataset preprocessing . . . 106

4.2.2. Prediction of secretion efficiencies 106

5. Conclusion **108**

5.1. Conclusion 109

A. Appendices **110**

A.1. Appendix A 111

A.1.1. Figures 111

A.1.1.1. Immunoblotting 111

A.1.2. Plots 120

A.1.2.1. Secretion quantification assays 120

A.1.3. Tables 121

A.2. Appendix B 123

A.2.1. Python code 123

A.2.1.1. Primer designer 123

A.2.1.2. Dataset random subselection 131

A.3. Appendix C 134

A.3.1. T3SE database 134

List of Figures

1.1. Structure of the Type III secretion system 8
1.2. Two hypothesis on the origin of the secretion signal have been proposed. . 15

3.1. Assembly of reporter plasmid pMP005. 41
3.2. SptP and SipA are secreted from plasmid-complemented strains. 42
3.3. Replacing the native N-terminus of SptP abolishes reporter expression. . . 44
3.4. The expression of SptP is regulated via RNA secondary structures in the N-terminus and contains an unusual start codon deviating from the Uniprot annotation. 45
3.5. The chimeric secretion signal-reporter fusion protein $SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$ is expressed and secreted via the T3SS. 46
3.6. Plasmid map of pMP028 with the native sptP sequence tagged C-terminally with the Nanoluc luciferase. 48
3.7. Secretion signals from different T3SS-secreted proteins display diverging secretion quantities. 49
3.8. Reporter variants are secreted in T3SS-mediated manner. 50
3.9. Residual Nanoluc luminescence observed in T3SS-deficient cultures can be attributed to cell lysis 52
3.10. Reporter variants with N-termini from other T3SS-carrying bacteria tested for T3SS-mediated secretion in the Nanoluc luminescence quantification assay (see Figure 3.11). 53
3.11. T3SS-mediated secretion of reporter fusion proteins with secretion signals from *Chlamydia*, *Escherichia*, *Shigella* and *Yersinia*. 54
3.12. Nanoluc luminescence signals converge due to sample mixing. 56
3.13. Arrangement of cell cultures in the 96-well deep well plate incubation setup. 57
3.14. Incubation times between 3.5 h to 5.5 h yield the highest Nanoluc signal-to-noise ratios. 58
3.15. The dual-reporter plasmid pMP049 expresses and secretes SptP-Nanoluc luciferase (NLuc) and SipA-red firefly luciferase (RFLuc). 60
3.16. The calibration of absorbance measurements enables the high-throughput measurement of cell culture optical densities using a platereader. 61
3.17. Application of a residual light filter reduces the average signal intensity. . 63
3.18. Enhanced dynamic range increases the resolution of luminescence signals. 64
3.19. Prior sample shaking reduces the average luminescence signal intensity. . 65

3.20. The Nanoluc luminescence intensity decays faster in high intensity samples. 66
3.21. The Nanoluc luciferase displays optimal activity at room temperature. . . 67
3.22. Optimized parameters of the secretion quantification assay in the current experimental setup 68
3.23. The levels of secreted control reporter SipA-RFLuc do not correlate with increasing levels of secreted SptP-NLuc reporter and display large variations in signal intensity. 71
3.24. Design of the single-reporter plasmid pMP059. 72
3.25. Nanoluc luminescence signals quantified from supernatants of the wildtype (WT) control strain SB905 Δ*sptP* + pMP059 display variation in quantity across 10 assays. 74
3.26. Nanoluc luminescence signals quantified from supernatants of the T3SS-deficient negative control strain SB905 Δ*invA* Δ*sptP* + pMP059 exhibit consistent signal intensities. 75
3.27. The relative secretion efficiency of secretion signal-reporter variants is altered by employing different induction strategies. 76
3.28. *Salmonella* effector N-termini promote different secretion efficiencies. . . 78
3.29. Exemplary fasta entry of the type 3 secretion system effector (T3SE)-database for *Salmonella* effector SipA. 82
3.30. Non-proteinaceous interaction partners targeted by T3SS-secreted proteins. 87
3.31. Experimentally validated localizations of T3SS-secreted proteins identified in our database. 88
3.32. Small, polar amino acids are enriched in the N-terminus of T3SS-secreted proteins. 90
3.33. *Shigella* T3SS effector of the IpaH class display strong homology within their N-terminal 10 residues. 91
3.34. The implementation of an embedding layer increases the models accuracy. 94

A.1. Expression and secretion of SptP from plasmid pMP005 as visualized by immunoblotting. 112
A.2. Expression and secretion of SptP from plasmid pMP008 as visualized by immunoblotting. 113
A.3. Expression and secretion of SipA from plasmid pMP0008 as visualized by immunoblotting. 114
A.4. Expression and secretion of $SipA_{1\text{-}25}$-$SptP_{26\text{-}543}$ and SptP from plasmids pMP009 and pMP010, respectively as visualized by immunoblotting. . . . 115
A.5. Expression and secretion of $SptP_{1\text{-}25}$-$SipA_{26\text{-}685}$ and SipA from plasmids pMP009 and pMP010, respectively as visualized by immunoblotting. . . . 116
A.7. Original western blots for expression of reporter variants in T3SS-deficient strain SB906 118
A.8. Original blot of Lysis control western blot. 119

A.9. The effect of shaking on absorbance measurements is negligible. 120

List of Tables

2.1. Devices used in this work. 20
2.2. Chemicals used in this work. 21
2.3. Consumables used in this work. 21
2.4. Enzymes & assay kits used in this work. 22
2.5. Antibodies used in this work. 22
2.6. Buffers, media & solutions used in this work. 23
2.8. Plasmids used in this work. 24
2.10. Primers used in this work. 27
2.12. Oligos used in this work. 30
2.13. *Salmonella typhimurium* strains used in this work. 31
2.14. *Escherichia coli* strains used in this work. 32
2.15. Touchdown-polymerase-chain reaction (PCR) cycle program. 35
2.16. Python libraries used for data analysis and machine learning pipeline development 37
2.17. Functions used for 2-class prediction of T3SE vs cytoplasmic proteins. . . 38

3.1. *Salmonella* proteins identified via liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) in *Salmonella typhimurium* SB905. . . . 47
3.2. N-termini of experimentally verified *Salmonella* T3SEs 68
3.3. Different sample types display diverging degrees of Nanoluc luminescence signal variation. 73
3.4. Current status of effector sequences in our T3SE database. 81
3.5. Experimental validation methods qualifying T3SS effector proteins for inclusion into the T3SE database, sorted by frequency. 84
3.6. Functional characterization of validated T3SS-secreted proteins. 85
3.7. Identified host species sorted by animal and plant hosts. 89

A.1. *Salmonella enterica* subsp. *enterica* Typhimurium LT2 codon table, frequency per thousand Kazusa database. * denotes the stop codons. 122
A.2. Search terms and annotations used for collecting and cleaning dataset 1. . 134
A.3. Functions associated with T3SE-secreted proteins. 135
A.5. Identified target proteins of T3SS-secreted proteins 149
A.7. Identified localizations of T3SS-secreted proteins 167

Glossary

ddH_2O	double-distilled H_2O
AMR	antimicrobial resistance
APS	ammoniumperoxodisulfate
ATP	adenosine triphosphate
BLA	β-Lactamase
bp	base pair
cAMP	cyclic Adenosinemonophospate
CBD	chaperone-binding domain
CDS	coding sequence
CV	coefficient of variation
CyaA	adenylate cyclase
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DOC	deoxycholate
DTT	dithiothreitol
EDR	enhanced dynamic range
EDTA	ethylendiaminetetraacetic acid
FL	Full length
FRET	Foerster resonance energy transfer

fT3SS . . . . . flagellar T3SS

GLuc . . . . . *Gaussia* luciferase

HR . . . . . hypersensitive response

HRP . . . . . horse radish peroxidase

LB . . . . . lysogeny broth

LC-MS/MS . . . . . liquid chromatography-coupled tandem mass spectrometry

LN . . . . . liquid nitrogen

LPS . . . . . lipopolysaccharide

LSTM . . . . . long-short-term memory network

MAD . . . . . median absolute deviation

MODEST . . . . . mage oligo design tool

mRNA . . . . . messenger ribonucleic acid (RNA)

NC . . . . . needle complex

nfT3SS . . . . . non-flagellar T3SS

NLuc . . . . . Nanoluc luciferase

NMR . . . . . nuclear magnetic resonance

OD . . . . . optical density

PAGE . . . . . poly-acrylamide gel-electrophoresis

PBS . . . . . phospate-buffered saline

PBS-T . . . . . phospate-buffered saline (PBS)-Tween 20

PCR . . . . . polymerase-chain reaction

PTB . . . . . protein transfer buffer

PVDF . . . . . polyvinylidene difluoride

rcf . . . . . relative centrifugal force

RFLuc . . . . . red firefly luciferase

RLuc . . . . . *Renilla* luciferase

RNA . . . . . ribonucleic acid

RNN . . . . . recurrent neural network

rpm . . . . . rotations per minute

RT . . . . . room temperature

SCV . . . . . *Salmonella*-containing vacuole/vesicle

SDS . . . . . sodium dodecyl sulfate

SPI-2 . . . . . *Salmonella* pathogenicity island-2

SPI-1 . . . . . *Salmonella* pathogenicity island-1

SS . . . . . Secretion signal

T1SS . . . . . type 1 secretion system

T2SS . . . . . type 2 secretion system

T3SE . . . . . type 3 secretion system effector

T3SS . . . . . type 3 secretion system

T4SS . . . . . type 4 secretion system

T5SS . . . . . type 5 secretion system

T6SS . . . . . type 6 secretion system

T7SS . . . . . type 7 secretion system

T9SS . . . . . type 9 secretion system

TAE . . . . . tris-acetate-EDTA

TCA . . . . . trichloroacetic acid

TD . . . . . touchdown

TEMED . . . . . tetramethylethylendiamin

WHO . . . . . World Health Organization

I. Abstract

Abstract

The non-flagellar T3SS (nfT3SS) is a large, multimeric complex that is widely distributed among pathogenic and commensal gram-negative bacteria alike. Spanning the inner and outer membrane of diderm bacteria, this syringe-like structure facilitates the unfolded transport of specific proteins to the extracellular space or directly into eukaryotic host cells. Upon entry into host cells, these type 3 secretion system effectors (T3SEs) modulate a diverse array of host cell pathways involved in immune response, cytoskeletal organization or host cell trafficking, promoting bacterial survival, invasion or infection. Although critically required for type 3 secretion system (T3SS)-specific transport of substrates, a secretion signal, located in the extreme N-terminus of the effectors, has yet eluded a clear characterization. Building up on previous studies that indicate a correlation between the composition of the secretion signal and the secretion efficiency of substrates, this work establishes a Nanoluc luciferase-based high-throughput secretion assay to quantitatively monitor the impact of the N-terminal secretion signal on the secretion efficiency of T3SS substrates. Towards obtaining comprehensive experimental data, differential secretion efficiencies for several *Salmonella* effector N-termini were confirmed, and the largest and best annotated database of T3SS-secreted proteins to date assembled.

Zusammenfassung

Das nicht-flagellare Typ 3 Sekretionssystem T3SS ist ein grosser, multimerer Proteinkomplex der unter pathogenen und kommensalen gram-negativen Bakterien gleichermassen weit verbreitet ist. Dieser einer Nadel ähnliche Proteinkomplex, der die innere und äußere Membran der didermen Bakterien umspannt, ermöglicht den Transport ungefalteter, spezifischer Proteine in den extrazellulären Raum oder direkt in eukaryotische Wirtszellen. Bei der Injektion der sogenannten Typ 3 Effektorproteine modulieren diese eine Vielzahl zellulärer Abläufe, die an der Immunreaktion, zytoskeletalen Organisationsprozessen, dem intrazellulären Transport sowie einer Reihe weitere Vorgänge beteiligt sind. Ein Sekretionssignal, das sich im extremen N-terminus dieser Effektoren befindet, konnte sich, obwohl essentiell fuer den T3SS-spezifischen Transport der Effektoren, bisher einer klaren Charakterisierung entziehen. Aufbauend auf früheren Studien, die eine Korrelation zwischen der Zusammensetzung des Sekretionssignals und der Sekretionseffizienz der Substrate andeuten, wird in dieser Arbeit ein Nanoluc-Luciferase basiertes Hochdurchsatz-Assay etabliert, welches es ermöglicht, den Einfluss verschiedener Sekretionssignale auf die Sekretionsquantität zu ermitteln. Zur Erfassung umfassender experimenteller Daten wurden unterschiedliche Sekretionseffizienzen für eine Reihe von *Salmonella* Effektoren bestätigt, ausserdem wurde die bisher größte und umfassendste Datenbank Typ 3 sekretierter Protein aufgebaut.

1. Introduction

1.1. Gram-negative pathogens

Communicable diseases transmitted by pathogenic bacteria remain one of the major global threats to human health in the 21st century [1]. In 2019, bacterial infections were associated with 7.7 million deaths making it the second largest cause of death worldwide [2]. The emerging spread of bacterial antimicrobial resistance (AMR) linked to an estimated 4.95 million deaths further aggravates the burden on health systems as effective treatment becomes increasingly challenging [3]. Driven by predictions of up to 10 million deaths by the year of 2050, the World Health Organization (WHO) has issued a priority list of bacterial species for which new antibiotics are urgently needed [4]. 9 out of 12 of the bacterial species are gram-negative, reflecting the high relevance of this group as pathogens [5].
In gram-negative bacteria, the distinct architecture of the diderm cell envelope with an inner membrane, a thin peptidoglycan layer and an asymmetric lipopolysaccharide (LPS)-phospholipid outer membrane provides advanced protection from antimicrobial drugs and antibiotics. Conceived as the main barrier of resistance against large, charged molecules the characteristic outer membrane forms a rigid layer intercalated by β-barrel porin proteins that allow passive uptake and efflux of nutrients and other molecules [6]. Active export of antibiotics via specialized drug efflux pumps contributes significantly to the high resistance of gram-negative bacteria to antibiotics and poses a significant threat towards the dissemination of resistant strains [7].

Mechanisms to deliver cargo across cell membranes are essential to the survival and growth of all living organisms and can be found in all domains of life [8]. In bacteria, a multitude of complex transport systems has evolved to traverse one or multiple membranes. The most abundant and ubiquitous export pathways are the Sec- and Tat pathways that direct transport from the cytoplasmic to the periplasmic or extracellular space. The importance of these pathways is reflected by the fact that in certain species up to 25 % of the proteome localizes to the periplasmic space and cell envelope, requiring appropriate means for transport [9]. In *E.coli* for example, the transport of unfolded proteins via the general secretory (Sec) pathway constitutes 98 % of all proteins to cross the inner membrane [10]. Some proteins are not viable for unfolded transport requiring specific ions, cofactors or a suitable folding environment [11]. These proteins are folded in the cytoplasm of the bacteria and directed to the Twin-Arginine translocation (Tat) pathway for export. To be recognized by the Tat translocase, the presence of a cleavable N-terminal signal peptide containing a specific twin-arginine motif at the N-terminus of the substrate is required [11]. This energy-consuming process contributes to a lesser degree to protein translocation but is still found in 77 % of all bacteria, in many archaeal species as well as in cyanobacteria and plants [12].

1.2. Secretion systems

In addition to the universal Sec and Tat-pathways prevalent in almost all bacterial species, many bacteria have evolved an arsenal of dedicated secretion systems to export substrates into the surrounding environment or into host cells. Out of 11 secretion systems identified to date, 10 have been identified in gram-negative bacteria [13][14][15][16]. A notable exception to this prevalence displays the type 7 secretion system (T7SS) that has been found in gram-positive bacteria of the family *Corynebacteria* and *Mycobacteria* [17][18]. While some of these systems seem to be restricted to specific bacterial phyla like the type 9 secretion system (T9SS) in *Bacteroidetes*, others like the type 2 secretion system (T2SS), T3SS or the type 6 secretion system (T6SS) are widespread among pathogenic and commensal gram-negative bacteria [14][19].

Functionally, secretion systems can be categorized according to multiple traits. The type 1 secretion system (T1SS) comprised of an inner membrane ATP-binding cassette (ABC) transporter, a periplasmic membrane-fusion and an outer-membrane porin, exports unfolded peptides across both diderm membranes in a one-step mechanism [8][20]. As opposed to the small size of substrates of classical ABC transporters, the T1SS has been shown to also export large substrates [20]. The T2SS and type 5 secretion system (T5SS) are unique in that they rely on the transport of cytoplasmic components via the Sec- or Tat pathways. While the T2SS facilitates the translocation of folded substrates delivered to the periplasm via the Tat pathway to the extracellular environment, type 5 substrates engage in a 2-step translocation procedure. They are translocated in unfolded manner via the Sec pathway before they fold and secrete themselves through the outer membrane forming a distinct β-barrel domain pore [8]. Similar to the type 1 secretion system, type 3 (T3SS), type 4 (T4SS) and T6SS moderate transport across both the inner and outer bacterial membrane. In contrast to the T1SS, these three systems are also capable of translocating cargo across a third membrane into other bacteria or eukaryotic host cells. Intriguingly, the highly diverse type 4 secretion system facilitates not only transport of proteins and protein-protein complexes, but also protein-DNA complexes both intercellularly and to the external environment [21]. Ancestrally related to DNA conjugation systems, this trait allows type 4 secretion system (T4SS)-harboring bacteria to exchange genetic information with other bacteria or induce bacterial killing providing a competitive advantage in microbial communities [22][23][24]. Another secretion system that has been associated with interbacterial communication is the highly conserved T6SS for which multiple roles such as involvement in bactericidal activity, growth competition in biofilms or self-versus-nonself discrimination has been reported [25][26][27]. Importantly, both the T4SS and the T6SS also play a role in virulence to eukaryotic hosts underlining their high versatility and the frequently occurring exaptation to specific tasks [19][28].

1.3. The type III secretion system

One of the best characterized secretion systems to date is the type 3 secretion system. T3SS can be divided into two distinct, ancestrally related systems; the flagellar T3SS (fT3SS) that drives locomotion and the non-flagellar T3SS (nfT3SS) or virulence-associated T3SS that presumably evolved as an exaptation of the related flagellar system [29][30]. The nfT3SS that is also referred to as the injectisome, is a large, membrane-embedded multimeric protein complex. It comprises more than 20 proteins and is widely distributed among gram-negative bacteria forming both pathogenic and symbiotic relationships with a broad range of eukaryotic hosts [31][32]. Bypassing the periplasm, it serves as a direct conduit to translocate proteins from the bacterial cytoplasm into the extracellular space or into eukaryotic host cells crossing two to three membranes, respectively [33].
In a global survey of 20000 bacterial genomes, Hu et al. identified nfT3SSs from 36 bacterial species that could be divided into 12 categories based on microsynteny organization and phylogenetic analysis [34]. The identification of 174 different T3SS from 109 genera emphasizes the wide spread of this secretion system among gram-negative bacteria. Often these bacteria not only harbor a single but multiple T3SSs, highlighting its importance to the bacterial survival [35].

1.3.1. T3SS structure

The structure of nfT3SS can be segmented into four major substructures: the basal body, the extracellular needle, the export apparatus and the cytoplasmic sorting platform (see Figure 1.1) [36]. Together, the basal body and the needle comprise the needle complex (NC), a 3.6 MDa syringe-like subcomplex, that coined the term injectisome [37].

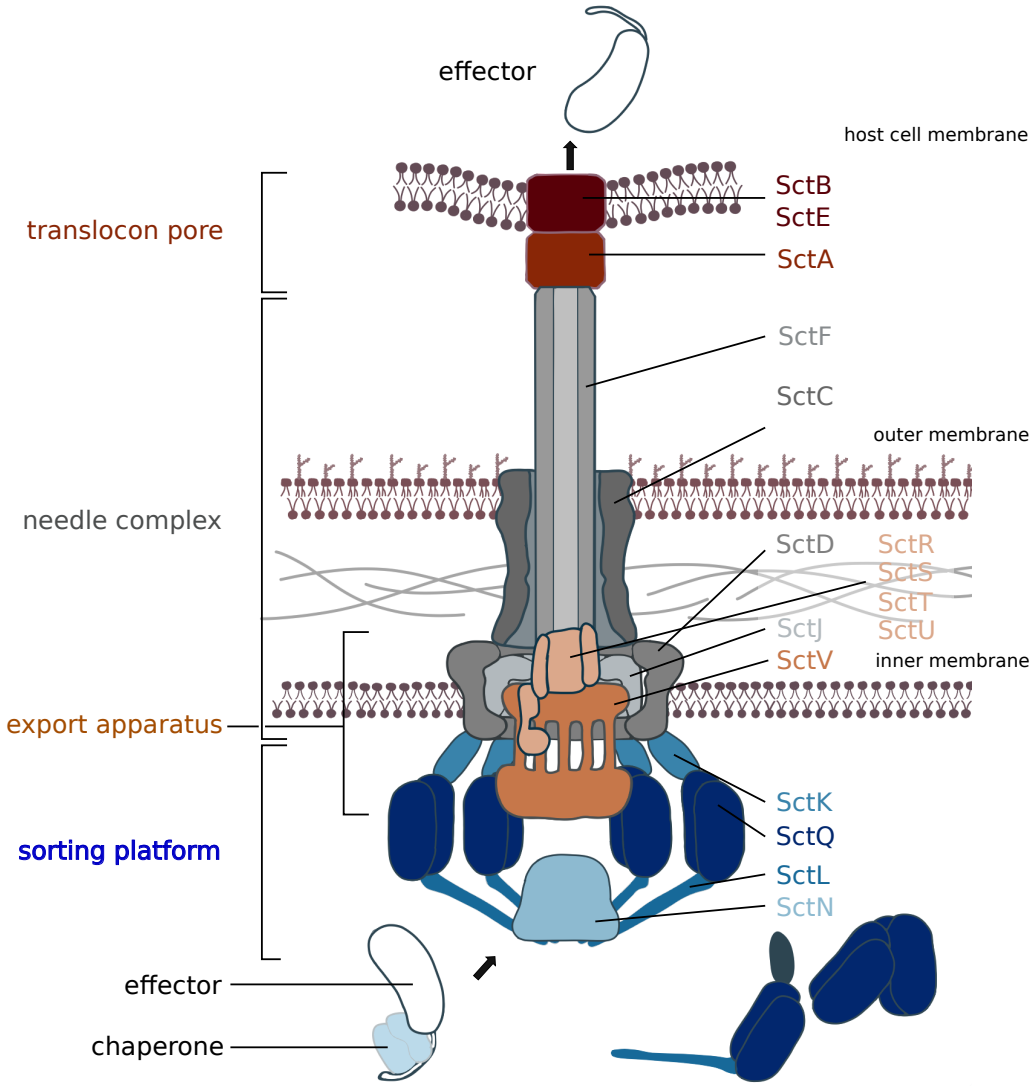


Figure 1.1.: Structure of the Type III secretion system adapted from Wimmi et al. [38]. The nfT3SS can be segmented into the subcomplexes, needle complex, export apparatus and sorting platform. At the distal end of the needle complex, a translocon pore allows secretion of substrates directly into host cells. The cytoplasmic sorting platform is thought to dynamically exchange effector-bound subcomplexes of SctK, SctQ and SctL as a shuttle mechanism during the secretion process[38].

1.3.1.1. The needle complex

The needle complex is composed of the basal body, formed by several membrane-embedded rings and the needle filament, protruding the bacterial cell to form contact with host cells. The basal body mainly acts as a scaffold anchoring the complex between the inner and outer bacterial membrane and providing a rigid basis for translocation [39][40]. Two inner rings with 24-fold symmetry composed of the proteins SctD and SctJ are connected to an outer ring comprised of SctC proteins [41][42]. The inner rings connects the complex to the inner bacterial membrane, while the outer ring embeds the complex in the outer membrane. Anchored to the basal body via the inner rod protein SctI, the needle is composed of the needle filament SctF, enclosing a narrow channel of roughly 2.5 nm that serves as a conduit for substrates necessitating transport in unfolded manner [41]. Arranged in helical fashion, the needle extends between 30 nm to 70 nm in length in a species-specific manner [43]. The α-helical arrangements results in an inner lumen composition of mostly polar residues with alternating positively and negatively charged regions that is assumed to promote secretion [43]. Interestingly, recent findings indicate potential translocation of partially unfolded substrates retaining α-helical secondary substructures during transport [39]. The length of the needle is regulated by SctP and the inner rod protein SctI that has also been connected to play a role in the selectivity of substrates [44]. At the distal end of the needle, the tip protein SctA provides the basis for translocon proteins SctB and SctE to insert into the host cell membrane forming a translocon pore [45]. In some bacterial species, the tip protein is replaced by a pilus or filament protein as shown e.g. for enteropathogenic *Escherichia* strains [46].

1.3.1.2. The export apparatus

The export apparatus is embedded within the needle complex and constitutes a decameric complex of the proteins SctR, SctS and SctT [39]. Four SctS proteins form the entry for substrates to traverse the complex, forming a hydrophilic interface of glutamine residues interacting with the substrate backbone (Q1-belt) [39]. Following this sidechain-independent translocation of the substrate, 5 SctR and a single SctT molecules constitute the hydrophobic, staircase-like M-gate. Opening of this gate allows passage of substrates while maintaining celullar homeostasis [39]. Located on top sits a second hydrophilic Q2-belt with the conformationally flexible SctT acting as a lid and two SctR glutamines promoting further transport through the channel.
Wrapping around the cytoplasmic end of the export apparatus, a single protein of SctU forms connections with SctR and SctT [47]. Hypothesized to induce conformational changes leading to an opening of the apparatus, SctU is often designated as a substrate switch [47]. In accordance with the assumed role for SctU, its directional position within the export apparatus supports speculations of an interaction with the cytoplasmically located nonameric ring of SctV that has been shown to bind to effector-chaperone pairs [48].

1.3.1.3. The sorting platform

The sorting platform is located at the cytoplasmic interface of the injectisome. Anchored by the cytoplasmic domain of the inner ring protein SctD, five proteins (SctK, SctL, SctN, SctO, SctQ) comprising this subcomplex form a pod-like structure with 6-fold symmetry [33][37][41]. SctK acts as a linker between SctD and SctQ and docks this cytosolic complex beneath the inner ring and the export apparatus [35]. Full-length SctQ shares homology to the flagellar C-ring protein FliM, an internal translation site produces a truncated SctQ homologous to FliN [49],[50]. In *Yersinia*, 22 copies of the full SctQ protein are found in non-secreting and secreting conditions [49]. Instead of forming a closed ring structure similar to the flagella system, the C-ring proteins of the sorting platform seem to adopt pod-like folds that dynamically cycle between cytosolic subcomplexes and injectisome-docked state, putatively acting as shuttles for effector-chaperone complexes [51][38]. Transient complexes of SctQ, SctK and SctL have been shown to comigrate with T3SS substrates that display differential affinities towards binding to the cytosolic complexes. The direct binding of effector proteins to SctQ that coincides with evidence for chaperone-independent secretion of effectors in *Shigella* as well as binding of effector-chaperone complexes to SctQ indicates several possible modes of action depending on the bacterial species [38][52][53]. SctL and SctK appear not to be involved in effector binding but seem to contribute to the stability of the cytosolic subcomplexes [38].

Located in the center of the sorting platform sits the homohexameric ATPase SctN. A recent cryo-EM structure of the *E.coli* ATPase EscN hexamer in complex with the central stalk protein EscO/SctO reported by Majewski et al. displays similarity to the rotational mechanism of F_1/V_1-ATPases [54]. The stalk protein SctO is thought to serve as a anchoring point between SctN and SctV relaying conformational changes within the T3SS as a result of SctN activity [55][56]. Linked via SctL dimers, SctN forms a central complex that serves as a docking point for effector-chaperone bound SctK-SctL-SctQ pods and assists in chaperone detachement, a process that has drawn comparisons to the functional role of AAA+-ATPases in protein complex dissociation or unfolding [57] [58]. The recognition of effector-chaperone complexes seems to occur independent of the presence of ATP, the dissociation of chaperones however requires the catalytic activity of SctN [57]. Reports of the ATPase-independent secretion of substrates question whether SctN acts as the major driving force in the translocation process or in an auxiliary role in the initial unfolding of effectors [59]. The export of flagellar proteins in the related fT3SS in absence of its ATPase FliL suggest the latter and underline the necessity of the proton motive force (PMV) as driver for translocation [60][61][62][63].

1.3.2. T3SS assembly and substrate secretion

To add more complexity to the open question of how the initial recognition and secretion of substrates is orchestrated, the assembly of T3SS systems appears to be a highly hierarchical

process that can be divided into four discrete steps [31].

For the initial stages of assembly two models have been proposed, the inside-out model and the outside-in model. In the inside-out model, the integration of the export apparatus to the inner membrane and the subsequent formation of the inner rings (SctI, SctD) is followed by the localization of the secretin SctC to the outer membrane. In the outside-in model, the onset of activity by peptidoglycan-cleaving enzymes is defined earlier allowing initial localization of the outer membrane ring in the outer membrane followed by integration of the export apparatus and the inner rings. Intriguingly, while the ancestral relationship of the fT3SS and the nfT3SS promotes the inside-out model, fluorescence-labeling experiments on the structural components of the T3SS suggest an alternative evolutionary path for the *Yersinia* T3SS that follows the outside-in model [64].

After assembly of the basal body and the export apparatus, subsequent steps necessitate the secretion of structural components in T3SS-dependent manner [35]. The complex architecture of the T3SS machinery requires the precise sequential secretion of substrates in order to guarantee the correct assembly of the complex. The first substrates to be T3SS-secreted are the structural proteins SctI and SctF that arrange into the inner rod and needle. Completion of the inner rod has shown to induce conformational changes that result in the first substrate specificity switch and prevent further secretion of early substrates [65]. In *Salmonella*, the regulatory protein SctP has been associated with a role as a needle length regulator and is critically required for assembly of the inner rod, serving in a stabilizing function promoting anchoring of the inner rod and needle in the basal body [44]. Curiously, the length of SctP correlates well with the length of the resulting needle leading to a divergent explanation on the mechanism of needle length regulation. Supported by experiments in *Yersinia*, this model assumes SctP to function as a molecular ruler that signals the substrate switch upon full extension [66]. Finally, the species-specific length of the needle is influenced by the stoichiometry of SctF and SctI, overexpression of either protein results in aberrantly long or short, yet functional needles [65]. The absence of SctP however, completely impaired the assembly of a functional T3SS emphasizing its importance in the assembly process. Thus, how SctP regulates needle length and substrate switch is still under debate and findings in *E.coli*, reporting an interaction of SctP with the gatekeeper protein SctW in a calcium-dependent manner indicate another additional role in external sensing [67].

Ensuing the assembly of the needle, the middle substrates comprised of the tip protein SctA and the translocon proteins finalize the assembly of the T3SS and establish host cell contact. In the process of switching from early to middle substrates, the switch protein SctU seems to act as binding partner for different structural components in the export apparatus as *in-vitro* studies suggest [68]. SctU belongs to a family of inner membrane proteins that contain a conserved motif for autocatalytic cleavage. Early speculations of this autocleavage as the trigger for substrate switching appear unlikely as the cleavage event occurs prior to incorporation of SctU into the T3SS base [69]. At this point a second

substrate specificity switch occurs resulting in the secretion of effector proteins (T3SE). This switch seems to be mediated by SctW, often termed gatekeeper protein that has been identified in multiple species such as *Pseudomonas*, *Salmonella* or *Shigella* [70][71][72]. In its role as gatekeeper SctW binds to translocon-chaperone complexes and recruits them to the sorting platform by docking them to SctN. Additionally, SctW is able to suppress secretion of later substrates. The exact mechanism how SctW is able to sense the assembly of the translocon pore and activate substrate switching to effector proteins is still debated [35][73]. After the establishment of host cell contact, secretion of T3SE proteins occurs that mainly exert their function in the host cell. The intricate manipulation of the host cells and the interplay of many of these effectors adopting redundant functions during infection or invasion furthermore stresses the importance of a highly regulated secretion process.

1.4. Type III effector proteins

Despite significant advances in the structural characterization of the nfT3SS several aspects of the substrate recognition and translocation mechanism remain elusive. For the study of these processes it has become imperative to not only focus on the structural components comprising the T3SS but to also thoroughly investigate features of the T3SS-secreted substrates that might be important during translocation. Concomitant to the elucidation of functional roles and activities of effector proteins in the context of host cell interaction and virulence, distinct patterns shared by the majority of effector proteins have emerged. Typically, T3SEs are multi-domain, sometimes multifunctional proteins that display a wide array of activities adapted to their specific role of interaction with the targeted host cells. Despite the large range of functions, T3SEs seem to share a modular architecture that has prompted speculations about their origin and evolution [74].

Commonly, type III effectors harbor one or multiple functional domains located in the central or C-terminal fraction of the protein that often mediate different, seemingly unrelated activities [74]. Exemplarily, the *Salmonella* effector SptP containing both a GTPase-activating domain (GAP) with homology to the *Pseudomonas* effector ExoS and a tyrosine-phosphatase domain related to the *Yersinia* effector YopH, these functional domains suggest a shared origin and subsequent adaptation to the respective host [74][75][76]. In respect to the fact that T3SS-harboring pathogens infect and manipulate eukaryotic hosts, it is not surprising that the evolution of the active domains often mimics structural and functional roles of eukaryotic proteins [77][78].

1.4.1. Role of the chaperone-binding domain

Separate from the functional diversity of the C-terminal domains, T3SE harbor one to two distinct domains located in the N-terminal region that play a role in facilitating secretion and determining the hierarchical secretion of different substrates. For many of the identified

effector proteins a chaperone-binding domain (CBD), located between residues 50 to 150, displays a prerequisite for T3SS-mediated secretion. The identification of single- and multi-cargo chaperones for a vast number of effectors and the interaction of structural components with chaperone-effector pairs implies a certain importance to the secretion process. In the fT3SS, the interaction of a chaperone-substrate complex to FlhA, a homolog of the export apparatus protein SctV highlights a potential mode of action with the substrate-chaperone interaction adopting a conformation suitable for bringing the substrate in close proximity to the opening of the export apparatus channel [79]. As recently demonstrated by Wimmi et al., chaperone-effector complexes also bind to the sorting platform protein SctQ, as well as to the ATPase SctN suggesting a common theme in priming effectors for translocation [38][80]. Supporting this hypothesis, the mechanical lability of some effectors as shown for *Salmonella* SopE2 and SptP stresses the important role of chaperones in keeping effector proteins in a stable, secretion-competent state and assisting in the translocation process [81]. The dynamics of this mechanism remain to be determined, differential expression patterns of chaperones and their cognate effectors support a recycling mechanism of chaperones after dissocciation from their effectors [62].

With a view to the intricate hierarchical secretion order required for T3SS assembly and manipulation of hosts, it has further been speculated whether different binding affinities of chaperone-effector complexes to the cytoplasmic sorting platform might contribute to determining the secretion order [82]. Studies monitoring the secretion kinetics of several effectors have reported both similar but also strongly diverging secretion rates for different T3SEs [83][84]. The fact that *Salmonella* effectors SopE and SipA displaying similar secretion kinetics bind to the same multi-cargo chaperone InvB could be indicative of the assisting role chaperones play in regulating the secretion process [83]. Alternatively, the real-time observation of functionally opposing *Salmonella* effectors SopE2 and SptP respectively activating or suppressing the host GTPase Cdc42, revealed that SopE2 is secreted 2-fold faster than SptP [84]. Remarkably, both effectors contain signals within their N-terminal domain that result in differential proteasomal degradation rates in the host cell, adding another layer of regulation [84][85].

Despite these findings, Ernst et al. have reported the chaperone-independent secretion for multiple *Shigella* effectors [52]. This is in congruence with the chaperone-independent secretion of the *Yersinia* effector YopO lacking its chaperone-binding domain (CBD) [86]. Finally, Lee and Galán proposed a role in conferring secretion-pathway specificity for the *Salmonella* chaperones SicP and InvB targeting the effectors SptP and SopE to the nfT3SS rather than the fT3SS [87]. The finding that these effectors are still secreted via a Type III secretion system further sparked interest into another essential domain identified in the extreme N-terminus of the effector proteins that functionally seems to resemble a secretion signal as observed for other secretion systems.

1.4.2. Secretion signals

Secretion signals have been identified for a number of secretion systems. Typically located in the amino acid sequence, these signal peptides guide the substrates to the designated export pathway via a distinct composition or pattern, either composed of conserved physico-chemical properties, structural elements or by a clear consensus sequence. For example, proteins targeted to the general Sec pathway contain a cleavable N-terminal signal sequence comprised of 3 distinct regions, a positively charged N-terminal region, a hydrophobic core and a polar carboxyterminal region, promoting association with components of the SecYEG machinery [88]. Substrates of the type I secretion system however, are characterized by a C-terminal secretion sequence and nonapeptide repeats N-terminal to the secretion sequence [89]. T1SS secretion signals display high variability in sequence, yet adopt a flexible structural fold consisting of 2 α-helices and an unstructured C-terminal domain [90].

For type III secretion system substrates, several studies have proclaimed the existence of a secretion signal in the extreme N-terminus of the T3SE. Although critically required for T3SS-mediated secretion, the lack of a distinct conserved secretion signal has fueled research towards its identification.

In the late 1990s, two opposing hypotheses were postulated proposing a T3SS secretion signal to either originate in the peptide sequence or the mRNA 5'-end of an effector protein, respectively [91][92][93][94]. The identification of two domains within the N-terminal 75 residues of the *Yersinia* effector YopE by Schesser et al. and Sory et al. that are required for T3SS-mediated secretion and translocation strengthened the idea of a signal rooted in the peptide sequence [91],[92]. Composed of a secretion signal at the extreme N-terminus and a translocation signal within the N-terminal 49 residues of YopE these domains facilitate transport across the bacterial or host cell membranes [91]. Mutational studies on the *Yersinia* effectors YopE, YopN and YopQ implying a tolerance to frameshift mutations in the N-terminal nucleotides led to the alternative theory of an mRNA-encoded secretion signal [93][94]. The limited effect of those frameshifts on the peptide sequence (Figure 1.2A) and ambiguous results on the *Yersinia* effector YopQ presented in subsequent works by Schneewind and colleagues did not lead to a clarification on the origin of the presumed secretion signal [95] [96]. Findings, like the abrogation of secretion via introduction of a silent mutation in the YopQ N-terminus have remained scarce and although an impact of the mRNA sequence on T3SS-mediated secretion should not be ruled out [96], subsequent works provided more evidence in support of a secretion signal localized in the peptide sequence. As Lloyd et al. showed in 2001, the alteration of more than half of the nucleotides in the N-terminal 10 codons of YopE did not impair its secretion as long as the amino acid sequence remained intact [97]. The expansion of research to other effectors and T3SS-harboring species such as *Salmonella* brought more evidence of a peptide-encoded signal and advanced its characterization [98][99]. It should be noted, that while more evidence for a peptide located secretion signal exists, a study from 2013 also identified the

potential involvement of 5'-Untranslated leader sequences and the RNA-binding protein Hfq in effector translocation [100].

The ability to facilitate T3SS-mediated secretion in heterologous bacteria as shown for

A

YopN WT	MTTLHNLSYGNTPLR
YopN -1	MRRFITYLMAIPRCV
YopN +2	MKRRFITYLMAIPRC
YopN +1	MNDASKPILWQYPAA
YopN -2	MDASKPILWAYPAAG
YopE WT	MKISSFISTSLPLPA
YopE -1	MKYHHLFLHHCPCRQ
YopE +2	MKKYHHLFLHHCPCR
YopE +1	MENIIIYFYITAPAG
YopE -2	MNIIIYFYITAPAGG

B

pattern	enriched/depleted
polar-hydrophobic-polar	enriched
alkaline-alkaline	depleted
threonine	enriched
serine	enriched
proline	enriched
polar	enriched
alkaline	depleted
acidic	depleted
hydrophobic-alkaline	depleted
polar-polar	enriched

Figure 1.2.: Two hypothesis on the origin of the secretion signal have been proposed. A N-terminal peptide composition of the *Yersinia* effectors YopE and YopN after introduction of frameshift mutations. Adapted from [93]. **B** The computational analysis of T3SEs N-termini exhibits discriminatory features within the peptide sequence against non-T3SE sequences with an enrichment bias towards small polar residues such as threonine and serine. Adapted from [101].

chlamydial and *Vibrio* effector N-termini in *Yersinia* or the identification of homologous effectors in *Salmonella* led to a systematic review into secretion signal properties [102][103] [104][105]. With the growing availability of sequenced bacterial genomes and the advent of sophisticated computational methods such as machine learning it became apparent that while a consensus pattern as identified for signal peptides of other secretion systems seems lacking, T3SS secretion signals share common features across bacterial species.

An analysis of the N-terminal domains of animal and plant pathogens revealed a significant enrichment for small polar residues such as serine and threonine and a depletion of leucine and acidic residues [106] [101][107]. Based on the propensity for an overall amphipathic sequence composition in combination with the enrichment of prolines it was speculated that the secretion signal might be intrinsically disordered as a key property [101][107]. In light of the fact that some ATPases specifically recognize unstructured peptides on their substrates, one could speculate about a role of the secretion signal in the binding to SctN [108]. Despite the lack of a concrete definition for the secretion signals several computational approaches have successfully predicted new effector proteins using solely the information of the N-terminal peptide sequence [101][109][110][111]. The concomitant biochemical characterization of more T3SS substrates revealed that effector proteins are secreted in hierarchical and competitive order [84][112]. Importantly, Sorg et al. demonstrated that during assembly of the export apparatus, autocleavage of the substrate switch protein SctU is specifically required to allow secretion of the tip and translocon

proteins SctA, SctB and SctE to form the needle tip and translocon pore without hindering secretion of late substrates [113]. Remarkably, the exchange of the N-terminus of SctA with the N-terminal sequence of the late substrate YopE restored its secretion in a strain incapable of SctU autocleavage. As opposed to the location of T3SS secretion signals in the extreme N-terminus, Login and Wolf-Watz identified a C-terminal secretion signal in the cleaved C-terminal peptide of SctU that is critical for its secretion. The lack of this signal resulted in increased secretion of the middle substrate SctF without compromising secretion of late substrates [114]. Considering that in both cases the secretion of late substrates were not affected, it seems natural to assume the existence of separate, specific secretion signals recognized by the T3SS machinery securing the secretion order of early, middle and late substrates. A work conducted in *E.coli* also supports this notion, postulating the existance of an additional translocon protein-specific secretion signal located downstream of the N-terminal secretion signal between residues 20 to 70 [115].

Diverging secretion rates as briefly described before (subsection 1.4.1), not only prompted the investigation of T3SE chaperones but also raised speculations whether the N-terminal secretion signal itself might contribute to the secretion efficiency. In general, effector proteins are secreted and translocated in highly different quantities depending on their activity in the host cell. Some effectors critically involved in promoting infection such as the *E.coli* effector Tir or the *Salmonella* effector SipA have been shown to be translocated in high abundance and very early during the infection process [116]. Other effectors however, are translocated at much lower rates, possibly due to the translocation of multiple effectors with redundant function inside the host. Aside from regulatory effects on the gene level, the translocation efficiency of many effectors is directly influenced both by the intrabacterial concentration as well as the availability of chaperones to stabilize them and guide them to the T3SS [117][118]. Interestingly, as shown for several *Shigella* effectors not all T3SEs require chaperones for secretion, raising the question how the secretion levels for these substrates are modulated [52]. Another study, that conducted shuffling experiments on the N-terminal residues of the *Yersinia* effector YopE, discovered that amphipathic sequences with alternating serine and isoleucine residues facilitate effective secretion, while longer streches of polar or very hydrophobic signal sequences diminished the quantity of effector secretion [106]. It therefore seems likely to assume that the sequence composition of the secretion signal itself also contributes to the finetuned deployment of effectors unrelated to expression levels and chaperone-mediated allocation to the sorting platform. The investigation of a correlation between secretion signal and effector secretion quantity/efficiency has not been tested and could narrow the search for why the secretion signal is essential for the translocation process. It will be part of this work.

1.5. Salmonella

Among the most intensively studied organisms for deciphering the role of the T3SS and its effectors in the context of virulence is the gram-negative bacterium *Salmonella* of the *Enterobacteriaceae* family. As the major cause for foodborne infections, *Salmonella* serovars are responsible for an estimated number of 200 million to 1.3 billion cases of disease worldwide and the increasing dissemination of drug-resistant *Salmonella* strains has prompted the WHO to include this genus on their priority list of monitored bacterial species [4][119].

During its life cycle as a facultative pathogen *salmonella* deploys several secretion systems, among them a T1SS, two T3SS, a T4SS and a T6SS [120]. The two nfT3SSs encoded by *Salmonella* are the *Salmonella* pathogenicity island-1 (SPI-1) T3SS-1 and the *Salmonella* pathogenicity island-2 (SPI-2) T3SS-2 [121]. During the initial stages of invasion, the SPI-1 T3SS is utilized to inject a range of T3SEs, notably SipA, SipC, SopE and others into the host cell to remodel the cells cytoskeleton and facilitate entry of the *Salmonella*. Inside the host, a *Salmonella*-containing vacuole/vesicle (SCV) is formed around the *Salmonella* that promotes its intracellular survival and replication. To persist in its intracellular niche and evade the host immune responses, effectors secreted via the SPI-2 T3SS mediate SCV maturation, regulate the hosts defense pathways and induce migration towards the perinuclear region of the host cell [121]. Tethered to the SCV via secreted SPI-2 effectors, the SCV remains in close proximity to the nucleus, promoting prolonged bacterial survival and replication [122]. While the T6SS also seems to contribute to the intracellular growth and replication of *Salmonella* in macrophage cells, the vast range of interactions and roles of T3SS-secreted proteins underlines the decisive importance of the type III secretion system in the context of invasion and surival [123].

1.6. Aim of this work

Using *Salmonella enterica* subsp. *enterica sv.* Typhimurium as a model organism, this work aims to advance the current knowledge how the N-terminal T3SE secretion signal promotes and regulates T3SE-mediated secretion.

In consideration of the limited success to identify key components within the secretion signal, we hypothesized that the purely qualitative analysis of sequence features of T3SEs will not be sufficient to shed further light on the signals key attributes. By investigating the putative connection between the secretion signal sequence and the secretion efficiency of a T3SE we strive to determine whether the secretion signal itself confers a modulating effect on the quantity and rate of a secreted substrate.
To address this, a quantitative T3SS secretion assay using a luminescence-based approach was established fulfilling the requirements for high-throughput, speed and simplicity of

handling. Given the large number of identified T3SEs with individual activities requiring an intricate regulation of secretion, it is envisioned to test as many secretion signals as possible to attain a finegrained vision of ranked secretion efficiencies. At this point, the quantification of signals is ongoing, therefore this work provides secretion efficiencies for the majority of identified *Salmonella* N-termini. Concomitant to the setup of the quantitative assay, the largest and most comprehensive database of experimentally verified type III secreted substrates to date was compiled, comprising both early, middle and late substrates. Initially serving the purpose of providing validated secretion signal sequences for our large-scale quantification setup, it also features annotated data about a proteins function, host cell targets and host cell localization directly connected to the information source. Overall, this makes it a valuable source not only for this work, but could also serve as a reference beneficial for future research on type III secreted substrates. The following sections describe the process of developing and establishing a quantification and analysis pipeline:

1. Endpoint quantification of secreted reporter proteins with secretion signal variants

We aim to complement a *Salmonella typhimurium* strain with plasmids carrying different N-termini fused to a reporter protein. By quantifying the amount of secreted protein after a specified timepoint, we rank the effect of individual secretion signals according to their propensity to promote T3SS-mediated secretion.

2. Database of experimentally confirmed T3SE

To test the maximal number of secretion signals and study their intricacies, a database of published, experimentally validated and non-redundant T3SS-secreted proteins comprising both structural/regulatory components and effectors proteins will be assembled.

3. Computational integration of secretion efficiencies with qualitative sequence features

Following experimental determination of secretion efficiency scores, the quantitative data needs to be associated with the respective peptide sequences. We intend to deploy machine learning frameworks for both classification and regression tasks in order to gain a deeper understanding of the signal.

2. Materials & Methods

2.1. Materials

Table 2.1.: Devices used in this work.

name	manufacturer
tuberoller	Starlabs
tuberoller SRT9D	Stuart
cell density meter WPA Biowave CO8000	Biochrom
centrifuge 5810 R	Eppendorf
centrifuge Biofuge fresco	Heraeus
tabletop centrifuge 5425R	Eppendorf
incubator HT Multitron Std	Infors
incubator Heratherm	Heraeus
Safety cabinet Maxisafe 2020	Thermo Scientific
PCR cycler T1	Biometra
spectrophotometer DS-11 Fx+	DeNovix
platereader Clariostar Plus	BMG Labtech
platereader Envision	Perkins Elmer
imager Chemostar	Intas
thermomixer	ThermoScientific
thermoblock	ThermoScientific
imager ChemiDoc MP	Bio-Rad
Western blotting transfer device	Bio-Rad
power supply PowerPac™HC	Bio-Rad
power supply PowerPac™Basic	Bio-Rad
pump	Bio-Rad
gel electrophoresis system MiniProtean	Bio-Rad

Table 2.2.: Chemicals used in this work.

name	manufacturer
agarose	Invitrogen
10 % APS	Thermo Scientific
L-(+)arabinose	Sigma
DMSO	Honeywell
20 % arabinose	Sigma
99 % methanol	Sigma
milk powder	Sucofin
4 % acrylamide solution	Roth
20 % acrylamide solution	Roth
ampicillin sodium salt	Roth
sodium chloride	Merck
sodium deoxycholate	Sigma
sodium dodecyl sulfate	Sigma
tricholoracetic acid	Sigma
Tween 20	Sigma
chloramphenicol	Sigma
kanamycin sulfate	Sigma
tetracyclin hydrochloride	Sigma
TEMED	Applichem
m-toluic acid	Sigma

Table 2.3.: Consumables used in this work.

name	manufacturer
pipette tips	Eppendorf
Nunc flat-bottom 96-well plate	Thermo Scientific
96-well LumiTrac™luminescence plate	Bio-Greiner
0.2 µm filter-membranes	Sigma
gas-permeable foil	Machery-nagel
PVDF-membrane	Merck
filter papers	Sigma
Sanger sequencing Ecoli Nightseq	Microsynth
Sanger sequencing Economy Run	Microsynth

Table 2.4.: Enzymes & assay kits used in this work.

name	manufacturer
2X repliQa HiFi ToughMix	Quantabio
PCR purification kit	Qiagen
Plasmid isolation kit QiaPrep Mini	Qiagen
In-Fusion®Snap Assembly Master Mix	Takarabio
20000x RedSafe™Nucleic acid stain	Intronbio
Gibson Assembly®Master Mix	New England BioLabs
Hifi assembly mix	QuantaBio
PageRuler™Prestained protein-ladder	Thermo Scientific
PageRuler™Plus Prestained protein-ladder	Thermo Scientific
GeneRuler 100bp DNA-ladder	Thermo Scientific
GeneRuler 1kb DNA-ladder	Thermo Scientific
GeneRuler 1kb Plus DNA-ladder	Thermo Scientific
Nano-Glo®Live cell assay kit	Promega
Nano-Glo®luciferase assay system kit	Promega
Nano-Glo®Dual-luciferase assay system kit	Promega
DpnI	New England Biolabs

Table 2.5.: Antibodies used in this work.

name	description/usage	manufacturer
α-SipA	1:5000, 1:10000	in-house
α-SptP	1:10000	in-house
α-GroEL	1:80000	Sigma
α-rabbit HRP	1:10000	Thermo Scientific
α-NC	1:10000	in-house

2.1.1. Buffers & solutions

Table 2.6.: Buffers, media & solutions used in this work.

buffer/medium	ingredient	quantity	
10x sodium dodecyl sulfate (SDS)-running buffer (V=2 L)			
	glycine	288.4	g
	Tris base	60.6	g
	SDS	20	g
	double-distilled H_2O (ddH_2O)	2000	mL
10x protein transfer buffer (PTB) (V=1.5 L)			
	glycine	58	g
	Tris base	116	g
	SDS	7.4	g
	ddH_2O	1500	mL
10x PBS (V=1 L)			
	NaCl	80	g
	KCl	2	g
	Na_2HPO_4	11.5	g
	KH_2PO_4	2	g
	ddH_2O	1000	mL
PBS-Tween 20 (PBS-T)			
	10 x PBS	1000	mL
	Tween 20	5	mL
mPBST V=50 mL			
	PBS-T	50	mL
	milk powder	2.5	g
5x Laemmli (V=60 mL)			
	SDS	6	g
	dithiothreitol (DTT)	4.64	g
	1 M Tris base (pH:6.8)	15	mL
	1 % bromphenol Blue	6	mL
	87 % glycerol	17.2	mL
	ddH_2O	21.8	mL
10x tris-acetate-EDTA (TAE)-buffer V=1 L			
	Tris	48.5	g
	glacial acetic acid	11.4	mL

Buffers, media & solutions used in this work (continued).

buffer/medium	ingredient	quantity	
	0.5 M ethylendiaminetetraacetic acid (EDTA) pH:8.0	20	mL
annealing-buffer			
	0.5 M EDTA	2	mL
	1 M Tris pH:8.0	10	mL
	1 M NaCl	50	mL
Wash buffer 1 CaCl2-competent cells (V=500 mL)			
	1 M $CaCl_2$	50	mL
	ddH_2O	450	mL
Wash buffer 2 CaCl2-competent cells (V=500 mL)			
	1 M $CaCl_2$	50	mL
	99 % glycerol	75	mL
	ddH_2O	375	mL

2.1.2. Plasmids & primers

2.1.2.1. Plasmids

Table 2.8.: Plasmids used in this work. All plasmids with initials p**JA**XXX were assembled by J.Ahrendt during her masters thesis. Plasmids with initials p**MP**XXX were assembled by myself and partly by B.Grueter.

plasmid	description/usage	reference
pWSK29		[124]
pT10_SptP_NL_myc		Wagner group, Tuebingen
pM2	pORTMAGE	[125]
pM3	pORTMAGE	[125]
pM4	pORTMAGE	[125]
pEC1	pORTMAGE	[126]
pHilA pacyc184	HilA induction	
pMP005	SptP	this work
pMP008	SptP and SipA	this work
pMP009	SptP and $SptP_{1\text{-}25}SipA_{26\text{-}685}$	this work
pMP010	$SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc and SipA	this work
pMP017	$SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc and SipA	this work
pMP028	SptP-NLuc and SipA	this work
pMP029	$AvrA_{1\text{-}25}$-$SptP_{26\text{-}543}$-NLuc and SipA	this work
pMP030	$SipB_{1\text{-}25}$-$SptP_{26\text{-}543}$-NLuc and SipA	this work

Plasmids used in this work (continued).

plasmid	description/usage	reference
pMP037	$SopD_{1-25}$-$SptP_{26-543}$-NLuc and SipA	this work
pMP049	SptP-NLuc, SipA-RFLuc	this work
pMP059	SptP-NLuc	this work
pMP071	$AvrA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP072	$GogA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP073	$GogB_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP075	$GtgE_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP076	$OrgC_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP077	$PipA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP079	$PipB2_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP080	$PrgI_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP081	$PrgJ_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP082	$SboA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP083	$SboC_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP084	$SboH_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP086	$SifA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP087	$SifB_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP088	$SipA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP089	$SipB_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP090	$SipC_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP091	$SipD_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP092	$SlrP_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP093	$SopA_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP094	$SopB_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP095	$SopD_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP096	$SopD2_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP097	$SopE_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP098	$SopE2_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP099	$SopF_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP102	$SrfJ_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP104	$SsaG_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP105	$SsaL_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP106	$SseB_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP107	$SseC_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP109	$SseF_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP111	$SseI_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP112	$SseJ_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP113	$SseK1_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP114	$SseK2_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP115	$SseK3_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP116	$SseL_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP118	$SspH2_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP120	$SteB_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP121	$SteC_{1-25}$-$SptP_{26-535}$-NLuc	this work
pMP122	$SteD_{1-25}$-$SptP_{26-535}$-NLuc	this work

Plasmids used in this work (continued).

plasmid	description/usage	reference
pMP124	$StoD_{1-25}$-$SptP_{26-535}$-NLuc	this work
pJA022	$AvrA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA023	$NleC_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA024	$GogB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA025	$GtgA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA026	$GtgE_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA027	$OrgC_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA028	$PipA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA029	$PipB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA030	$PipB2_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA031	(prgI) $PrgI_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA032	(prgJ) $PrgJ_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA033	$SboA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA034	$SboC_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA035	$SboH_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA036	$SboI_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA037	$SifA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA038	$SifB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA039	(SipA) $SipA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA040	$SipB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA041	$SipC_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA042	$SipD_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA043	$SlrP_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA044	$SopA_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA045	$SopB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA046	$SopD_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA047	$SopD2_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA048	$SopE_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA049	$SopE2_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA050	$SopF_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA053	VsdE/$SpvD_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA056	$SsaG_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA057	$SsaL_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA058	$SseB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA061	$SseF_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA063	$SseI_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA064	$SseJ_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA065	$SseK1_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA066	$SseK2_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA067	$SseK3_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA068	$SseL_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA070	$SspH2_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA072	$SteB_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA073	$SteC_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work
pJA074	$SteD_{1-25}$-$SptP_{26-535}$-NLuc, SipA-RFLuc	this work

Plasmids used in this work (continued).

plasmid	description/usage	reference
pJA076	$StoD_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc, SipA-RFLuc	this work

2.1.2.2. Primers & oligos

Table 2.10.: Primers used in this work.

primer	description/usage	sequence
MP_A09_SicP_SptP_rv	isolation SicP-SptP	TCAGCTTGCCGTCGTCATAA
MP_A10_SicP_SptP_fw	isolation SicP-SptP	AAGTAAATTGCAAGCA-CACCAG
MP_A52_SptP_ko_test_fw	sequencing SptP N-terminus	GCGTATATTGATGCCGCA-GAGA
MP_A53_SptP_ko_test_rv	sequencing SptP N-terminus	GCACCACTTCAGTGTTTT-TAAATAACG
MP_A70_SptP_upstream01_fw	gibson assembly pMP005	TCCGATGCGTAGTGAATGG
MP_A73_SptP_up01_GA_fw	gibson assembly pMP005	CGCTCTAGAACTAGTG-GATCTCCGATGCGTAGT-GAATGGCTATT
MP_A76_SptP_up_GA_rv	gibson assembly pMP005	GAGAGTCATTACC-CCAGGCGTCAGCTTGC-CGTCGTCATAAGC
MP_A77_rrnB_T1_fw	amplification of rrnB T1 terminator region	CGCCTGGGGTAATGACTCTCT
MP_A78_rrnB_T1_rv	amplification of rrnB T1 terminator region	ATTTGTCCTACTCAGGA-GAGCGTTC
MP_A79_rrnB_T1_GA_fw	gibson assembly pMP005	TTATGACGACGGCAAGCT-GACGCCTGGGGTAATGACTCT
MP_A80_rrnB_T1_GA_rv	gibson assembly pMP005	ACGGTATCGATAAGCTTCA-GATTTGTCCTACTCAGGA-GAGCGT
MP_A81_pWSK_bb_GA_fw	amplification pWSK backbone, gibson assembly pMP005	CTCTCCTGAGTAGGA-CAAATCTGAAGCTTATC-GATACCGTCGACCTC
MP_A82_pWSK_bb_GA_rv	amplification pWSK backbone, gibson assembly pMP005	GCCATTCACTACGCATCGGA-GATCCACTAGTTCTAGAGCG-GCC
MP_A85_SipA_fw	amplification SipA	CAGAAGAGGATATTAATAATG-GTTACAAGTG
MP_A86_SipA_rv	amplification SipA	CGCATCTTTCCCGGTTAAT-TAAC
MP_A87_pMP005_back-bone_GA_fw	gibson assembly pMP008	AATTAACCGGGAAAGAT-GCGCGCCTGGGGTAAT-GACTCTCTAGCTTG

Primer used in this work (continued).

primer	description/usage	sequence
MP_A88_pMP005_backbone_GA_rv	gibson assembly pMP008	ATTATTAATATCCTCTTCTGTCAGCTTGCCGTCGTCATAAGCA
MP_A89_SipA_GA_fw	gibson assembly pMP008	TTATGACGACGGCAAGCTGACAGAAGAGGATATTAATAATGGTTACAAGTGTAAGGAC
MP_A90_SipA_GA_rv	gibson assembly pMP008	GAGAGTCATTACCCCAGGCGCGCATCTTTCCCGGTTAATTAACGCT
MP_A91_SptP_Nterminus_seq_fw	sequencing SptP N-terminus	CTGCGAATAATGAAGGTACGTTAGCGTATAT
MP_A92_SptP_Nterminus_seq_rv	sequencing SptP N-terminus	TGGTCCTGTACTCTGATATTTTCCGTATGT
MP_A93_SipA_Nterm_seq_fw	sequencing pMP008	GGCGGCCCTTGTACTTAAGGATAAT
MP_A94_SipA_Nterm_seq_rv	sequencing pMP008	CAGCGCGGGAAAATCTTCCAG
MP_A95_SipA_Cterm_seq_fw	sequencing pMP008	GGCTGACCAGGCTAAAA
MP_A96_pMP009_bb_GA_fw	gibson assembly pMP009	agatcaaaacgcaggccacgGCCCGACTTTATATTGCTAAGGAAAATACT
MP_A97_pMP009_bb_GA_rv	gibson assembly pMP009	gtccttacacttgtaaccatATTCCTGCAGTATGTTTTTGAGCGCTTCCTG
MP_A98_pMP010_bb_GA_fw	gibson assembly pMP010	AAGTTGGTGTGTCGAATGATAATCTTGCGGCGAATCTTTCCG
MP_A99_pMP010_bb_GA_rv	gibson assembly pMP010	TTCTCTCCTCATACTTTAGCATTATTAATATCCTCTTCTGTCAGCTTGCC
MP_B01_SipA_Nterm_seq_fw	sequencing of SipA region	GCGCAGGAAGAAAATATGAAAACCACA
MP_B02_SipA_seq_rv	sequencing of SipA region	AAAGTTATGTTCAATGCAGCTGGCA
MP_B23_M13_fw	common primer M13 forward	TGTAAAACGACGGCCAG
MP_B24_M13_rv	common primer M13 reverse	CAG GAA ACA GCT ATG ACC
MP_B27_pMP017_bb_GA_fw	gibson assembly of pMP018	AATACTGATAAGGCATATGTTGCGCCTGAAAAATTTTCGT
MP_B28_pMP017_bb_GA_rv	gibson assembly of pMP018	TTTTCTCTCCTCATACTTTAGCATATTCCTGCAGTATGTT

Primer used in this work (continued).

primer	description/usage	sequence
MP_B29_pMP018_bb_GA_fw	gibson assembly of pMP018	tagatattagaggtagtgcgTTAATGTCCCGTATAAATATGAACAAACCCCT
MP_B30_pMP018_bb_GA_rv	gibson assembly of pMP018	gataaagttatttttgtcacTTTTCTCTCCTCATACTTTAGCATATTCCTGCA
MP_B31_pMP018_SopE_GA_fw	amplification SopE N-terminus (1-100 aa)	TAAAGTATGAGGAGAGAAAAgtgacaaaaataactttatctcccca
MP_B32_pMP018_SopE_GA_rv	amplification SopE N-terminus (1-100 aa)	ATATTTATACGGGACATTAAcgcactacctctaatatctatatcatt
MP_B47_pMP028_NL_fw	isolation of	TGCTTATGACGACGGCAAGCATGGTCTTCACACTCGAAGATTTCGTT
MP_B48_pMP028_NL_rv	assembly of pMP028	ATTAATATCCTCTTCTGTCACGCCAGAATGCGTTCGCACA
MP_B49_pMP028_bb_GA_fw	assembly of pMP028	TGTGCGAACGCATTCTGGCGTGACAGAAGAGGATATTAATAATGGTTACAAGT
MP_B50_pMP028_bb_GA_rv	assembly of pMP028	TCTTCGAGTGTGAAGACCATGCTTGCCGTCGTCATAAGCA
MP_B51_pMP030_SipB_insert_fw	assembly of pMP030	GCGGATATACCCAAAATCCGCGCCTCGCTGAGGCGGCTTTTGAAAATACTGATAAGGCATATGTTGC
MP_B52_pMP030_SipB_insert_rv	assembly of pMP030	TTTTGGGTATATCCGCTACGGCTAATGCTACTTGCGTCATTTACCATTTTTCTCTCCTCATACTTTAGC
MP_B59_pMP037_SopD_insert_fw	assembly of pMP037	TCAAAATTATACGCTTAATGAAAGTCGGCTTGCTCATCTGTTAAGCAATACTGATAAGGCATATGTTGCGC
MP_B60_pMP037_SopD_insert_rv	assembly of pMP037	AAGCGTATAATTTTGATGATTACCGAAGCTTAAAGTGACTGGCATTTTTCTCTCCTCATACTTTAGCATATTCC
MP_C14_pMP028_bb_GA_fw	gibson assembly of pMP049	ctgattagcgaagaagatttaTAATTAACCGGGAAAGATGCGCGCCT
MP_C15_pMP028_bb_GA_rv	gibson assembly of pMP049	gtccttacacttgtaaccatTATTAATATCCTCTTCTGTCACGCCAGAATGCGTT
MP_C16_SipA_RFL_GA_fw	gibson assembly of pMP049	CGTGACAGAAGAGGATATTAATAatggttacaagtgtaaggactca

Primer used in this work (continued).

primer	description/usage	sequence
MP_C18_SipA_RFL_GA_rv	gibson assembly of pMP049	ATCTTTCCCGGTTAATTAcatcttg-gccacgggtttct
MP_C19_pMP028_bb_GA_fw	gibson assembly of pMP049	gaaacccgtggccaagatgTAATTAAC-CGGGAAAGATGCGCGCCT
MP_C20_pMP049_bb_fw	backbone-PCR pMP049	aacccgtggccaagatgTAATTAAC-CGGGAAAGATGCGCGCCTGG
MP_C21_pMP049_bb_rv	backbone-PCR pMP049	atcttggccacgggtttcttcaggatctcccg-gatggccctg
MP_C22_InvI_seq_fw	sequencing InvI N-terminus	GCGTAAATTAAT-GACGCGTTTGGAAG
MP_C23_InvI_seq_rv	sequencing InvI N-terminus	GTCGCGATCTTTTTTATCACCG-GAATA
MP_C24_pMP049_SipA_exci-sion_fw	creation of plasmid pMP059, pMP071-pMP124	TGGCGTGACGCCTGGGGTAAT-GACTCTCTAGCT
MP_C25_pMP049_SipA_exci-sion_rv		CCCAGGCGTCACGCCAGAAT-GCGTTCGC

Table 2.12.: Oligos used in this work.

oligo	description/usage	sequence
BY_pORTMAGE_InvA_ko	pORTMAGE mutagenesis InvA	TGACGAACATAGAAAT-GATCATCACCATTAG-TACCAGAATCtaTcAT-TaAGGTCGTAAACGAG-CACTGTTAAGTAGA-GAAAGCAGCACTA
MP_pORTMAGE_InvC_3stop	pORTMAGE mutagenesis InvC	TAATTGGGCCGGT-TATTTTTTGTGGGTAG-GCCAGATATTGCtatTAt-CaAGGTGTTTTCATCT-CATTAGCGACCGAC-TAAAAACTTCCAG
MP_pORTMAGE_SptP_ko	pORTMAGE mutagenesis SptP	ATATAAAGTCGGGCAT-CATTCGACACAC-CAACTTTTGAAAAC-tAttACtACGTTAAAT-TATTCAATTTTCTCTCCT-CATACTTTAGCATA

2.1.3. Bacterial strains

2.1.3.1. Strain selection

Salmonella strains used in this work are listed in Table 2.13. Strain M2433 was used for most of the proof-of-principle experiments. This strain lacks the genes for flagellar

assembly and T3SS effectors SopE, SopE2, SipA, SopB, SopA, SptP, SpvB and SpvC [127]. To reduce the risk of undesired effects caused by the genotype of this strain, the flagellar knockout strain *S. typhimurium* SB905 was modified by introducing triple stop codons immediately downstream of the start codon of the respective gene using pORTMAGE mutagenesis (section 2.2.1.0.1) [125], [126]) [128]. All generated *Salmonella* strains were isogenic derivatives from either *Salmonella enterica enterica* serovar Typhimurium LT2 or SL1344 [129]. The *S. typhimurium* strain SB905 served as the basis for the majority of generated strains [128]. The resulting strains SB905 Δ*sptP*, SB905 Δ*sipA* and SB905 Δ*sptP* Δ*sipA* strains did not express the knocked-out proteins and were subsequently used for secretion quantification assays. Additionally, a knockout strain not expressing SptP, SipA and InvA was prepared as a T3SS-deficient negative control strain (SB905 Δ*sptP* Δ*sipA* Δ*invA*) to replace the T3SS-deficient SB906 strain that still expresses SptP and SipA.

Table 2.13.: *Salmonella typhimurium* strains used in this work. All generated strains were confirmed via sequencing.

strain	**genotype**	**reference**
SB905	Flg-	[128]
SB906	Flg-, Δ*prgH*	[128]
M2433	Flg- Δ*sopE*, Δ*sopE2*, Δ*sipA*, Δ*sopB*, Δ*sopA*, Δ*sptP*, Δ*spvB*	[127]
SL1344		[129]
SB905 Δ*sipA*	Δ*sipA*	this work
SB905 Δ*sptP*	Δ*sptP*	this work
SB905 Δ*sptP* Δ*invA*	Δ*sptP* Δ*invA*	this work
SB905 Δ*sptP* Δ*invC*	Δ*sptP* Δ*invC*	this work
SB905 Δ*sipA* Δ*sptP*	Δ*sipA* Δ*sptP*	this work
SB905 Δ*sipA* Δ*sptP* Δ*invA*	Δ*sipA* Δ*sptP* Δ*invA*	this work

Commercial *Escherichia* strains as listed in Table 2.14 were used for plasmid amplification and assembly using appropriate protocols described below.

Table 2.14.: *Escherichia coli* strains used in this work.

strain	genotype	reference
E. coli DH5α	F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80 *dlacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsdR17*(rK-mK+), Λ–	[130]
E. coli Top10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*dlacZ*ΔM15 Δ*lacX74* *nupG* *recA1* *araD139* Δ(ara-leu)7697 *galE15* *galK16* *rpsL*(StrR) *endA1* Λ-	

2.2. Experimental procedures

2.2.1. Cell culture

All *Salmonella* cultures were grown in lysogeny broth (LB)-0.3 M NaCl with appropriate antibiotics, if not specified differently. Strains complemented with the plasmid pHilA pacyc184 were induced with 0.012 % arabinose. *Escherichia* cultures were grown in LB with appropriate antibiotics, if not specified differently.

Glycerol stocks 5 mL *Salmonella* strains of choice were grown overnight at 180 rpm, 37 °C. 5 mL day cultures were inoculated to a starting optical density (OD) of 0.1 and incubated for 2 to 3 h at 180 rpm, 37 °C to an OD: 0.8 to 0.9. 500 µL day culture was mixed with 500 µL 50 % glycerol, labelled and shockfrozen in liquid nitrogen (LN) for storage at −80 °C until use.

T3SS secretion assays 5 mL *Salmonella* cultures inoculated from plated culture or directly from glycerol stocks were grown overnight 180 rpm, 37 °C. The next day, 10 mL day cultures (100 mL Erlenmeyer flask) were inoculated without antibiotics to a starting OD of 0.05 and incubated for 5 h at 180 rpm, 37 °C. After incubation, OD was measured and cultures normalized to OD:1.0 using LB-0.3 M NaCl. 5 mL normalized culture was transferred to round-bottom tubes and centrifuged for 10 min at 3214 rcf, 4 °C. The supernatant was filtered through a 0.2 µm membrane and samples processed as described in section 2.2.3.

Secretion quantification assays 5 mL *Salmonella* cultures of choice inoculated from plated culture or directly from glycerol stocks were grown overnight at 180 rpm, 37 °C. The next day, cultures were centrifuged for 10 min at 3214 rcf, 20 °C. The supernatant was discarded and the pellets resuspended in 5 mL LB-0.3 M NaCl before measuring OD.

For each strain carrying a different plasmid variant, 8 x 1.5 mL cultures were inoculated without antibiotics to a starting OD of 0.1 in 96-well deep well plate. The deep well plate was sealed with a gas-permeable foil and incubated for 4 h, 180 rpm, 37 °C. 30 min prior to the end of incubation, the Nanoluc substrate-mix was prepared according to manufacturers instructions and equilibrated to room temperature. After incubation, 3 x 100 µL samples of each cell culture were transferred to a transparent 96-well plate for OD-measurement. The remaining culture volume was centrifuged for 10 min at 3214 rcf, 37 °C. 3 x 100 µL supernatant per sample were transferred to 96-well luminescence plates. The remaining supernatant was either discarded or processed for western blot analysis (section 2.2.3). Pellets were resuspended in 1200 µL LB-0.3 M NaCl and 3 x 100 µL samples transferred to 96-well luminescence plates. All samples and the Nanoluc substrate-mix were equilibrated to 20 °C. Using a multipipette, 25 µL substrate-mix was added to each 100 µL sample, subsequently the 96-well plate was sealed with a PCR foil and directly quantified using a plate reader.

Plasmid variants with N-termini of all verified *Salmonella* effectors were constructed and tested. Using an adaptation of the touchdown (TD) PCR protocol, new overhangs were created to introduce secretion signals replacing the native N-terminus (section 2.2.2). Our T3SE database includes the amino acid sequences of experimentally verified effectors and not the original nucleotide sequence. To convert the secretion signal peptide sequence into a corresponding DNA sequence, a python script was written using the *Salmonella typhimurium* LT2 codon table as a basis to automatically generate overhang primers (subsubsection A.2.1.1). By default, the most frequent codon for each amino acid was selected for assembly of the present plasmids. All generated primers were manually reviewed and if necessary adjusted to required melting points. The correct assembly of the plasmids was confirmed via sequencing.

2.2.1.0.1. Time-curve quantification asssay 5 mL *Salmonella* cultures of choice inoculated from plated culture or directly from glycerol stocks were grown overnight at 180 rpm, 37 °C. The next day, cultures were centrifuged for 10 min at 3214 rcf, 20 °C. The supernatant was discarded and the pellets resuspended in 5 mL LB-0.3 M NaCl before measuring OD. A 1.5 mL culture was prepared for every 30 min of incubation time, thus for a 6 h assay 12 cell cultures per strain were prepared. The deep well plate was sealed with a gas-permeable foil and incubated for the specified incubation time at 180 rpm, 37 °C. Every 30 min, a single culture per strain was harvested and directly processed as described for the secretion quantification assays.

pORTMAGE mutagenesis Mutant strains were generated using an adaptation of the pORTMAGE protocol as described by Wannier et al. [126]. Target strains were complemented with plasmid pEC1 via electroporation. Complemented cells were grown in 10 mL

LB with appropriate antibiotics in 100 mL erlenmeyer flasks at 180 rpm, 37 °C until an OD of 0.4 to 0.6 was reached. At this point, induction of the pEC1 plasmid was started by adding 1 mL 1 M m-toluic acid. After 30 min incubation was stopped, cells transferred to round-bottom tubes and centrifuged for 10 min at 3214 rcf, 4 °C. Cells were washed in 10 mL chilled ddH_2O and centrifuged again for 10 min. This process was repeated for a second cycle before resuspension in 100 µL chilled ddH_2O was done. 81 µL cells were mixed with 9 µL 100 µM pORTMAGE oligo and kept on ice for 10 min. All oligos were designed using the mage oligo design tool (MODEST)-server. Electroporation was conducted as described in section 2.2.2. After electroporation, 910 µL prewarmed LB was added to the cells and the whole mixture added to 4 mL prewarmed LB in a 100 mL erlenmeyer flask. Cells were incubated for 30 min to 60 min prior to addition of 5 mL LB and 10 µL appropriate antibiotics. If conducting multiple pORTMAGE cycles this whole process was repeated up to 3 times per day. Typically, 4 cycles for a single mutation were conducted.

2.2.2. Cloning

Electropermeabilization *Salmonella* strains of choice were made competent for electroporetic transformation via repeated washing. 5 mL *Salmonella* cultures of choice inoculated from plated culture or directly from glycerol stocks were grown overnight at 180 rpm, 37 °C in LB and appropriate antibiotics. The next day, 5 mL day cultures were inoculated with overnight culture in LB without antibiotics to a starting OD of 0.05 and incubated at 180 rpm, 37 °C. Upon reaching an OD of 0.6-0.8, cultures were centrifuged briefly for 5 min at 3214 rcf, 4 °C. Pellets were resuspended in 5 mL chilled ddH_2O and centrifuged again. After 2 cycles of washing, cells were resuspended in 50 µL chilled ddH_2O and kept on ice until further use. A total of 100 ng to 200 ng of plasmid was added to 50 µL cells and left to rest for 10 min. Electroporation was conducted in 0.1 cm cuvettes at a constant 1800 kV in an Eppendorf 2510 electroporator. After electroporation, cells were supplied with 950 µL prewarmed LB medium and left to regenerate for 30 min to 60 min at 37 °C. Cells were centrifuged for 1 min at 4000 rpm, 800 µL supernatant removed and the pellet resuspended. 20 µL resuspended pellet were mixed with 180 µL prewarmed LB and plated onto prewarmed LB agar plates with respective antibiotics. After incubation overnight, grown colonies were picked and tested in cell culture.

Chemical transformation 5 mL *E. coli* Top10 cultures inoculated from plated culture or directly from glycerol stocks were grown overnight at 180 rpm, 37 °C in LB and appropriate antibiotics. The next day, a 250 mL culture was inoculated with 2.5 mL overnight culture and incubated to an OD:0.6 at 180 rpm, 37 °C. The culture was transferred on ice to rest for 20 min prior to centrifugation for 10 min, 3214 rcf, 4 °C. Supernatant was discarded and pellets resuspended in 15 mL ice-cold wash buffer I (see Table 2.6). The centrifugation was repeated and pellets resuspended in 15 mL ice-cold wash buffer II. Following a third

Table 2.15.: Touchdown-PCR cycle program.

phase 1	step	temperature [°C]	time	cycles
1	Denaturation	98	2 min	
2	Denaturation	98	30 s	
3	Annealing	T_m +10	30 s	10 to 15[a]
4	Elongation	72	15 s/kbp to 30 s/kbp	
phase 2				
5	Denaturation	98	30 s	
6	Annealing	T_m or T_m −5	30 s	20 to 25
7	Elongation	72	15 s/kbp to 30 s/kbp	
termination				
8	elongation	72	5 min	
9	hold	4	10 min to 15 min	

[a] **Step 3** Decrease of annealing temperature 1 °C per cycle.

centrifugation, pellets were resuspended in 4 mL wash buffer II and split into 100 µL aliquots. Aliquots not directly used for chemical transformation were frozen in liquid nitrogen (LN) and stored at −80 °C.
For immediate use, a total of 10 ng to 100 ng of plasmid was added to 50 µL competent cells. Cells were incubated at 37 °C for 45 s and left to rest on ice for 4 min. After addition of 500 µL prewarmed LB, cells were incubated for 10 min at 37 °C. Subsequently, cells were centrifuged for 1 min at 5000 rpm and the pellet resuspended in 200 µL prewarmed LB. The mix was incubated for at least20 min at 850 rpm, 37 °C, before plating onto LB agar plates with required antibiotics. Plates were incubated overnight at 37 °C, grown colonies were selected the next day for further analysis.

Touchdown-PCR All PCR reactions were conducted using an adaptation of the protocol as described by Korbie and Mattick [131]. Briefly, 50 µL PCR-mixes were assembled and subjected to the PCR protocol displayed in Table 2.15. For colony-PCR, the DNA-template for amplification was substituted with 1 µL diluted cell culture.

Gibson assembly For Gibson assembly, plasmids were linearized via PCR using suitable primers. Inserts were annealed by mixing equimolar amounts of single-stranded DNA in annealing-bufer, heated to 95 °C and subsequently cooled to room temperature (RT). Following a DpnI digest, plasmid backbone and inserts were assembled following the manufacturers instruction with ratios between 1:3 to 1:10. 25 µL Gibson assembly mix was added to 100 µL competent *E. coli* Top10 cells and processed as described in section 2.2.2.

2.2.3. Sample preparation

Secretion assay samples If not specified differently, samples from secretion assays were prepared in the following way. Pellets were resuspended in 200 µL PBS. 100 µL resuspended pellet was mixed with 100 µL 87 % glycerol and 50 µL 5 x Laemmli-buffer. 1 mL supernatant was subjected to TCA precipitation (section 2.2.3).

TCA precipitation 1 mL supernatant sample was incubated for 10 min, RT with 100 µL 0.15 % deoxycholate (DOC). Subsequently, 195 µL 100 % trichloroacetic acid (TCA) was added to the sample for a final concentration of 15 %. Samples were kept overnight at −20 °C. After overnight incubation, samples were spun down for 30 min, 28 000 rpm at 4 °C. Using a vacuum aspiration system, the supernatant was removed and the protein pellet resuspended in 500 µL ice-cold 99 % acetone. Following a 20 min centrifugation step at 28 000 rpm 4 °C. the acetone was removed using vacuum aspiration and the step repeated. After the second washing step, the protein pellet was resuspended in 80 µL PBS. 20 µL 5x Laemmli-buffer was added. If a color shift as a result of residual TCA was visible, pH was adjusted using 1 M NaCl.

2.2.4. Imaging

Agarose gel electrophoresis Depending on desired percentage 50 mL TAE-buffer was mixed with 0.5 g to 1 g LE-agarose and dissolved by heating for 50 s to 70 s in a microwave. Dissolved agarose-mix was cooled to 50 °C before adding 2.5 mL 20000x gel stain. Gels were cast and left to polymerize for a minimum of 20 min.

SDS poly-acrylamide gel-electrophoresis (PAGE) By default, handcast 4 % to 20 % gradient polyacrylamide gels were used for SDS PAGE. To 50 mL of 4 % and 20 % acrylamide stock solutions 50 µL and 20 µL TEMED was added, respectively. Gels were cast by adding 5.5 mL of each mixture to a peristaltic pump setup and adding 55 µL 10 % APS to each solution prior to casting the gel. Cast gels were left to polymerize for 45 min before usage or storage at 4 °C. After sample application gels were run for 40 min to 50 min at 55 mA constant current per gel.

Western blotting After SDS PAGE gels were left to equilibrate in PTB for 10 min. 7X9 cm PVDF-membrane were activated for 3 min in pure methanol and subsequently stored in PTB. Filter papers were soaked in PTB. Assembly of the transfer sandwich was conducted using 3 filter papers, adding the PVDF-membrane, the polyacrylamide gel and finally 3 more filter papers. Transfer was conducted for 70 min at 20 V. After transfer, membranes were blocked in 45 mL 5 % milk-PBS-T for 1 h on a tube roller at RT. Antibody solutions were freshly prepared, 10 mL of the first antibody solution was applied to each membrane and incubated overnight or for a minimum of 2 h. Membranes were washed in PBS-T 3 times for at least 20 min before applying 10 mL of the secondary antibody solution.

After incubation for 2 h,60 rpm, 20 °C on a tuberoller, chemiluminescent detection and imaging was performed.

2.2.5. Computational procedures

All generated code was written in python 3 using Visual Studio Code. Data analysis procedures and machine learning pipelines were written using JupyterNotebooks or plain python files.

Table 2.16.: Python libraries used for data analysis and machine learning pipeline development

library	version
matplotlib	3.7.1
numpy	1.24.3
pandas	1.5.3
propy3	1.1.1
rdkit	2022.09.05
seaborn	0.12.2
scikit-learn	1.2.2
scipy	1.10.1
pytorch	1.13.1
pytorch-lightning	2.0.2
ray	2.4.0
keras	2.6.0
transformers	4.28.1
uncertainties	3.1.7

Database assembly The database of Type III effector proteins was assembled in entirely manual fashion. Literature obtained from a GoogleScholar search on species-specific effectors was investigated for experiments providing evidence of T3SS-mediated secretion or translocation. Initially, all organisms mentioned in [34] were investigated for known effectors, through the literature search this was later expanded to more species. If available, protein sequences were selected directly from the publication presenting the experimental evidence. Sequences from experiments that did not display suitable negative controls were excluded. Additional information on effectors was included in the dataset if available, specifiying functions, host cell or other targets, localizations, hosts and references. Furthermore, available online databases on T3SEs from species-specific effectors were used for supplementation of the data.

Machine learning Using the Keras library, long-short-term memory network (LSTM) models were utilized for binary classification. The following settings were used.

Table 2.17.: Functions used for 2-class prediction of T3SE vs cytoplasmic proteins.

parameter	
loss function	binary crossentropy
optimizer	RMSProp
activation function	sigmoid

Statistical analysis Statistical analysis was performed using python3 and jupyter notebooks.

Secretion quantification assay analysis For individual cell cultures, the mean was calculated from triplicate measurements. For each sample variant, the median ± median absolute deviation (MAD) was calculated from 8 biological replicates using their mean signal. All subsequent calculations (e.g. normalization procedures, secretion efficiency score) were performed with error propagation.

2.2.5.0.0.1. Calculation of signal-to-noise ratio For calculation of signal-to-noise ratios in the time-curve secretion quantification experiments, the following formula was applied.

$$\text{signal-to-noise ratio} = \frac{\text{luminescence positive control}}{\text{luminescence negative control}} \tag{2.1}$$

3. Results

Disclaimer: The following section includes experimental work conducted by Jiline Ahrendt during her master thesis as part of the project. Furthermore the lab technician Barbara Grueter assisted with data generation. These experiments are marked accordingly. All analysis was conducted by me.

3.1. T3SS effector quantification

3.1.1. Qualitative detection of T3SS-secreted proteins

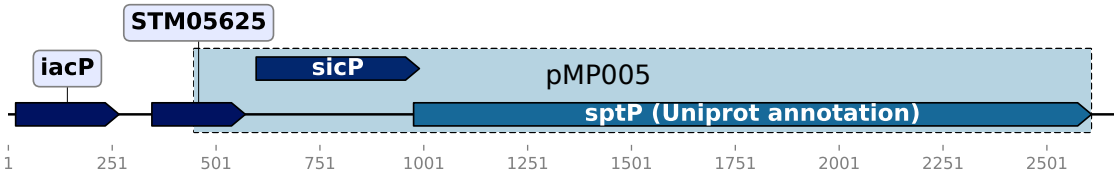
A simple and frequently applied method to monitor T3SS-mediated secretion is the Type 3 secretion assay. This type of assay enables the qualitative and semi-quantitative detection of T3SEs after incubating the bacterial strains of interest for a determined period of time [132][133][134]. The *Salmonella* T3SS secretion assay established in our lab (section 2.2.1) was modified to identify suitable conditions for conducting high-throughput quantitative detection of T3SEs. Our initial strategy encompassed complementing a *Salmonella* strain with a specific inactivated effector gene with plasmids expressing fusion proteins of various secretion signals to the respective effector.

3.1.1.1. T3SS reporter system design

Based on the low-copy plasmid pWSK29 a reporter vector was constructed [124]. The reporter protein serves as a scaffold by replacing its native N-terminal secretion signal with other confirmed secretion signals and assessing the impact on T3SS secretion quantity.
In the past, different T3SS effector proteins from *Salmonella* have been identified using classical immunoblotting techniques that rely on specific antibodies or tags fused to the candidate effector protein. To study the temporal and hierarchical regulation of the T3SS, other research groups have routinely applied the SPI-1 effector SptP as a reporter protein [85][84][135]. In the chromosome, the tyrosine-phosphatase SptP is translationally coupled to its chaperone SicP (Figure 3.1A) [136]. To mimic the native gene expression conditions of the T3SS machinery, cistrons *sicP* and *sptP* as well as a segment of the sicP upstream sequence were included on the plasmid in order to incorporate the native promoter region. The construct pMP005 was assembled, complemented into the *S. typhimurium* strain SB905 Δ*sptP* (Table 2.13) and assessed for secretion. Stable expression and secretion of SptP was confirmed, plasmid pMP005 was subsequently used as the basis for further modification (Figure 3.1B).
A common strategy to enhance the expression of *Salmonella* T3SS invasion genes utilizes the expression of the transcriptional activator protein HilA from a vector [137], [138]. In *Salmonella* HilA acts a central coordinative regulator that activates expression of invasion genes upon stimulation by environmental factors such as oxygen, osmolarity or pH [137].

Using an arabinose-inducible plasmid carrying the hilA sequence (denoted as pHilA in the following, see Table 2.8) we compared the signal strength of native and enhanced induction levels.

A



B

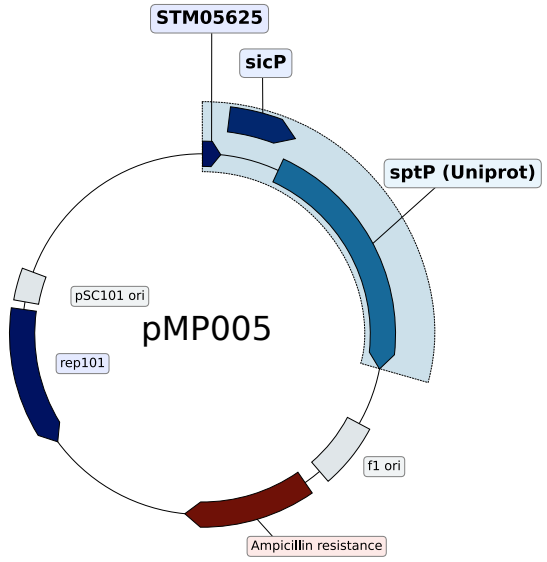


Figure 3.1.: Assembly of reporter plasmid pMP005. A Chromosomal organisation of the SicP-SptP genomic region. Construct pMP005 was assembled encompassing the genetic region upstream of *sicP*, *sicP* and *sptP*. ((highlighted in light blue ●)). **B** Plasmid map of pMP005. The nucleotide sequence highlighted in blue in Figure 3.1A was introduced into the pWSK29 backbone.

As an internal protein expression control, the *Salmonella* T3SE SipA with its 18 basepair upstream non-coding region was introduced downstream of the SptP coding sequence (CDS). SipA is encoded in the multicistronic *sipBCDA* operon downstream of SipD. By creating an artificial operon consisting of *sicP-sptP-sipA*, we intended to monitor any secretion changes of our primary reporter reflected in relation to the secretion of the unaltered effector SipA. Expression of both reporters was confirmed via Western Blotting (Figure 3.2, lanes 5-6) from strain M2433 (Δ*sopE* Δ*sopE2* Δ*sipA* Δ*sopB* Δ*sopA* Δ*sptP* Δ*spvB* Δ*spvC*), that was used in lack of a Δ*sptP* Δ*sipA* strain at that time. Secretion levels were slightly higher compared to the wildtype strain, presumably originating from the higher gene dosage of the pWSK29 low-copy plasmid.

After confirmation of stable expression and T3SS-mediated secretion of both reporter proteins from plasmid pMP008, the N-terminus of the native SptP peptide sequence was

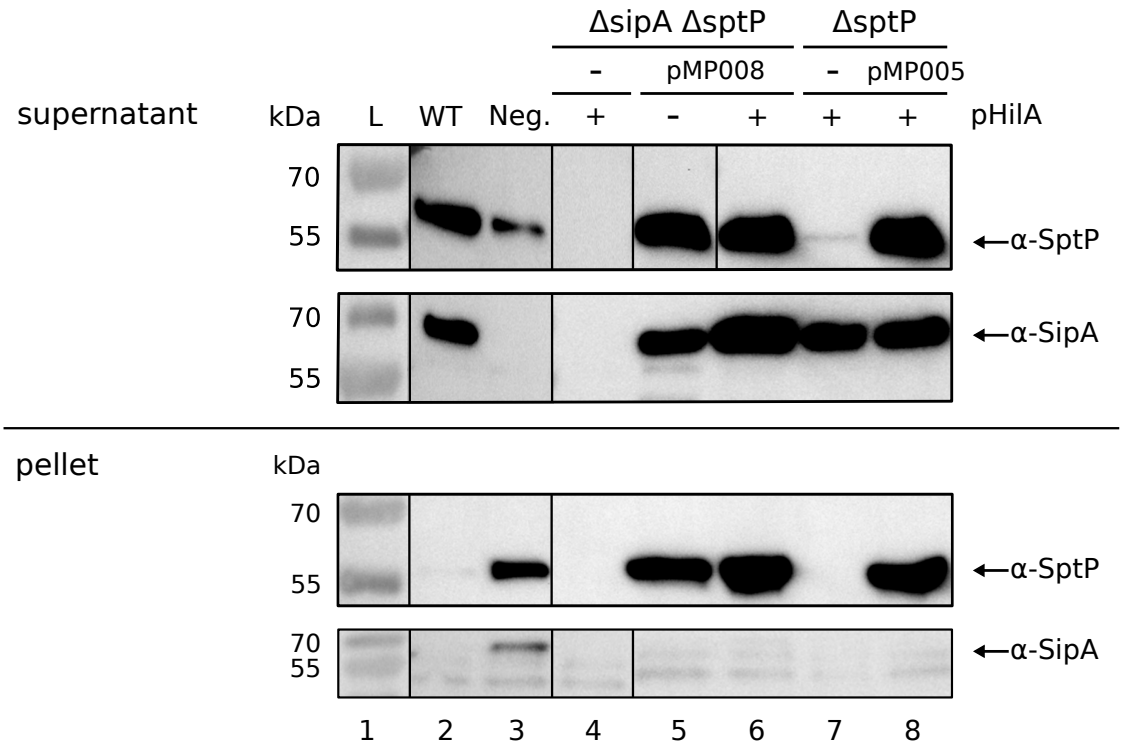


Figure 3.2.: SptP and SipA are secreted from plasmid-complemented strains. *S.typhimurium* strain M2433 (Δ*sptP* Δ*sipA* Δ*sopE* Δ*sopE2* Δ*sopA* Δ*sopB* Δ*spvB* Δ*spvC*, for brevity denoted here as Δ*sipA* Δ*sptP*) complemented with pMP008 expresses and secretes both SipA and SptP (lanes 5-6). HilA-induction leads to a slight increase of secretion. *S. typhimurium* strain SB905 ΔSptP complemented with the precursor plasmid pMP005 displays similar levels of expressed and secreted SptP (lane 8). All complemented strains expresses and secrete at higher levels than the wildtype (WT) strain *S. typhimurium* SB905 (lane 2). The slight SptP signal visible in the supernatant of the T3SS-deficient strain SB906 (ΔprgH) is a result of carryover from the WT lane. All strains were grown for 4 h at 37 °C and 180 rpm, separated into pellet and supernatant and detected via polyclonal antibodies by western blot analysis as described in section 2.2.4. Blots cropped for clarity, original blots are available in appendix A Figure A.2 and Figure A.3.

replaced with the N-terminal secretion signal of the effector SipA. We chose to replace the first 25 residues of the N-terminus to exclusively capture the secretion signal and avoid disrupting downstream chaperone binding regions.
A construct carrying a *sipA*$_{1\text{-}25}$-*sptp*$_{26\text{-}543}$ sequence (pMP009) did not express the respective fusion protein (Figure 3.3, lane 6) but only showed expression and secretion of the SipA control reporter. Upon shifting the N-terminal replacement site 24 basepairs downstream of the annotated SptP start codon (Uniprot entry: P74873 (SPTP_SALTY), 2024-03-25) to an alternative start codon, expression and T3SS-mediated secretion for a SipA$_{1\text{-}25}$-SptP$_{26\text{-}535}$ fusion protein was established (Figure 3.5, lanes 7-8). This unusual *TTG*-encoded start codon for SptP was first described by Button and Galán [136] and confirmed using a series of mutagenesis experiments (Figure 3.4). Thus, in this instance the automated sequence annotation of the Uniprot database is erroneous. We also compared native induction conditions with pHilA-supplemented induction conditions. Expectedly, HilA expression leads to an enhanced expression and secretion of the reporter protein (see Figure 3.5).

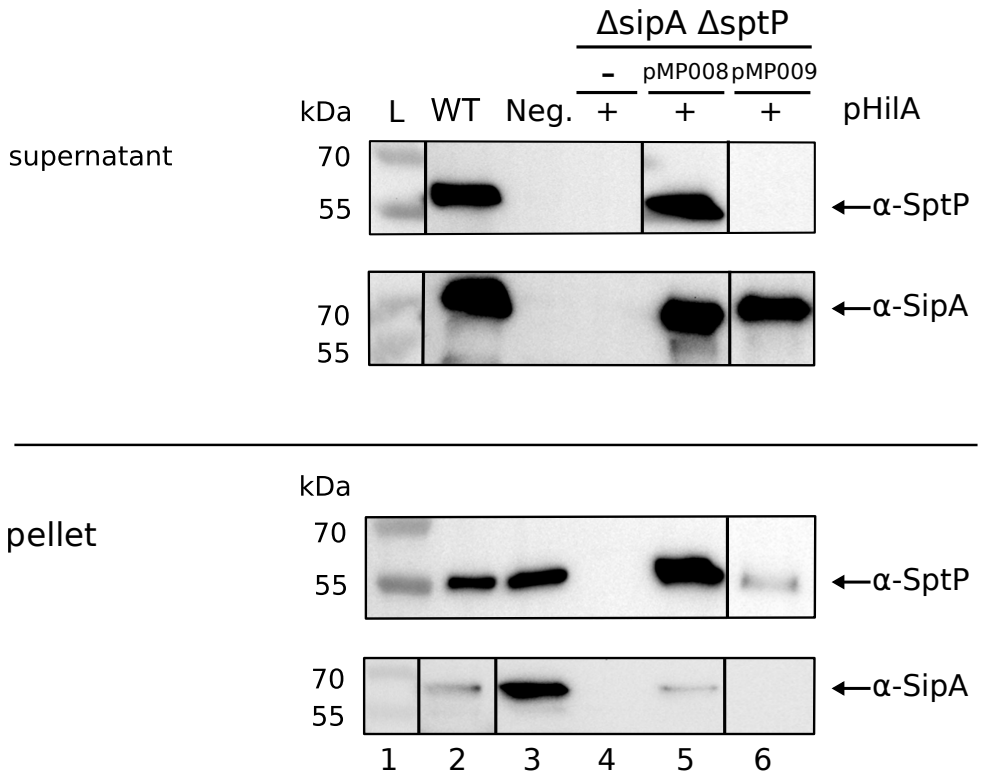
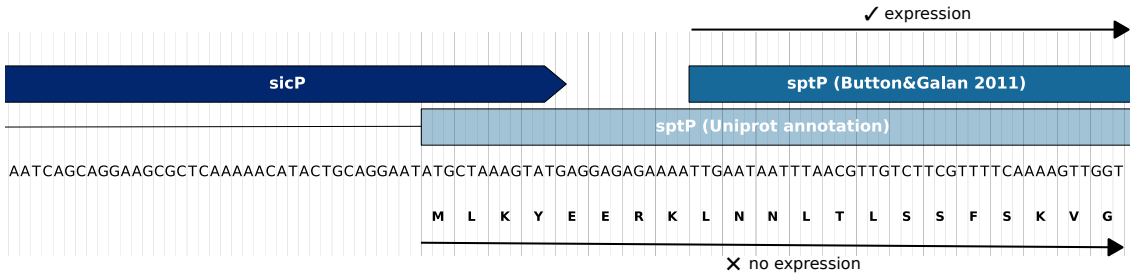


Figure 3.3.: Replacing the native N-terminus of SptP abolishes reporter expression. *S.typhimurium* strain M2433 (for brevity denoted as ΔSipA ΔSptP) complemented with plasmid pMP009 carrying the chimeric fusion protein $SipA_{1\text{-}25}$-$SptP_{26\text{-}543}$ (pMP009) did not express (lane 6), whereas strain M2433 complemented with pMP008 (lane 5) shows stable expression and secretion of native SptP and SipA. The T3SS-deficient strain negative control strain SB906 (Neg., lane 3) did not secrete expressed SptP. All strains were grown for 4 h at 37 °C and 180 rpm, separated into pellet and supernatant and detected via polyclonal antibodies by western blot analysis as described in section 2.2.4. Blots cropped for clarity, original blots are available in appendix A Figure A.4 and Figure A.5.

A



B

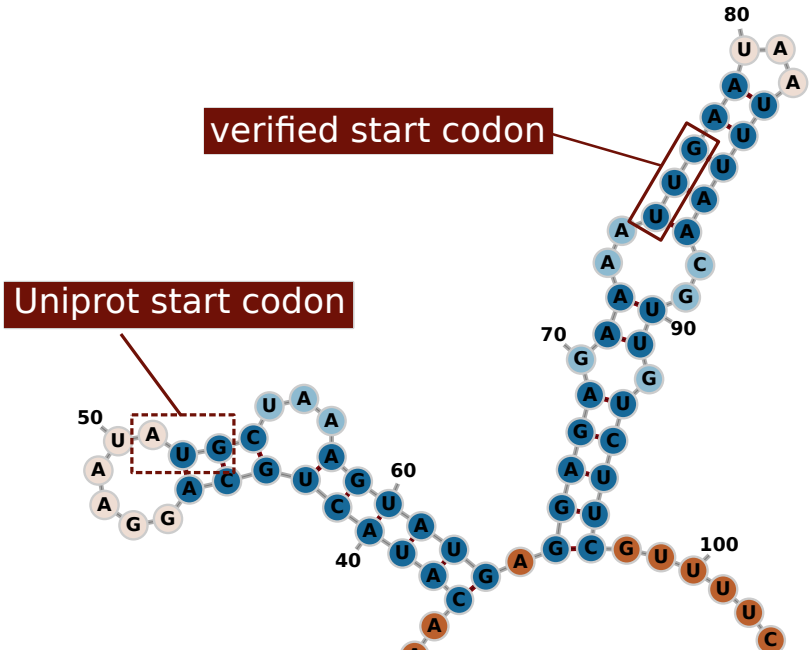


Figure 3.4.: The expression of SptP is regulated via RNA secondary structures in the N-terminus and contains an unusual start codon deviating from the Uniprot annotation. A Nucleotide sequence and open reading frames of the *sptP* N-terminal region as annotated by Uniprot (state: 2024-03-25) and by Button and Galán [136]. **B** RNA secondary structure prediction of the *sptP* N-terminus with 2 stem loops involved in regulation of SptP translation (calculated via the Vienna RNAfold webserver) [139]. The start codon of the Uniprot-deposited SptP peptide sequence is depicted in the dashed box, the experimentally verified start codon in the continuous box. This shortens the SptP effector by 8 residues [136].

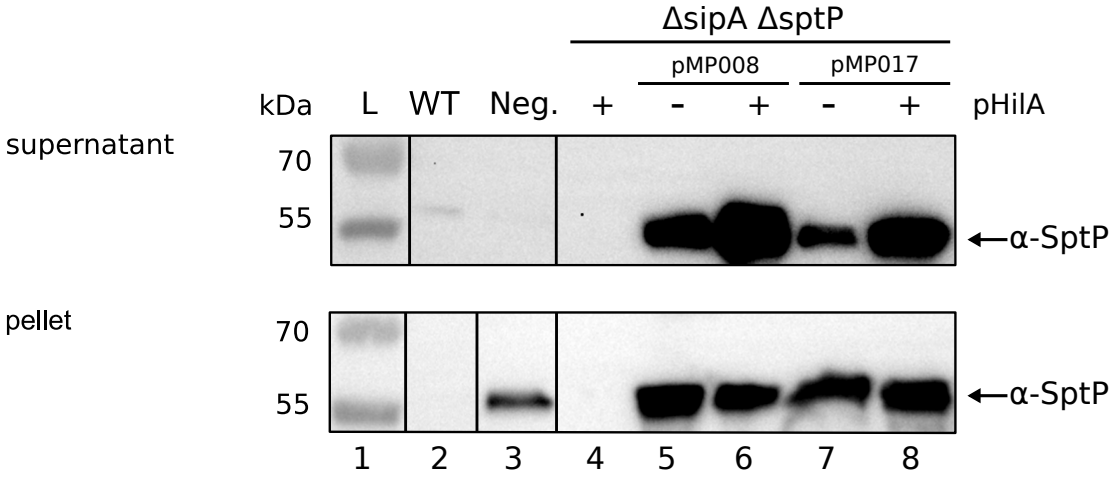


Figure 3.5.: The chimeric secretion signal-reporter fusion protein $SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$ is expressed and secreted via the T3SS. *S.typhimurium* strain M2433 complemented with plasmids carrying the wildtype SptP reporter (pMP008, lane 5-6) and the $SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$ fusion protein (pMP017, lane 7-8) expresses and secretes both proteins. To enhance expression and secretion additional complementation with the pHilA pacyc184 plasmid was tested. In comparison, the wildtype (WT) strain *S.typhimurium* SB905 expresses and secretes weaker amounts of SptP (lane 2), the T3SS-deficient negative control strain SB906 (Neg., lane 3) does not secrete SptP. All strains were grown for 4 h at 37 °C and 180 rpm, separated into pellet and supernatant and detected via polyclonal antibodies by western blot analysis as described in section 2.2.4. Blots cropped for clarity, original blots are available in appendix A Figure A.6.

3.1.2. Quantitative detection of T3SS-secreted proteins

Ensuing the successful detection of a secretion signal-reporter fusion protein using our reporter system, a suitable method to quantify T3SS-secreted proteins was investigated.

3.1.2.1. Mass spectrometric quantification of T3SE secretion

Initially, the application of a mass spectrometric approach to monitor the entire secretome of T3SS-mediated proteins was pursued [82][117]. Mass spectrometric analysis of the secretome of *Salmonella* strains SB905, SL1344 and M2433 was performed by the UKE mass spectrometric facility and to confirm the chromosomal knockout of specified T3SEs. A small fraction of *Salmonella* proteins could be identified from TCA-precipitated supernatant, the majority of proteins however were of *Saccharomyces* origin, indicating significant contamination from the LB growth medium. Of the *Salmonella* proteins identified, two hits were directly (SipA, marked bold in Table 3.1) or indirectly (DnaJ) associated with host cell invasion (Table 3.1) [140]. The constraints arising from the choice of growth medium and the elevated workload, time frame and cost of a single experiment led us to abandon a mass spectrometric approach as the main quantification method in favour of a simpler and faster approach using a reporter-tagging strategy.

Table 3.1.: *Salmonella* proteins identified via LC-MS/MS in *Salmonella typhimurium* SB905. Type III effector proteins are marked in bold.

protein name	gene name	Uniprot ID	coverage [%]
Major outer membrane lipoprotein Lpp 1	*lpp1*	Q7CQN4	18
10 kDa chaperonin	*groS*	P0A1D5	14
DNA-binding protein HU-α	*hupA*	P0A1R6	14
glyceraldehyde-3-phosphate dehydrogenase	*gapA*	P0A1P0	11
EIIAB-Man	*manX*	Q8ZP03	7
elongation factor Ts	*tsF*	P64052	7
hyperosmotically inducible periplasmic protein	*osmY*	Q7CP68	5
L-seryl-tRNA(Sec) selenium transferase	*selA*	Q8ZL69	4
Ribosome-recycling factor	*frr*	P66738	4
chaperone protein DnaJ	*dnaJ*	P0A1G7	3
glycerol-3-phosphage dehydrogenase	*glpD*	Q8ZLH4	3
Lipopolysaccharide core heptose(I) kinase RfaP	*rfaP*	Q06995	3
60 kDa chaperonin	*groL*	P0A1D3	2
cell invasion protein SipA	*sipA*	P0CL52	2
putative periplasmic binding protein	*vieP*	Q8ZKB2	1

3.1.2.2. Luminescence-based quantification of T3SE secretion

An alternative method that allows precise quantification has been described by Westerhausen et al. by tagging effector proteins with luciferase proteins [141]. Bioluminescent luciferases offer high sensitivity, enhanced signal-to-noise ratios and suitability for real-time kinetic applications. Due to its small size of 19 kDa, high brightness and stability, we set out to test the Nanoluc luciferase as the C-terminal effector tag [142] (in the following abbreviated as NLuc).

To express a Sptp-NLuc fusion protein, SptP was C-terminally tagged with the Nanoluc luciferase (Figure 3.6). Based on the created plasmid, 3 plasmid variants were assembled replacing the native N-terminus of SptP with the secretion signals of the *S. typhimurium* effectors AvrA and SopD, as well as the secreted translocon protein SipB.

In a first 8 h secretion quantification assay, secretion of the native SptP-NLuc (denoted as $_{FL}$SptP in Figure 3.7) as well as the generated reporter variants ($AvrA_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc, $SipB_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc, $SopD_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc) was quantified in the supernatant using a commercial Nanoluc kit (Table 2.4) following the methodology described in section 2.2.1.

All of the secreted fusion proteins were quantitatively detectable to different degrees in the supernatant of the cell cultures (Figure 3.7). To ascertain that the detected proteins were secreted via the T3SS, the T3SS-deficient *S.typhimurium* strain SB906 was complemented with each of the tested constructs. In comparison to the secretion pattern of the T3SS-competent *S.typhimurium* M2433 expressing native SptP-NLuc, all T3SS-deficient strains display diminished amounts of reporter fusion protein in the supernatant regardless of the

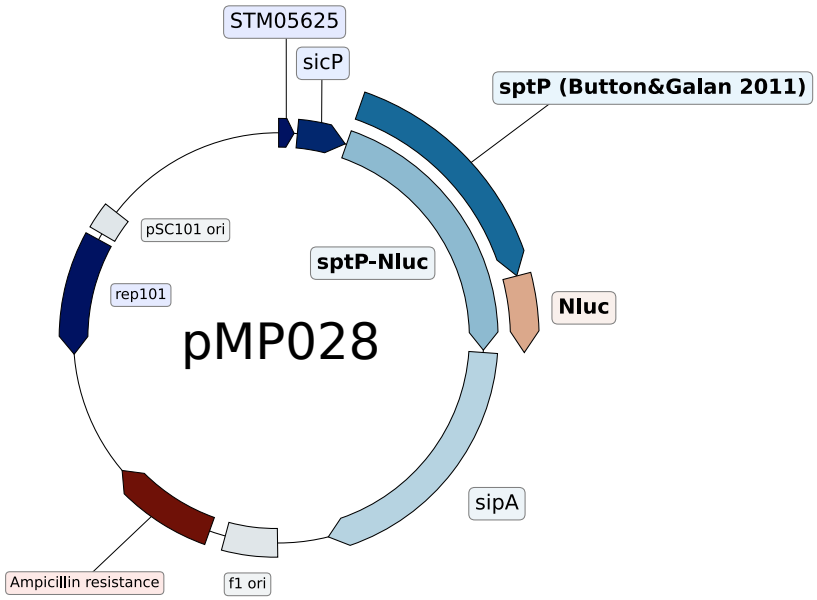


Figure 3.6.: Plasmid map of pMP028 with the native sptP sequence tagged C-terminally with the Nanoluc luciferase. This plasmid serves as the scaffold for subsequent secretion-signal variants.

---

complemented plasmid (Figure 3.8A). Qualitative detection of samples from the same assay confirms that the decrease in supernatant luminescence levels can not be attributed to reduced expression levels (Figure 3.8B). Notably, the T3SS-deficient SB906 strain still expresses the native untagged SptP protein resulting in a second band around 55 kDa.

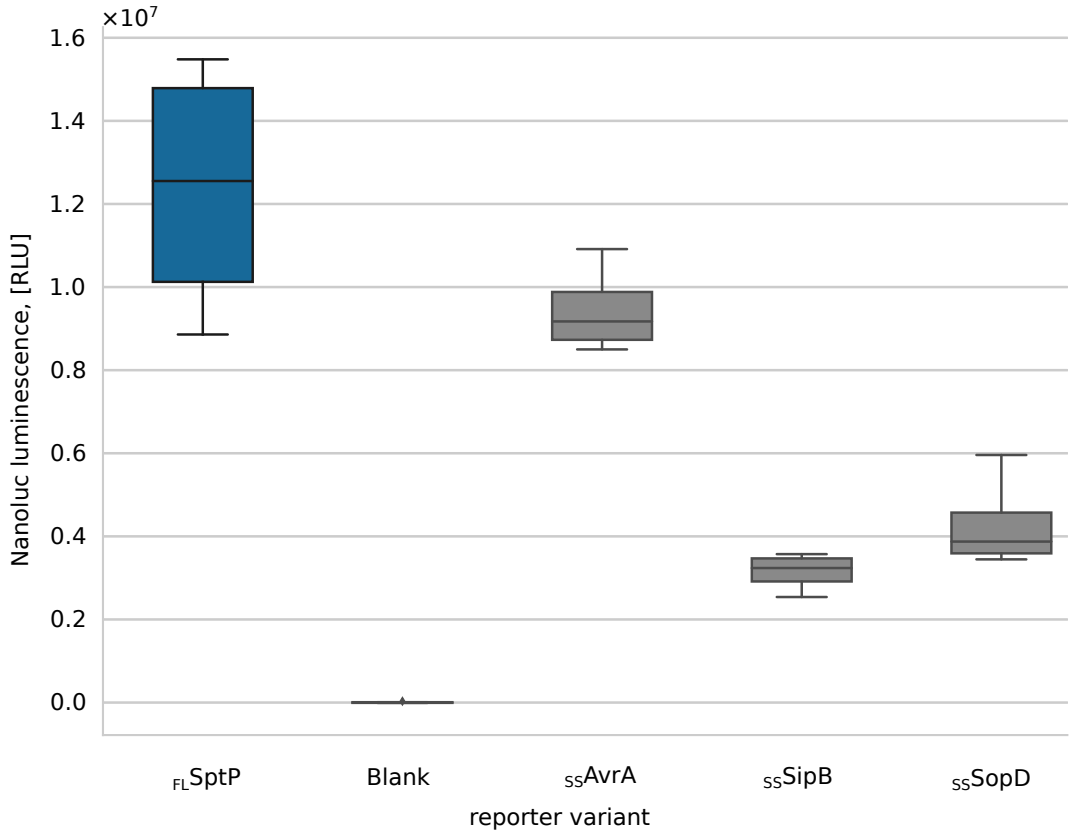


Figure 3.7.: Secretion signals from different T3SS-secreted proteins display diverging secretion quantities. *S.typhimurium* strain M2433 complemented with plasmids expressing native SptP-NLuc (denoted as $_{\text{Full length (FL)}}$SptP) or chimeric fusion proteins $AvrA_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc , $SipB_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc and $SopD_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc respectively, were grown for 8 h at 37 °C and 180 rpm. Reporter variants are denoted according to the effector protein the secretion signal (Secretion signal (SS)) is derived from. Uncomplemented cultures of M2433 were grown as a blank control. Following incubation, the amount of secreted reporter protein was quantified in the supernatant using the commercial NanoGlo Live cell assay kit (as described in section 2.2.1). Each box denotes 4 cultures per sample variant (n=4) measured in triplicate, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are displayed rhombs.

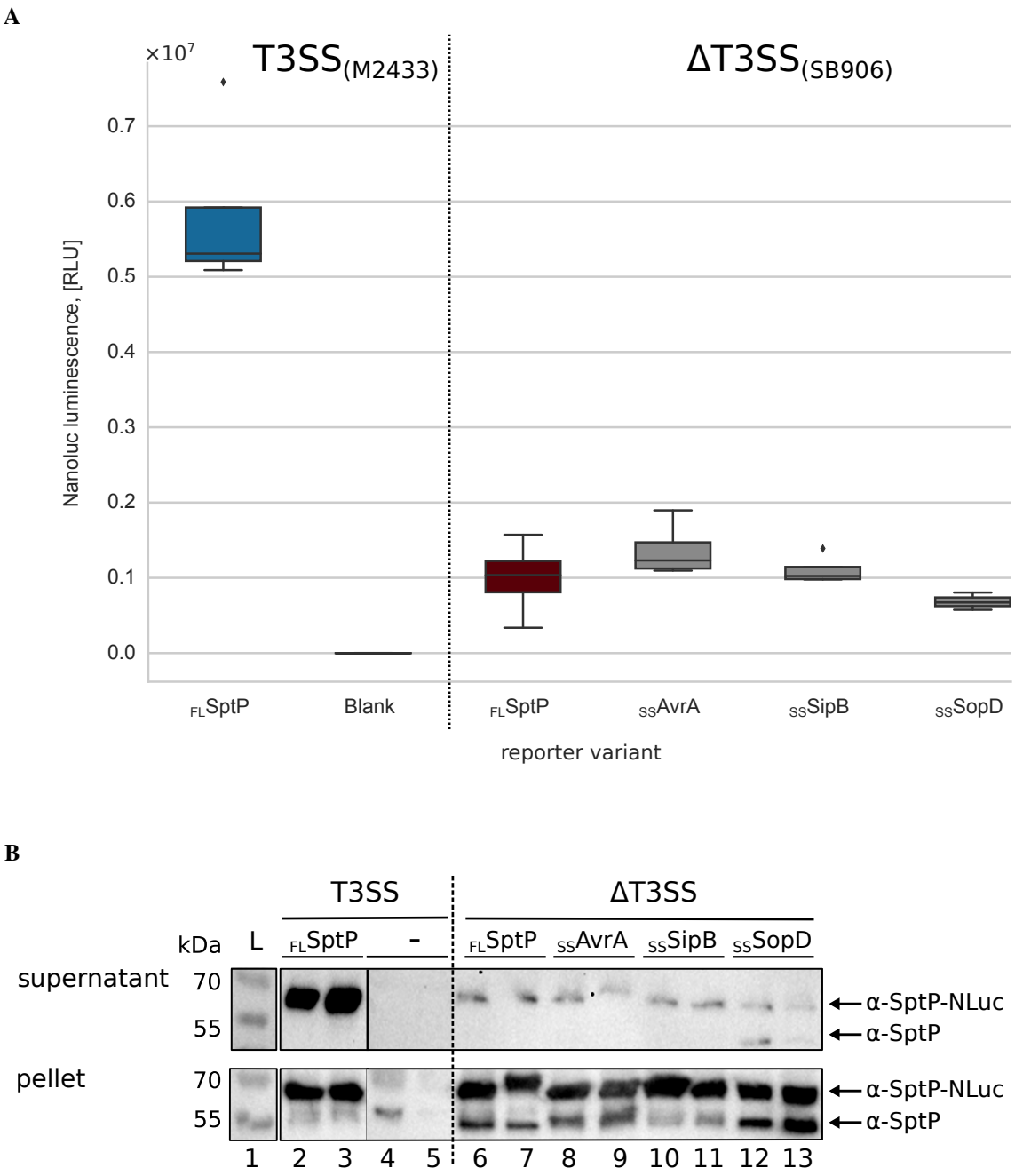


Figure 3.8.: Reporter variants are secreted in T3SS-mediated manner. A The T3SS-deficient *S.typhimurium* strain SB906 (denoted as ΔT3SS) complemented with plasmids expressing native SptP-NLuc ($_{FL}$SptP) or chimeric fusion proteins AvrA$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc, SipB$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc and SopD$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc, respectively were grown for 8 h at 37 °C and 180 rpm. Reporter variants are denoted according to the effector protein the secretion signal (SS) is derived from. T3SS-competent *S.typhimurium* M2433 expressing native SptP-NLuc ($_{FL}$SptP) and uncomplemented cultures of M2433 were grown as positive and blank control, respectively. After sample preparation, the amount of secreted reporter protein was quantified in the supernatant using the NanoGlo Live cell assay kit (as described in section 2.2.1). All tested variants display reduced signal intensities compared to the positive control. Each box denotes 4 cultures per sample variant (n=4) measured in triplicate, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are shown as rhombs **B** Additional samples of supernatant and pellet were prepared as described in section 2.2.3 and visualized using polyclonal antibodies against SptP following the protocol described in section 2.2.4. All annotation follows Figure 3.8A, except for the blank control that is denoted with a −. Original blots available in appendix A Figure A.7

A comparison of the luminescence signals of the T3SS-competent M2433 strain expressing $_{FL}$SptP-NLuc in Figure 3.7 and Figure 3.8A also revealed significant variance in signal strength between assays despite an identical experimental setup. To maintain comparability across assays we included this strain as as wildtype (WT) control strain in all subsequent assays and set the luminescence signal of each reporter variant in relation to it. The clearest indication for T3SS-mediated secretion marks the reduction of the Nanoluc luminescence signal in supernatants of the T3SS-deficient strain SB906 expressing $_{FL}$SptP-NLuc to 24 % of the signal detected in supernatants from strain M2433 expressing $_{FL}$SptP-NLuc (Figure 3.9A). The other tested reporter variants also display moderate (AvrA, SopD) to marginal (SipB) signal reductions in signal intensity between the T3SS-competent and deficient strain. Considering the incubation time of 8 h, the residual signals detected in strain SB906 might be attributed to cell lysis as an additional blotting against the cytoplasmic marker groEL indicates (Figure 3.9B).

A

	secretion ratio, [%]	
reporter variant	$T3SS_{(M2433)}$	$\Delta T3SS_{(SB906)}$
$_{FL}$SptP-NLuc	100 *a*	24 *b*
$_{SS}$Avra-SptP-NLuc	83	33
$_{SS}$SipB-SptP-NLuc	26	25
$_{SS}$SopD-SptP-NLuc	39	17

a = wildtype control (WT), all values are calculated in relation to this signal
b = negative control (Neg.), residual signal obtained from supernatants of the T3SS-deficient strain SB906.

B

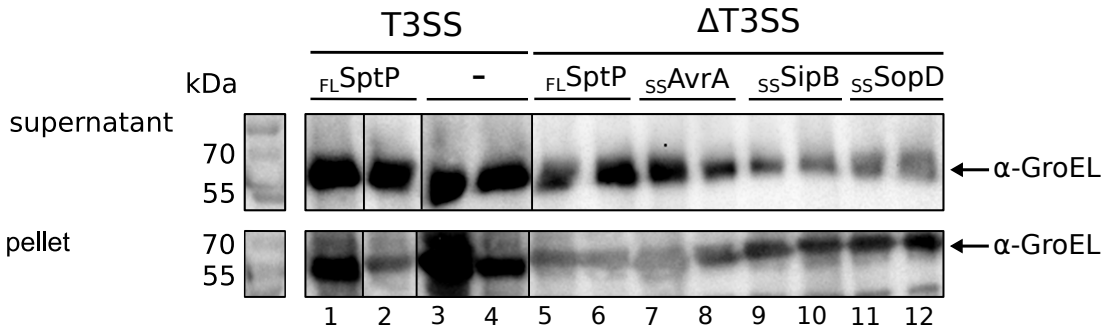


Figure 3.9.: Residual Nanoluc luminescence observed in T3SS-deficient cultures can be attributed to cell lysis. A The relative Nanoluc luminescence signal is reduced in T3SS-deficient strains. The mean Nanoluc luminescence signal from each reporter variant displayed in Figure 3.7 and Figure 3.8A was set in relation to the mean luminescence signal of strain *S. typhimurium* M2433 expressing $_{FL}$SptP-NLuc of the respective assay. In comparison, luminescence signals in the T3SS-deficient strain SB906 are reduced. Ratios were calculated using the mean value of each reporter variant. **B** Immunoblotting of the samples displayed in Figure 3.8B against the cytoplasmic marker GroEL indicates elevated levels of cell lysis. Sample annotation is identical to Figure 3.8B. Samples were prepared as described in section 2.2.3 and visualized against GroEL following the protocol described in section 2.2.4. Original blots available in appendix A Figure A.8.

3.1.2.3. Species-independent T3SS secretion

To demonstrate the wide applicability of our reporter system beyond *Salmonella* N-termini we expanded the choice of secretion signals to effectors from *Yersinia*, *Escherichia*, *Shigella* and *Chlamydia*. We confirmed varying degrees of T3SS-mediated secretion for all tested secretion signal variants(Figure 3.11). In this setup, a specifically engineered *Salmonella* strain lacking Δ*sptP* and Δ*sipA* and a strain lacking Δ*sptP* Δ*sipA* and Δ*invA* were utilized as wildtype (WT) and negative control (Neg.), respectively.

Figure 3.10.: Reporter variants with N-termini from other T3SS-carrying bacteria tested for T3SS-mediated secretion in the Nanoluc luminescence quantification assay (see Figure 3.11).

abbreviations	**reporter variant**	***Salmonella* strain**	**effector organism**
WT	$_{FL}$SptP-NLuc	SB905 Δ*sipA* Δ*sptP*	*Salmonella*
Neg.	$_{FL}$SptP-NLuc	SB905 Δ*invA* Δ*sipA* Δ*sptP*	*Salmonella*
$_{SS}$YopH	$YopH_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Yersinia*
$_{SS}$YopO	$YopO_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Yersinia*
$_{SS}$Map	$Map_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Escherichia*
$_{SS}$NleE	$NleE_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Escherichia*
$_{SS}$OspB	$OspB_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Shigella*
$_{SS}$CT_115	$CT\_115_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Chlamydia*
$_{SS}$CT_223	$CT\_223_{1\text{-}25}$-$SptP_{26\text{-}535}$-NLuc	SB905 Δ*sipA* Δ*sptP*	*Chlamydia*

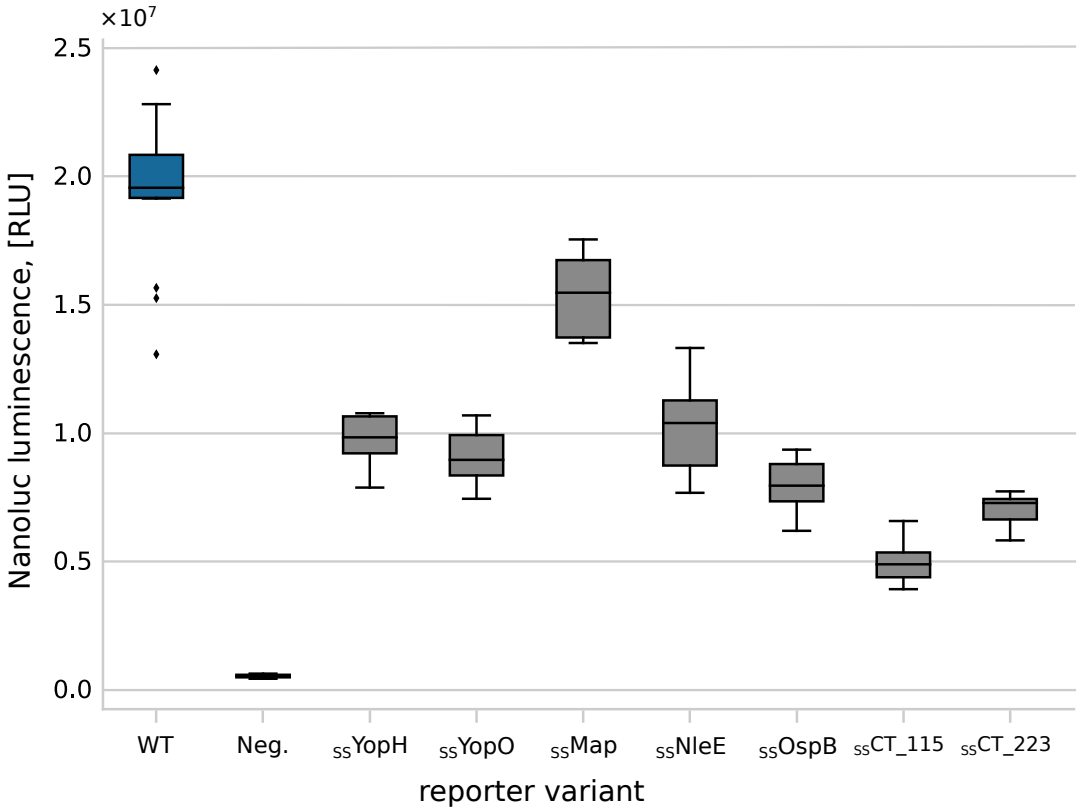


Figure 3.11.: T3SS-mediated secretion of reporter fusion proteins with secretion signals from *Chlamydia*, *Escherichia*, *Shigella* and *Yersinia*. *S.typhimurium* strain SB905 Δ*sipA* Δ*sptP* complemented with plasmids expressing $_{FL}$SptP-NLuc (WT) or reporter variants with secretion signals from *Yersinia* effectors YopH and YopO, *Escherichia* effectors Map and NleE, *Shigella* effector OspB as well as *Chlamydia* effectors CT_115 and CT_223 were grown for 4 h, 180 rpm, 37 °C (variants are annotated by the origin of their secretion signal (SS)). The T3SS-deficient strain SB905 Δ*sipA*Δ*sptP* Δ*invA* complemented with the $_{FL}$SptP-NLuc reporter and an uncomplemented SB905 Δ*sipA* Δ*sptP* strain were grown as negative (Neg.) and blank control, respectively. All strains were additionally induced via the complemented pHilA pacyc184 plasmid. After sample preparation, Nanoluc luminescence was quantified in the supernatant using the commercial NanoGlo Live cell assay kit (as described in section 2.2.1). The signal of the blank control strain was subtracted from all samples, the individual Nanoluc luminescence signals normalized to the optical density of the respective culture. Each box denotes 8 cultures per sample variant (n=8) measured in triplicate, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are shown as rhombs. Assay partly conducted by J.Ahrendt.

3.1.3. Optimization of assay conditions

After having confirmed the principal applicability of our quantification approach, a set of optimization experiments were performed to enhance the validity and reproducibility of the assay. Cell culture conditions were optimized regarding incubation time and culture volume to obtain a robust luminescence signal whilst preserving the feasibility of conducting the experimental setup. Furthermore, both the luminometer device settings as well as the luminescence assay conditions were examined to increase the reproducibility of the assay.

3.1.3.1. Culture volume & sample number

To enable a larger sample throughput, all quantification assays were performed in 96-well deep well plates suitable for cell culture. To detect the progression of T3SS secretion and determine an optimal endpoint for incubation cell culture volumes of 1.5 mL and 2 mL were monitored in a 6 h timecurve experiment. Despite sealing the plates with lids, we observed converging signal intensities in the 2 mL culture volume setup, as exemplified by the increase in signal intensity of the blank control (Figure 3.12). This effect was not detectable in the 1.5 mL culture setup. Consequently, subsequent cell cultures were adapted to 1.5 mL culture volume, as an additional precaution the sample culture variants were separated by an empty well column in the 96-well format to minimize mixing effects (Figure 3.13).

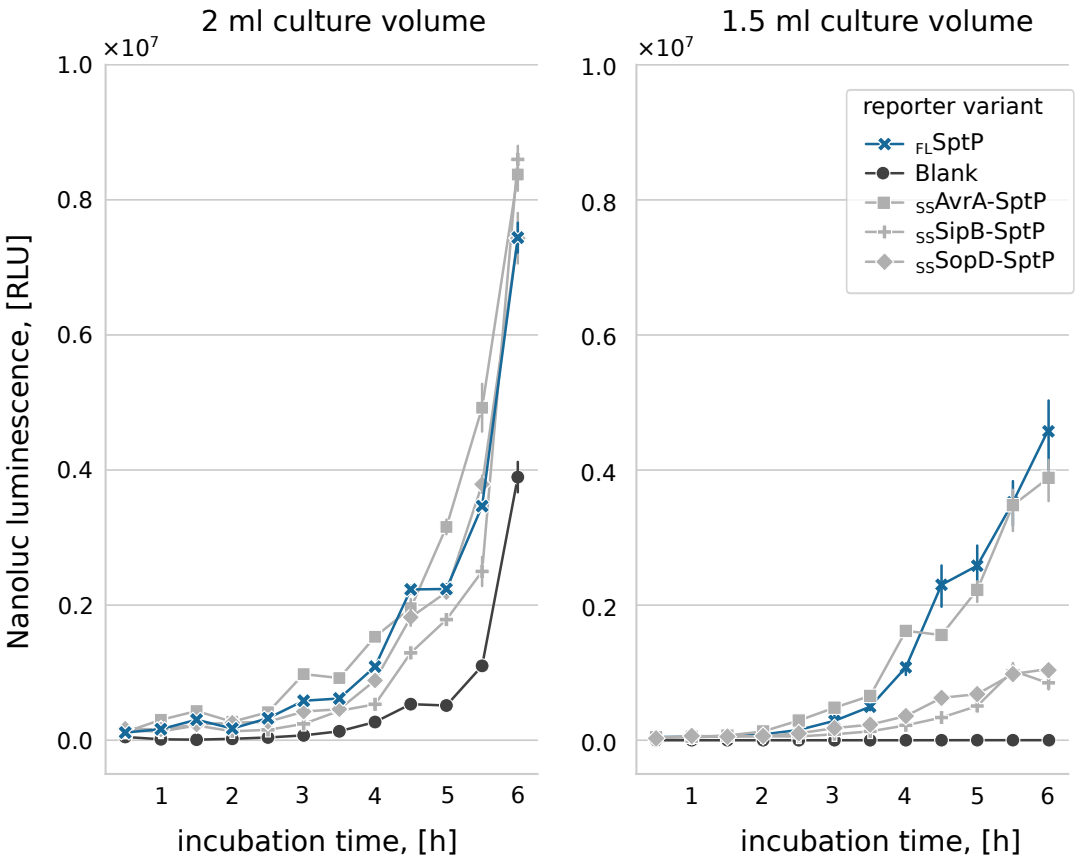


Figure 3.12.: Nanoluc luminescence signals converge due to sample mixing. 12 x 2 mL cultures of *S.typhimurium* strain M2433 complemented with plasmids expressing $_{FL}$SptP-NLuc or chimeric fusion proteins AvrA$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc, SipB$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc and SopD$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc were grown for 6 h at 180 rpm, 37 °C in a 96-well deep well plate. An uncomplemented M2433 culture was grown as a blank control. Every 30 min one culture was harvested and processed (see paragraph 2.2.1.0.1). After sample preparation, the amount of secreted reporter protein was quantified in the supernatant using the commercial NanoGlo Live cell assay kit (as described in section 2.2.1). All strains display similar secreted Nanoluc luminescence quantities throughout incubation, the increase of signal in the blank control indicates sample mixing. Each datapoint denotes a single culture measured in triplicate with standard deviation.

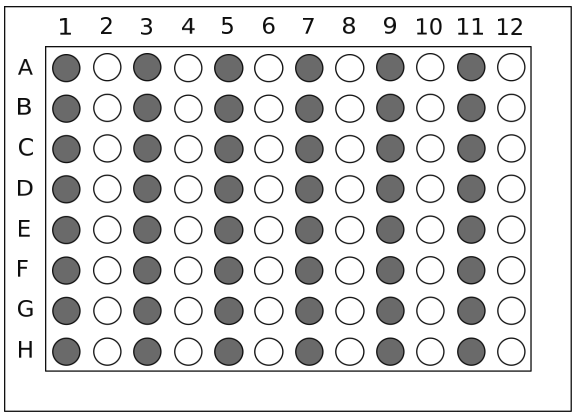


Figure 3.13.: Arrangement of cell cultures in the 96-well deep well plate incubation setup. Incubation of cell cultures in the planned 96-well format requires gap columns between strains to avoid sample mixing. Each column containing cultures of a single *S.typhimurium* strain variant are separated by an empty well column to avoid mixing of cell cultures as observed in Figure 3.12. Wells filled with cell culture are marked in dark grey.

3.1.3.2. Cultivation time

The observed cell lysis during cultivation (appendix A Figure A.8) indicated unfavourable assay conditions for *Salmonella* survival. Naturally, the dimensions of the deep well plate limit the surface area for air exchange resulting in oxygen-deprived culture conditions. In order to find an acceptable compromise between between signal intensity and cell lysis levels, a timecurve experiment was performed with sample collections every 30 min (see section 2.2.1). After normalization to an optical density of 1.0, the signal-to-noise ratio was determined by dividing the luminescence signal of the T3SS-competent M2433 strain expressing $_{\mathrm{FL}}$SptP-NLuc with the luminescence signal of the T3SS-deficient negative control SB906 expressing $_{\mathrm{FL}}$SptP-NLuc (see Equation 2.1). Signal ratios with a ratio > 2.0 were defined as sufficient and obtained for timepoints between 3.5 to 5.5 hours incubation time with an optimal ratio after 5 hours incubation (Figure 3.14). As a tradeoff, between signal-to-noise ratio and the necessicity of increasing sample throughput, we set the cultivation endpoint to 4 h to duplicate the number of samples processible in one experiment.

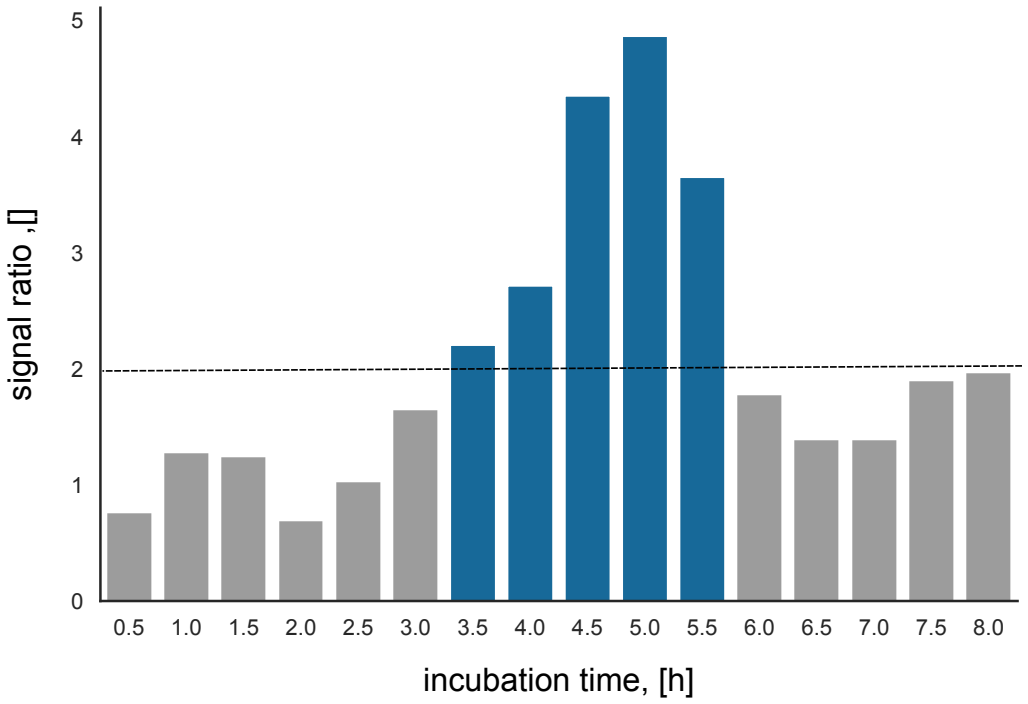


Figure 3.14.: Incubation times between 3.5 h **to** 5.5 h **yield the highest Nanoluc signal-to-noise ratios.** 16 cultures of strains M2433 and T3SS-deficient SB906 both expressing $_{FL}$SptP-NLuc were incubated over 8 h at 180 rpm, 37 °C. Every 30 minutes, one culture was harvested and processed for quantification (as described in paragraph 2.2.1.0.1). All data shown displays the mean from single biological replicates, measured in triplicate. The chosen threshold of 2.0 denotes the range between 3.5 h to 5.5 h that was determined suitable for analyzing culture endpoints (●).

3.1.3.3. Reporter expression control

The construction of effector fusion proteins alters the nucleotide sequence composition in the N-terminal region of the fusion construct. As shown previously, the N-terminal region of *sptP* within the *sicP-sptP* genomic region is tightly regulated via RNA secondary structures [136]. To monitor whether an alteration of this region causes differential secretion patterns, the second reporter protein SipA was C-terminally tagged with the Red Firefly luciferase (Figure 3.15A, described in subsubsection 3.1.1.1). The quantification of this effector should provide a consistent signal strength allowing to determine changes in the secretion patterns of the primary reporter variants. Although smaller tags seem more appropriate than the 60 kDa Red Firefly luciferase, many luciferases are not compatible with the Nanoluc luciferase. The Nanoluc substrate furimazine is a derivate of coelenterazine, a substrate metabolized by common luciferases such as *Gaussia* or *Renilla* luciferase. Therefore, crossreactivity between the luciferases and their substrates cannot be excluded, which limits the range of applicable luciferases. The Red Firefly luciferase requires D-luciferin, oxygen and ATP for activity. Expression and secretion of both reporters from a plasmid expressing $_{FL}$SptP-NLuc and SipA-RFLuc was confirmed via Western blotting (Figure 3.15B).

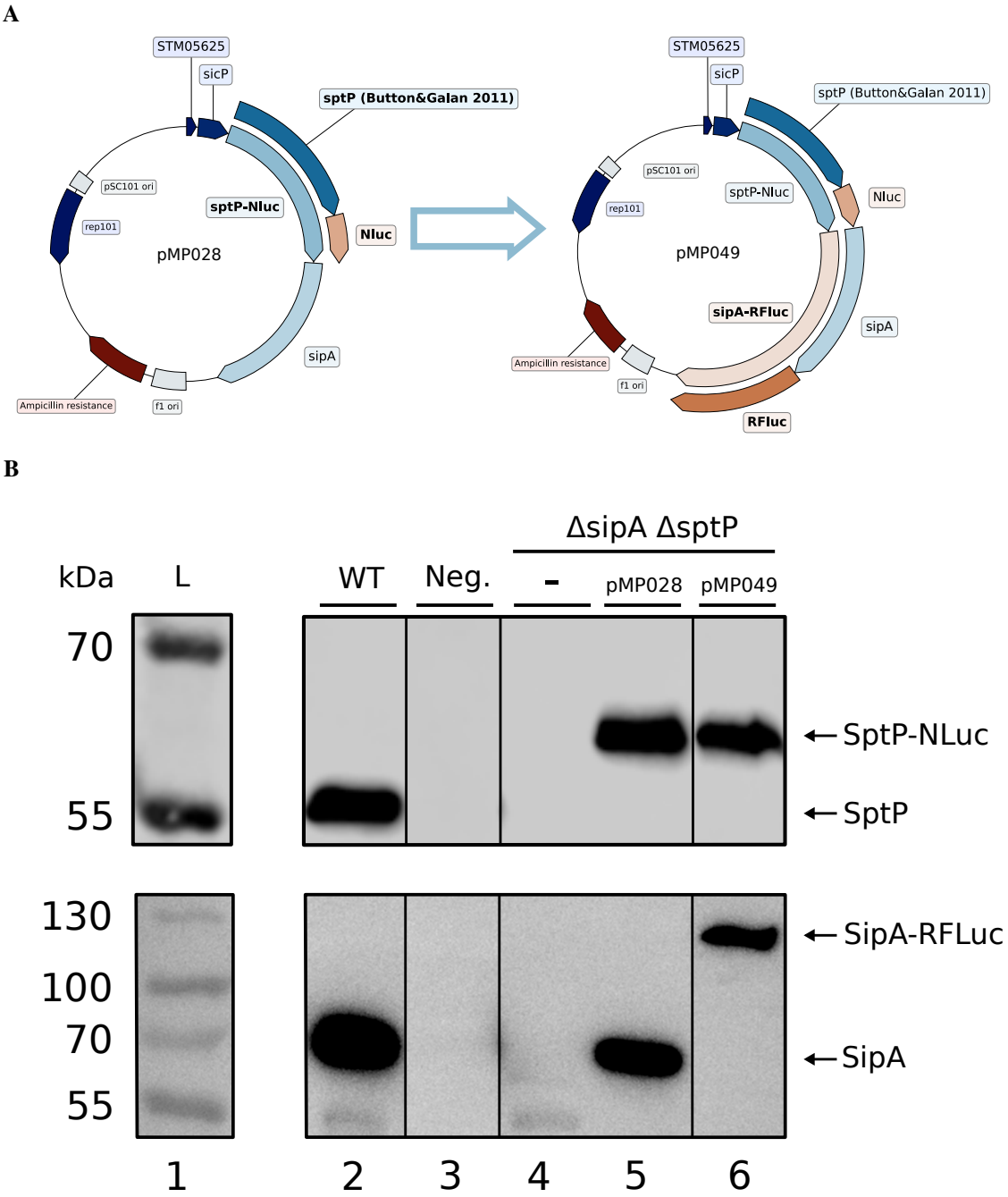


Figure 3.15.: The dual-reporter plasmid pMP049 expresses and secretes SptP-NLuc and SipA-RFLuc. A Assembly of plasmid pMP049 with Nanoluc (NLuc)-tagged SptP and RedFirefly luciferase (RFLuc)-tagged SipA using plasmid pMP028 as a template. **B** *S. typhimurium* strain SB905 Δ*sipA*Δ*sptP* complemented with pMP028 or pMP049 expresses and secretes $_{FL}$SptP-NLuc and SipA or $_{FL}$SptP-NLuc and SipA-RFLuc, respectively. All strains were complemented with pHilA pacyc184 (pHilA), including wildtype (WT) strain SB905 and T3SS-deficient strain SB906 (Neg.). All strains were grown for 4 h, 180 rpm, 37 °C, separated into pellet and supernatant and visualized using specific antisera against SptP and SipA by western blot analysis (as described in section 2.2.4. Blots cropped for clarity, assay conducted by J.Ahrendt.

3.1.3.4. Absorbance calibration

For a more robust statistical validity of the assay, all culture variants were grown in multiple biological replicates (n= 4 to 8). To account for differences in growth between individual cultures, an endpoint absorbance measurement was performed to calculate the optical density of each individual culture. The optical density measurement was then applied to normalize the obtained luminescence signals. To preserve the high-throughput nature of the quantification assays, absorbance values (600 nm) obtained from platereader measurements in the 96-well format were calibrated to the optical density cuvette measurement in a cell density meter (Figure 3.16). All subsequently obtained platereader absorbance values were converted using the derived calibration formula (Equation 3.1).

$$f(x) = 3.85 * x + 0.03 \tag{3.1}$$

During preparation of the luminescence samples, the samples for the absorbance measure-

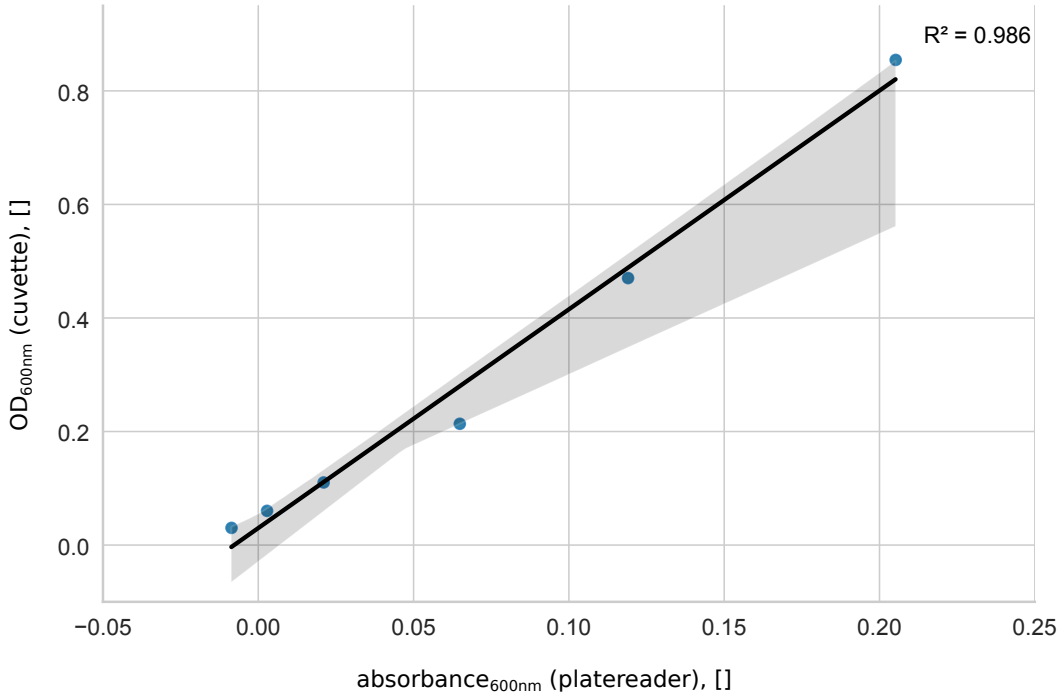


Figure 3.16.: The calibration of absorbance measurements enables the high-throughput measurement of cell culture optical densities using a platereader. After 4 h incubation at 180 rpm, 37 °C, cell culture dilutions of strain *S. typhimurium* SB905 Δ*sptP* expressing $_{\text{FL}}$SptP-NLuc ($\frac{1}{2}$, $\frac{1}{5}$, $\frac{1}{10}$, $\frac{1}{20}$, $\frac{1}{50}$, diluted in LB-0.3 M NaCl) were measured in a cuvette cell density meter (pathlength 1 cm) to obtain optical density values ($OD_{600\,nm}$). Subsequently, $absorbance_{600\,nm}$ was measured in 100 µL aliquots of the identical sample (n=6) in a flat-bottom 96-well plate using a platereader. The formula (Equation 3.1) obtained from fitting a linear regression was utilized to convert all platereader absorbance measurements into optical density values.

---

ment were taken before further processing of the cell culture and separation into supernatant and pellet samples. To assess whether the resting period from sample taking to sample

measurement influences the absorbance measurement, we compared absorbance signals of rested samples with samples that were shaken immediately before the measurement. The effect of shaking on the measured absorbance is negligible (Figure A.9), to ensure appropriate mixing of the samples we nonetheless retained a preceding shaking step in all subsequent assays.

3.1.3.5. Device settings

For the simultaneous measurements of all samples in 96-well plates, the Clariostar Plus platereader from BMG Labtech was used. Different device settings were tested to optimize the luminescence measurement protocol. A backlight filter is commonly applied to minimize the influence of residual light on luminescence and fluorescence measurements and enhance the signal-to-noise ratio. The application of such a filter reduced the average signal intensity up to 15.55 % but also decreased residual light by a magnitude (see Figure 3.17, WT sample). The uncomplemented M2433 strain not expressing a luminescent protein displays an average luminescence signal decrease from 1058 RLU to 101 RLU.

The Clariostar platereader offers the option to use an Enhanced-Dynamic Range (EDR) feature allowing signal detection over 8 concentration decades. In comparison to the standard dynamic range, EDR displays elevated signal intensities for strong luminescence signals, effectively increasing the resolution of the measurement (Figure 3.18). We also assessed the impact of a sample shaking step prior to the luminescence measurement as previously performed for the absorbance measurement. As seen in Figure 3.19, shaking decreases the average signal intensity of the samples with increasing effect on larger signal intensities. A homogeneous sample was deemed desirable in terms of reproducibility of our assay, therefore all subsequent assays included a shaking step prior to each luminescence measurement.

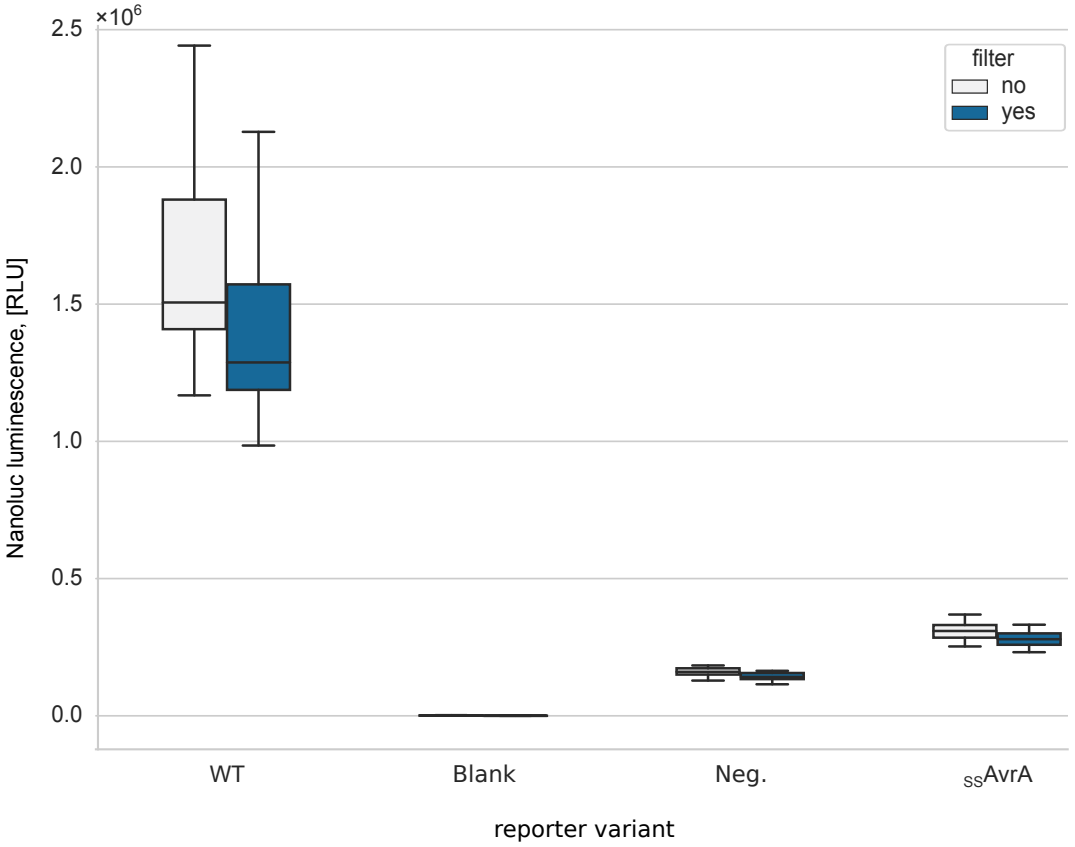


Figure 3.17.: Application of a residual light filter reduces the average signal intensity. Nanoluc luminescence from supernatants of *S. typhimurium* strain SB905 Δ*sptP* expressing $_{FL}$SptP-NLuc or $AvrA_{1\text{-}25}$-SptP26-535-NLuc, an uncomplemented SB905 Δ*sptP* as a blank signal control and the T3SS-deficient SB905 Δ*sptP*Δ*invA* as negative control (Neg.) was quantified first without applying a filter reducing residual light signals. Subsequently, the same samples were remeasured using a residual light filter. The usage of such filter reduces the average signal up to 15.55 % (WT signal). Measurements were conducted using a BMG Labtech Clariostar Plus platereader. Each box denotes 8 biological replicates measured as triplicates, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are shown as rhombs.

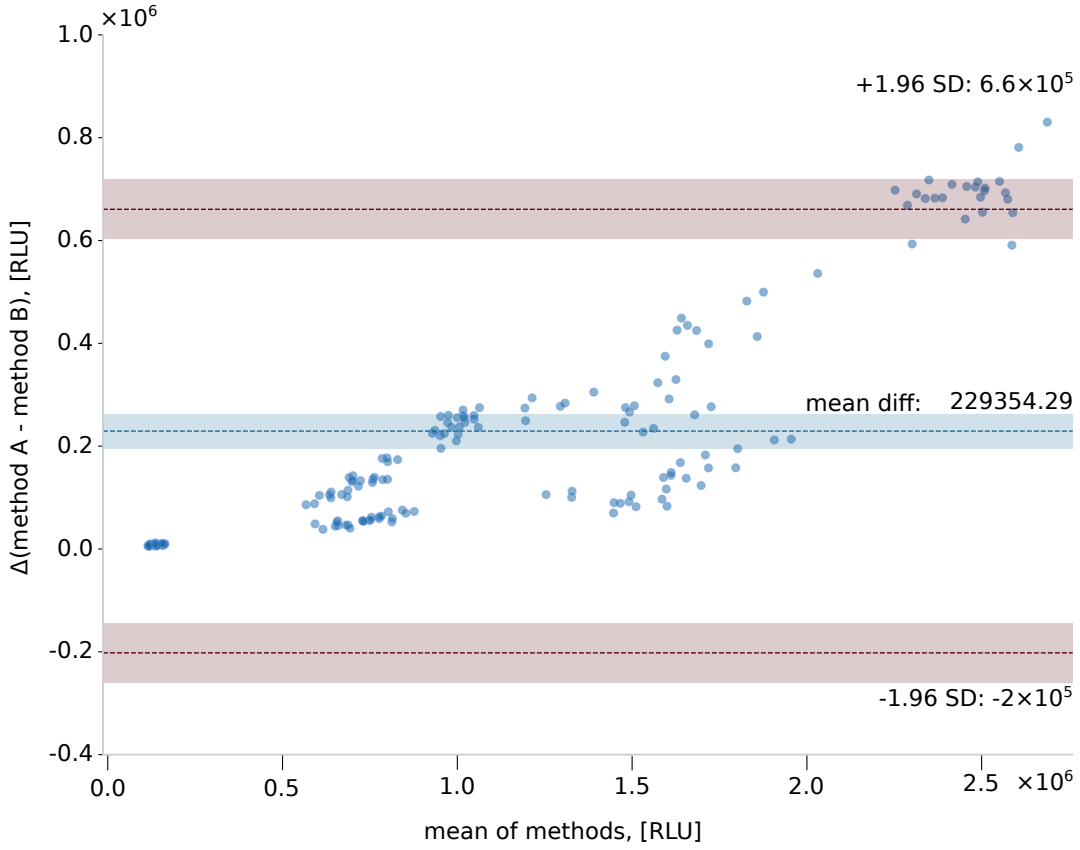


Figure 3.18.: Enhanced dynamic range (enhanced dynamic range (EDR)) increases the resolution of luminescence signals. The Nanoluc luminescence signals from supernatants of 8 sample variants (192 individual measurements in total) were quantified with and without using the Enhanced Dynamic Range (EDR) feature of the BMG Labtech Clariostar Plus platereader (EDR=method A, no EDR=method B). The EDR setting of the platereader enhances the detection range of the luminescence signal enabling concomitant measurement of both very weak and very high signal intensities. With increasing signal intensities. the divergence between method A and method B becomes more pronounced; the mean difference over all sample points amounts to 229 354 RLU. The mean value of both measurement methods (X-axis) is plotted against the measurement difference between both methods (Y-axis) for a given data point (mean_diff=mean difference, SD=standard deviation).

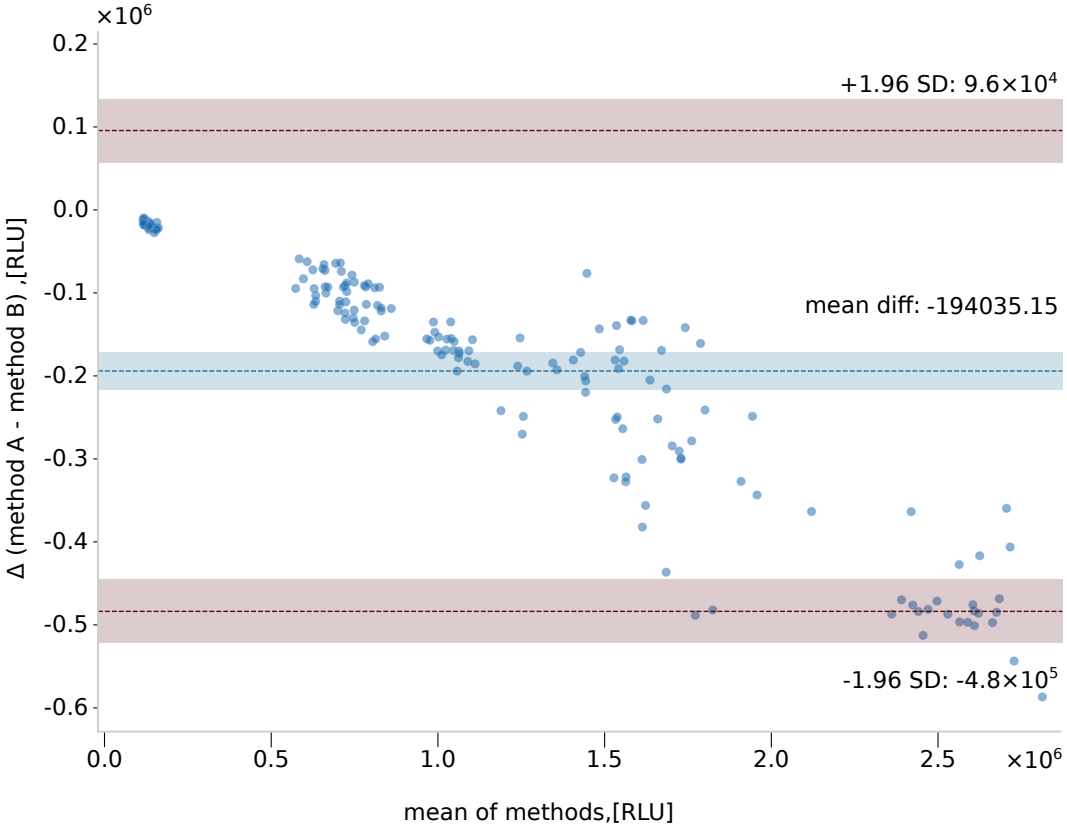


Figure 3.19.: Prior sample shaking reduces the average luminescence signal intensity. The Nanoluc luminescence signals from supernatants of 8 sample variants (192 individual measurements in total) were quantified with and without a sample shaking routine (30 s, 300 rpm) prior to luminescence quantification (shaking = method A, no shaking = method B). With increasing luminescence signal intensity, the effect of shaking becomes more pronounced, reducing the average signal strength by an average of 194 035 RLU over all sample points. The mean value of both measurement methods (X-axis) is plotted against the measurement difference between both methods (Y-axis) for a given data point (mean_diff=mean difference, SD=standard deviation).

3.1.3.6. Time & temperature effects

The Nanoluc luciferase is a glow-type luciferase with a declared signal half-life time of 120 min at room temperature [143]. We examined the signal stability in our assay conditions in a 120 min-timescale experiment with measurement intervals of 30 min (Figure 3.20). As opposed to the manufacturers specifications, we observed a 4-fold signal decay in the wildtype (WT) sample expressing $_{FL}$SptP-NLuc. The detected reduction is more pronounced at higher initial signal intensities.

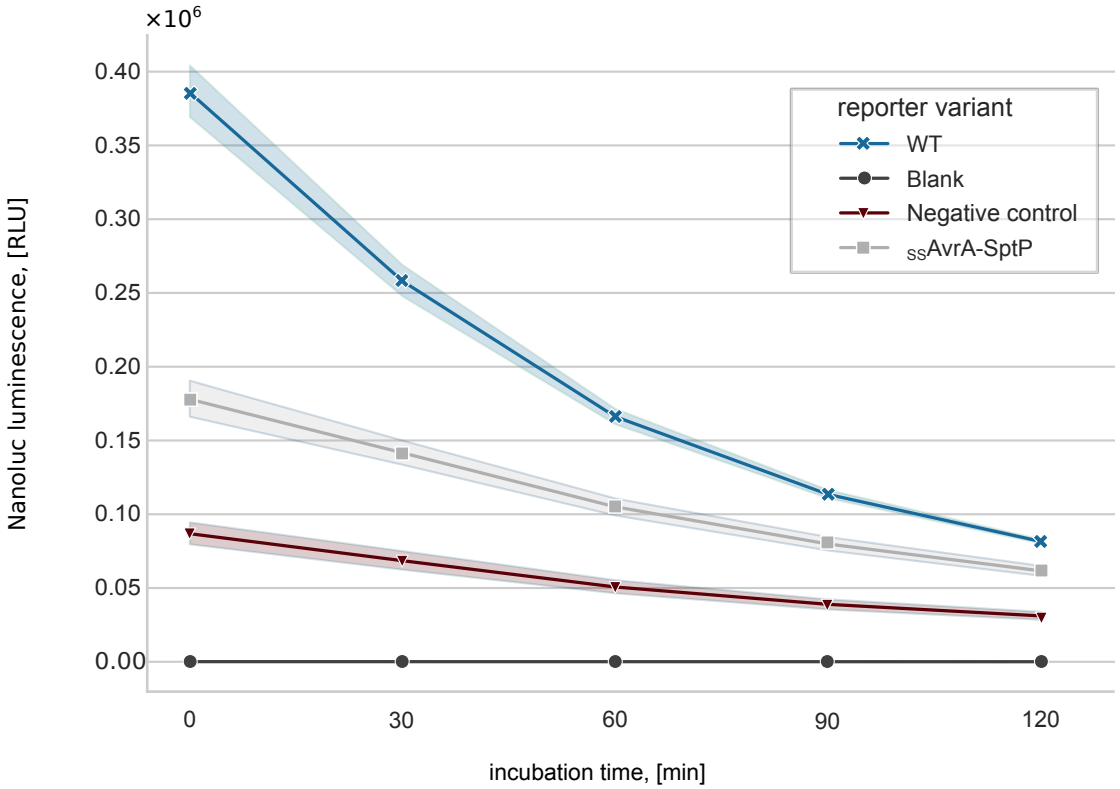


Figure 3.20.: The Nanoluc luminescence intensity decays faster in high intensity samples. Nanoluc luminescence signals in the supernatant of 4 samples with different initial luminescence intensity (WT = SB905 Δ*sptP* expressing $_{FL}$SptP-NLuc, SB905 Δ*sptP* expressing AvrA$_{1\text{-}25}$-SptP$_{26\text{-}535}$, uncomplemented SB905 Δ*sptP* (Blank), negative control = SB905 Δ*invA* Δ*sptP*) were quantified in 30 min intervals for a duration of 120 min. Prior to each measurement, the 96-well plate was shaken for 30 s at 300 rpm. Each datapoint represents 8 biological replicates measured in triplicate, displayed is the median value and standard deviation.

We also tested the temperature dependency of the Nanoluc luciferase during the luminescence measurement, for 20 °C and 37 °C, respectively. A selection of samples was first measured in a single 96-well luminescence plate equilibrated at room temperature, the same plate incubated for 15 min at 37 °C after the measurement and remeasured again at 37 °C. A signal decrease was detected in all samples incubated at 37 °C compared to measurement at 20 °C (Figure 3.21), thus the influence of higher measurement temperatures seems detrimental. Notably, the wildtype sample that exhibited the strongest initial luminescence

signal at 20 °C also displayed the most pronounced signal reduction (5-fold decrease). As a consequence, all subsequent experiments were conducted at 20 °C with an equilibration period of 30 min for samples and substrate prior to performing the luminescence assay.

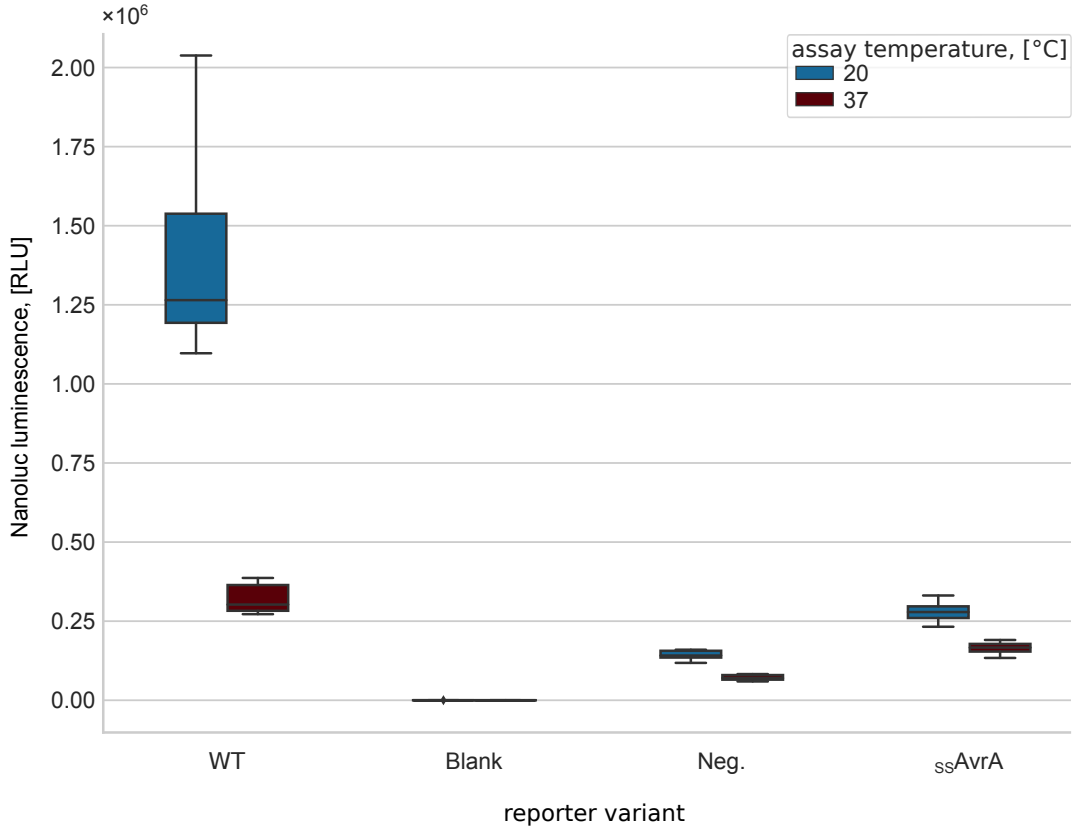


Figure 3.21.: The Nanoluc luciferase displays optimal activity at room temperature. The Nanoluc luminescence signal of 4 samples (WT = SB905 Δ*sptP* expressing $_{FL}$SptP-NLuc, SB905 Δ*sptP* expressing AvrA$_{1\text{-}25}$-SptP$_{26\text{-}535}$, uncomplemented SB905 Δ*sptP* (Blank), negative control = SB905 Δ*invA* Δ*sptP*) was quantified first at 20 °C. Following a 15 min incubation period at 37 °C, samples were remeasured at 37 °C. Prior to each measurement, the 96-well plates were shaken for 30 s at 300 rpm. Each box represents 8 biological replicates measured in triplicate, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are shown as rhombs.

3.1.3.7. Optimized parameters

After performance of all optimization tests, the conditions displayed in Figure 3.22 were applied to all subsequent secretion quantification assays.

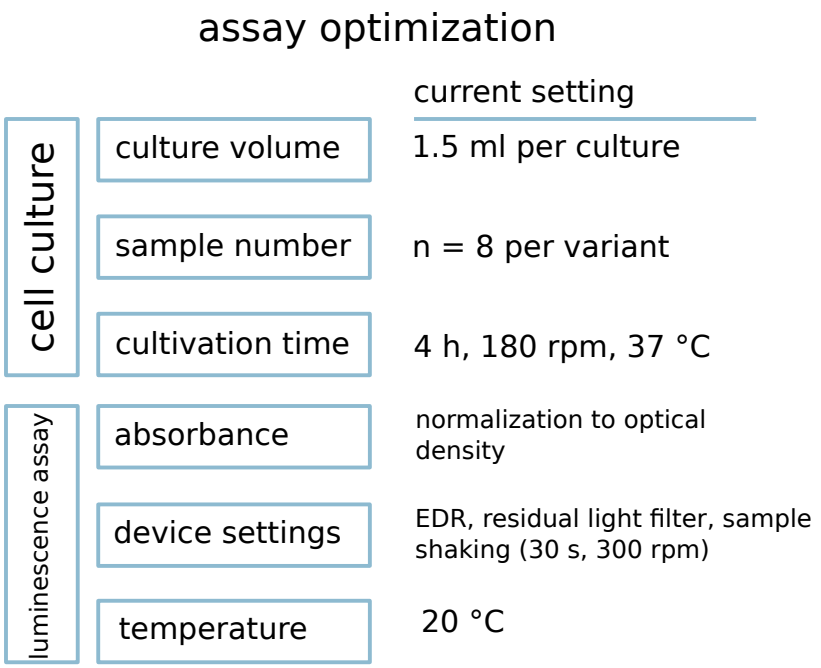


Figure 3.22.: Optimized parameters of the secretion quantification assay in the current experimental setup

---

3.1.4. Quantification of secretion signal effects

Subsequent to the optimization of the assay conditions, the amount of tested secretion signals was expanded to all experimentally verified *Salmonella* effector N-termini. In this first setup the dual-reporter strategy was used, additionally expression was enhanced using the arabinose-inducible HilA plasmid (pHilA).

3.1.4.1. Generation of plasmid variants

All N-termini of *Salmonella* T3SS effectors are shown in Table 3.2.

Table 3.2.: N-termini of experimentally verified *Salmonella* T3SEs. For 42 of the 54 verified effectors secretion quantities could be obtained. The plasmids of the remaining 12 N-termini failed to assemble despite manual modification of the generated plasmids. The majority of variants were assembled and quantified by J.Ahrendt.

Effector	Uniprot ID	N-terminus	tested
AvrA	Q8ZMI3	MIFSVQELSCGGKSMLSPTTRNMGA	✓
NleC	A0A0F6B537	MPAGIKPIFINNMMSIYGLSHPHDS	✓
GogB	Q8ZN18	MTYRLKKRMKIGFQPAILQYAYTSN	✓
GtgA	A0A0F6AZI6	MPTGIKPIFINNMMSTYGLSHPHDS	✓
GtgE	A0A0H3N9Y3	MLRHIQNSLGSVYRSNTATPQGQII	✓
OrgC	A0A0H3NF83	MIPGTIPTSYLVPTADTEATGVVSL	✓
PipA	A0A0F6AZQ0	MLPVTYRLIPQSGVSTYRLNTADTP	✓

Effector	Uniprot ID	N-terminus	tested
PipB	Q8ZQ59	MPITNASPENILRYLHAAGTGTKEA	✓
PipB2	Q8ZMM8	MERSLDSLAGMAKSAFGAGTSAAMR	✓
PrgI	P41784	MATPWSGYLDDVSAKFDTGVDNLQT	✓
PrgJ	P41785	MSIATIVPENAVIGQAVNIRSMETD	✓
SboA	A0A0K0H8V0	MNISSSRINFSTIPFQVKKLVKTIH	✓
SboC	A0A0K0HD42	MNVIKNCLSSLNNLLGISCRSYAVS	✓
SboH	A0A0K0HC32	MNISSSGINISTIPTQVKKSVETIR	✓
SboI	A0A0K0H9B7	MFDINSAHVSIRSINIPPQPSSTRS	✓
SifA	Q56061	MPITIGNGFLKSEILTNSPRNTKEA	✓
SifB	Q9KIB9	MPITIGRGFLKSEMFSQSAISQRSF	✓
SipA	P0CL52	MVTSVRTQPPVIMPGMQTEIKTQAT	×
SipB	Q56019	MVNDASSISRSGYTQNPRLAEAAFE	✓
SipC	P0CL47	MLISNVGINPAAYLNNHSVENSSQT	✓
SipD	Q56026	MLNIQNYSASPHPGIVAERPQTPSA	✓
SlrP	Q8ZQQ2	MFNITNIQSTARHQSISNEASTEVP	✓
SopA	Q8ZNR3	MKISSGAINFSTIPNQVKKLITSIR	✓
SopB	O30916	MQIQSFYHSASLKTQEAFKSLQKTL	✓
SopD	P40722	MPVTLSFGNHQNYTLNESRLAHLLS	✓
SopD2	Q8ZQC8	MPVTLSFGNRHNYEINHSRLARLMS	×
SopE	O52623	MTKITLSPQNFRIQKQETTLLKEKS	✓
SopE2	Q7CQD4	MTNITLSTQHYRIHRSDVEPVKEKT	✓
SopF	Q8ZPY9	MLKPICHSGSIKVPEYLETDKEKNA	✓
SpvC	P0A2M9	MPINRPNLNLNIPPLNIVAAYDGAE	×
VsdE/SpvD	P0A2N2	MPINRPNLNLNIPPLNIVAAYDGAE	✓
SrfJ	A0A0H3NPQ1	MKGRLISSDPYRQQFLVERAVSFSH	×
SpiC	P0CZ04	MSEEGFMLAVLKGIPLIQDIRAEGN	×
SsaG	A0A0H3NKW1	MDIAQLVDMLSHMAHQAGQAINDKM	✓
SsaL	H9L496	MSVVPVSTQSYVKSSAEPSQEQINF	✓
SseB	Q7BVH7	MSSGNILWGSQNPIVFKNSFGVSNA	✓
SseC	O84947	MNRIHSNSDSAAGVTALTHHHLSNV	×
SseD	Q9R803	MEASNVALVLPAPSLLTPSSTPSPS	×
SseF	H9L407	MKIHIPSAASNIVDGNSPPSDIQAK	✓
SseG	H9L486	MKPVSPNAQVGGQRPVNAPEESPPC	×
SseI	A0A0F6AZL3	MPFHIGSGCLPAIISNRRIYRIAWS	✓
SseJ	Q9FD10	MPLSVGQGYFTSSISSEKFNAIKES	✓
SseK1	Q9L9J3	MIPPLNRYVPALSKNELVKTVTNRD	✓
SseK2	Q8ZNP4	MARFNAAFTRIKIMFSRIRGLISCQ	✓
SseK3	P0DUJ7	MFSRVRGFLSCQNYSHTATPAITLP	✓
SseL	Q8ZNG2	MNICVNSLYRLSIPQFHSLYTEEVS	✓
SspH1	D0ZVG2	MFNIRNTQPSVSMQAIAGAAAPEAS	×
SspH2	P0CE12	MPFHIGSGCLPATISNRRIYRIAWS	×
SteA	Q8ZPD7	MPYTSVSTYARALSGNKLPHVAAGD	×

Effector	Uniprot ID	N-terminus	tested
SteB	Q8ZPA6	MPISICKHGAPFVVQHENRYGSGAS	✓
SteC	Q8ZP57	MPFTFQIGNHSCQISERYLRDIIDN	✓
SteD	Q8ZNP2	MNVTSGVNAQTPLLPPSERGNDEKP	✓
SarA	A0A0F6B506	MMRFVYIYILVIYGSYLWFSLGGNM	×
StoD	Q8Z7T2	MFLTFPNVAITRDNRIDKLSENDLE	✓

3.1.4.2. Dual reporter secretion quantification

For 42 of the 54 *Salmonella* N-termini secretion quantities could be retrieved (Figure 3.23). In comparison, we obtained up to 1000-fold larger signal intensities for the Nanoluc luminescence reporter compared to the Red Firefly reporter. Ideally, the SipA-RFLuc reporter should exhibit a reproducible, constant signal regardless of the upstream Secretion signal (SS)-SptP-NLuc variant. The wildtype controls expressing $_{\text{FL}}$SptP-NLuc and SipA-RFLuc however displayed a significant degree of variation. A relationship between the Nanoluc luminescence intensity and the Red Firefly luminescence could not be identified. As a consequence, we concluded that using the SipA-RFLuc construct as a reporter control was not a feasible approach to identify and normalize differences in T3SS secretion differences between secretion signal variants.

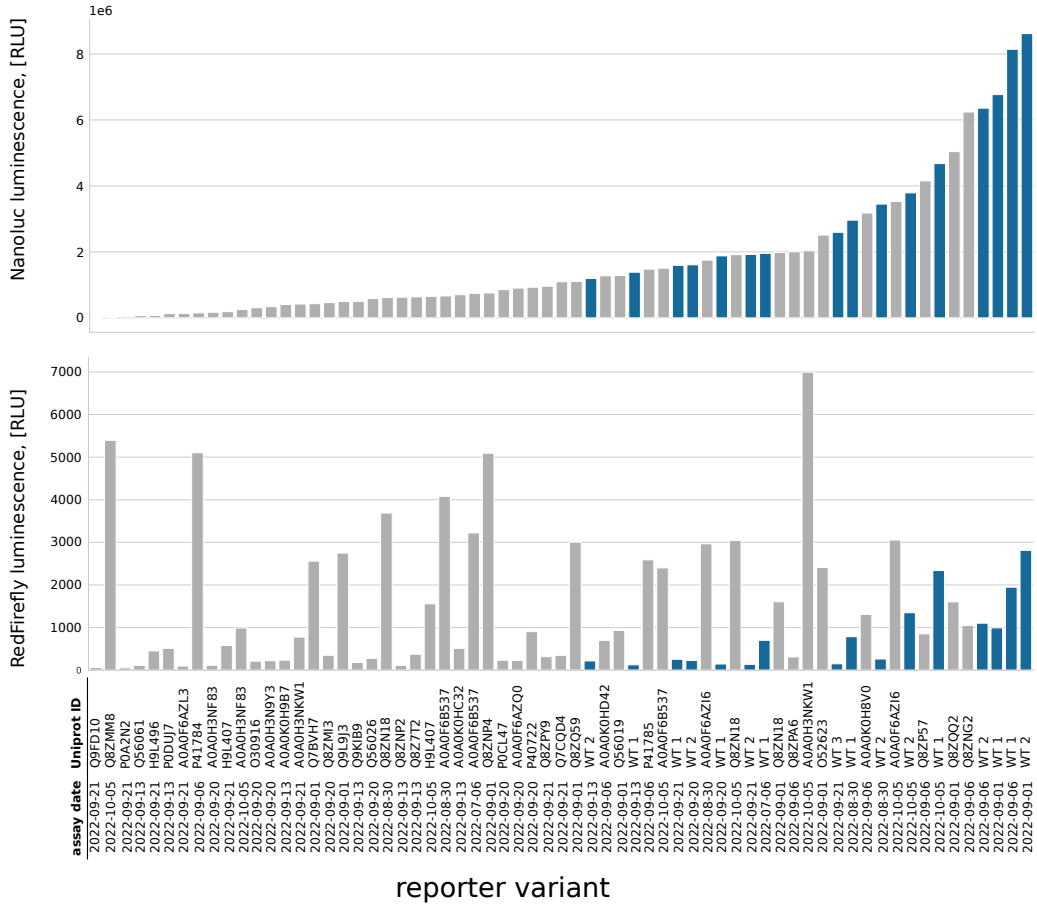


Figure 3.23.: The levels of secreted control reporter SipA-RFLuc do not correlate with increasing levels of secreted SptP-NLuc reporter and display large variations in signal intensity. Nanoluc and Red Firefly luminescence signals obtained from 8 conducted secretion quantification assays were ranked ascendingly, following the order of the Nanoluc luminescence signal. The luminescence signal from the Red Firefly luciferase is displayed below the Nanoluc luminescence signal. All luminescence signals were normalized to the optical density of the respective culture. Each assay included the quantification of two wildtype samples (*S. typhimurium* SB905 Δ*sipA* Δ*sptP*, expressing $_{\text{FL}}$Sptp-NLuc and SipA-RFLuc ●). Each bar displays the median value of the respective sample variant. The displayed data was generated in parts by J.Ahrendt as part of her master thesis.

3.1.4.3. Single reporter secretion quantification

The initial idea of assessing reporter expression by quantification of a second unaltered reporter did not provide reliable results (subsubsection 3.1.4.2). The weak signal strength of the second effector with the RFLuc-tag resulted in a large margin of error that did not allow normalization against the control effector (Figure 3.23). To enable an estimation of the amount of secretion whilst taking the total reporter expression levels into account, a new strategy was pursued. Using a new plasmid variant containing solely the SptP-NLuc

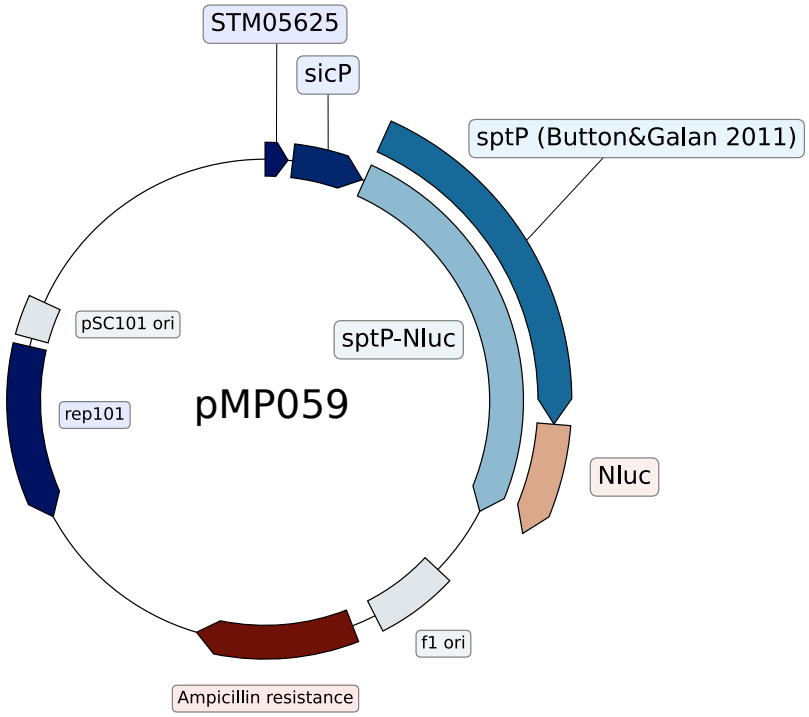


Figure 3.24.: Design of the single-reporter plasmid pMP059. Plasmid pMP059 expressing wildtype SicP and SptP-NLuc induced from the native promoter was used as a template for generating plasmid variants with secretion signals from *Salmonella* effectors. The N-terminal 25 residues of SptP-NLuc were replaced with the respective N-terminal residues of the selected T3SE.

reporter, we instead expanded the sample protocol to include preparation of cell culture samples and pellet samples (Figure 3.24). The quantification of either cell culture or pellet luminescence signal should allow an extrapolation of the total expressed reporter protein levels. A control assay was performed to assess the sensitivity and robustness of both samples types. The coefficient of variation (CV) was calculated as a measure of relative variability within the triplicate measurement of a biological replicate of a sample. Compared to the average CV of the supernatant samples, pellet and cell culture samples display elevated degrees of variation of 0.097 and 0.296, respectively (Table 3.3). To estimate the fraction of samples that exhibit higher variability, we defined a threshold CV of 0.2 as the maximum acceptable degree of variation. Of 48 samples tested in this experiment, 5 pellet samples and 26 cell culture samples exhibit a CV above 0.2, in contrast to only 2 samples of the supernatant samples.
Thus, for all subsequent assays, pellet and supernatant samples were prepared. The average

coefficient of variation over all conducted assays showed a distinct increase for pellet samples to an average CV of 0.271 and a marginal increase for supernatant samples to 0.073. Again, for all *Salmonella* N-termini new plasmid variants based on pMP059 (Figure 3.24) were assembled and quantified. In this setup, additional induction via expression of HilA was omitted to preserve native expression conditions.

Table 3.3.: Different sample types display diverging degrees of Nanoluc luminescence signal variation. A secretion quantification assay with strains *S.typhimurium* SB905 Δ*sptP* expressing $_{FL}$SptP-NLuc, SipA$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc, SopD2$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc and SopE2$_{1\text{-}25}$-SptP$_{26\text{-}535}$-NLuc and the T3SS-deficient SB905 Δ*invA* Δ*sptP* expressing $_{FL}$SptP-NLuc was performed as described in 2.2.1. 48 samples were prepared for all sample types and quantified in triplicate. For strain SB905 Δ*sptP* expressing $_{FL}$SptP-NLuc 16 samples were prepared as this strain served as a wildtype reference on two luminescence plates. For each of the other strains 8 samples were prepared.

Sample type	Average coefficient of variation	No. of samples with CV >= 0.2, [%]
supernatant	0.059	4.2
pellet	0.097	10.4
optical density	0.249	31.2
cell culture	0.296	54.2

3.1.4.4. Reproducibility

The previously conducted quantification procedure exhibited high degrees of variation for identical sample variants that could not be attributed to differences in growth. To assess the reproducibility of our assay and develop a methodology for signal comparison across assays, we monitored the signal strength of the control strain SB905 Δ*sptP* expressing wildtype $_{FL}$SptP-NLuc. Despite identical growth, culture and sample preparation conditions, the signal strength of this control varied by a significant amount from assay to assay (2.89×10^6 RLU to 1.31×10^7 RLU) with an average luminescence signal of 6.44×10^6 RLU. The negative controls of these assays (SB905 Δ*invA* Δ*sptP* expressing $_{FL}$SptP-NLuc) displayed mean residual signals ranging from 1.46×10^5 R to 1.08×10^6 RLU with an average luminescence signal of 358 263.00 (Figure 3.26). We attributed this to cell lysis.

By comparing the results of the first quantification assay using HilA-induced cultures with the current assay relying on the native induction mechanism, we attempted to examine differences and similarities regarding the quantitative effect of the N-termini. Using the Nanoluc luminescence signal of each N-terminus normalized against optical density and the respective positive control of the assay, a ranking of each secretion signal was compiled. It became evident that some secretion signal-reporter variants are secreted to different levels in natively grown cultures in contrast to HilA-induced cultures. For other secretion signal variants the relative secretion levels remain similar, overall no clear conclusion could be drawn (Figure 3.27).

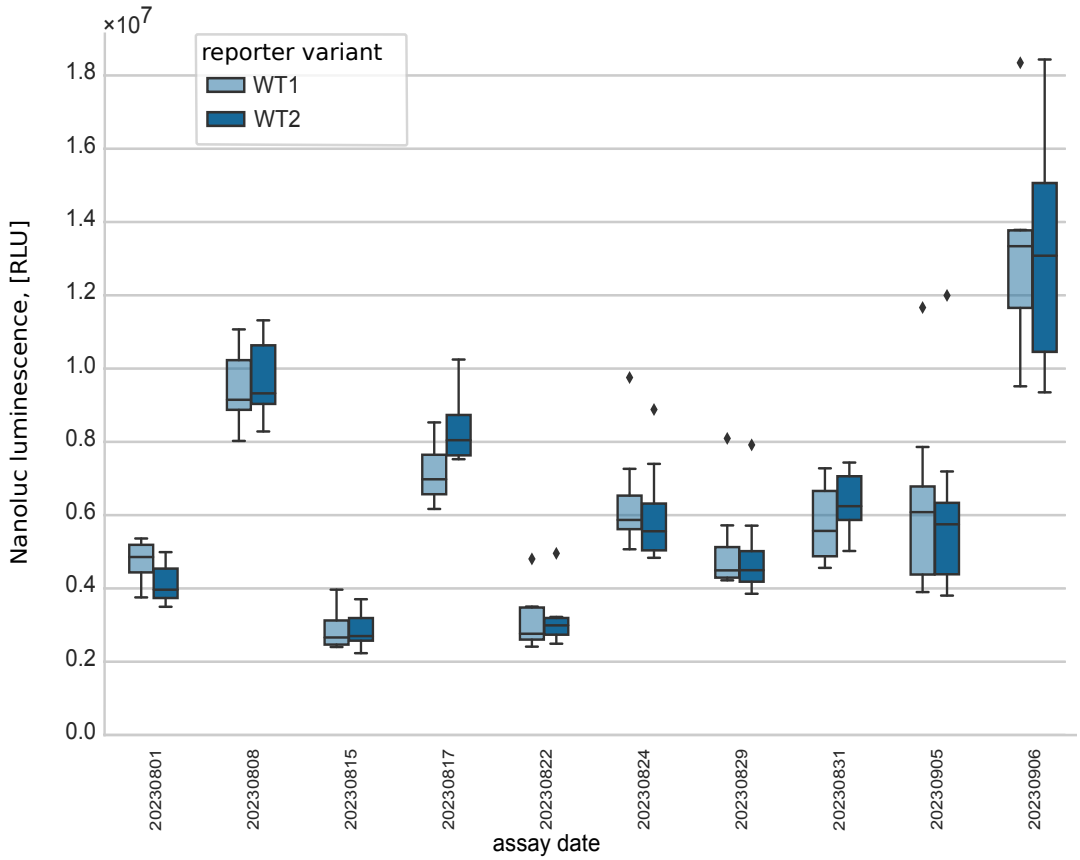


Figure 3.25.: Nanoluc luminescence signals quantified from supernatants of the wildtype (WT) control strain SB905 Δ*sptP* + pMP059 display variation in quantity across 10 assays. Nanoluc luminescence measured from supernatants of the wildtype control strain SB905 Δ*sptP* + pMP059 expressing $_{\mathrm{FL}}$SptP-NLuc were quantified over the course of 10 assays. All assays were conducted identically. Cell cultures were grown for 4 h, 180 rpm, 37 °C and separated into pellet and supernatant samples. For each assay, the control strain was measured in 2 separate LumiTrac plates for comparability. All supernatant luminescence signals are normalized to the optical density of the respective culture. Each box represents 8 identical biological replicates measured in triplicate, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are shown as rhombs.

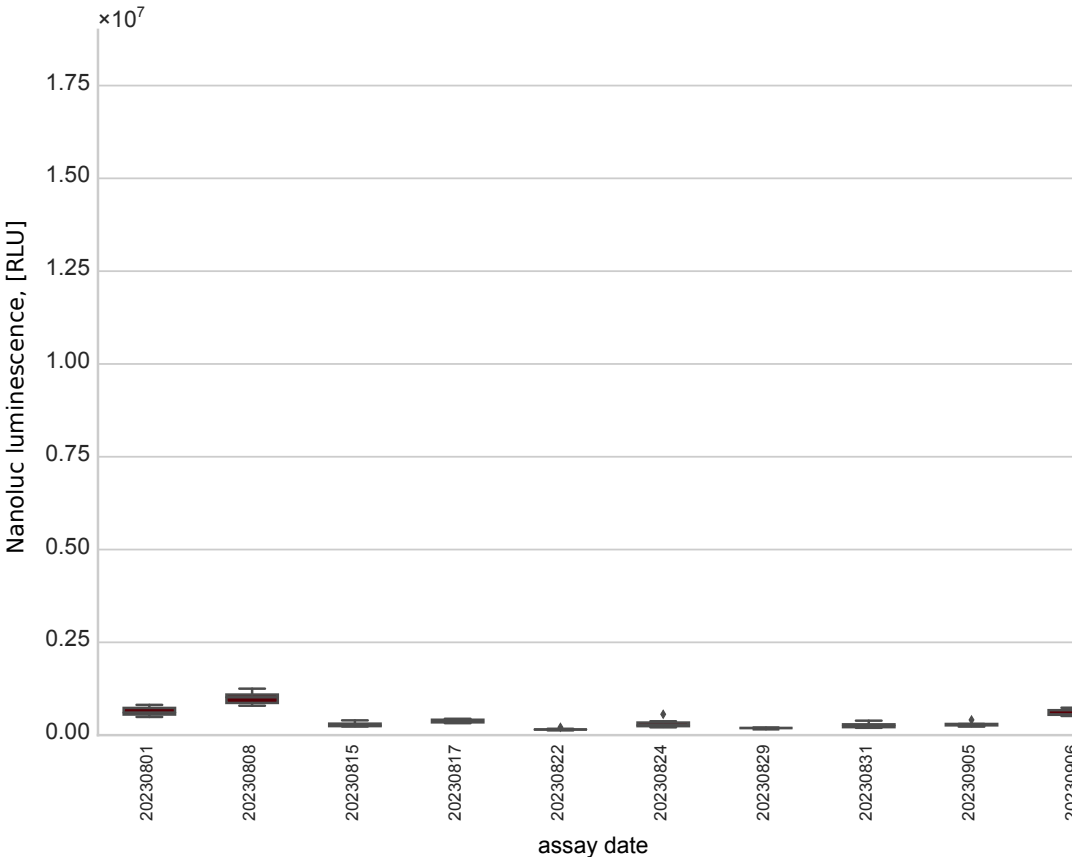


Figure 3.26.: Nanoluc luminescence signals quantified from supernatants of the T3SS-deficient negative control strain SB905 Δ*invA* Δ*sptP* + pMP059 exhibit consistent signal intensities. Nanoluc luminescence measured from supernatants of the T3SS-deficient negative control strain SB905 Δ*invA* Δ*sptP* + pMP059 expressing $_{\mathrm{FL}}$SptP-NLuc were compared across 10 assays. All assays were performed identically. Cell cultures were grown for 4 h, 180 rpm, 37 °C and subsequently separated into pellet and supernatant samples. All supernatant luminescence signals are normalized to the optical density of the respective culture. Each box represents 8 identical biological replicates measured in triplicate, the box displays the quartiles of the data distribution, whiskers denote the rest of the distribution, outliers are shown as rhombs.

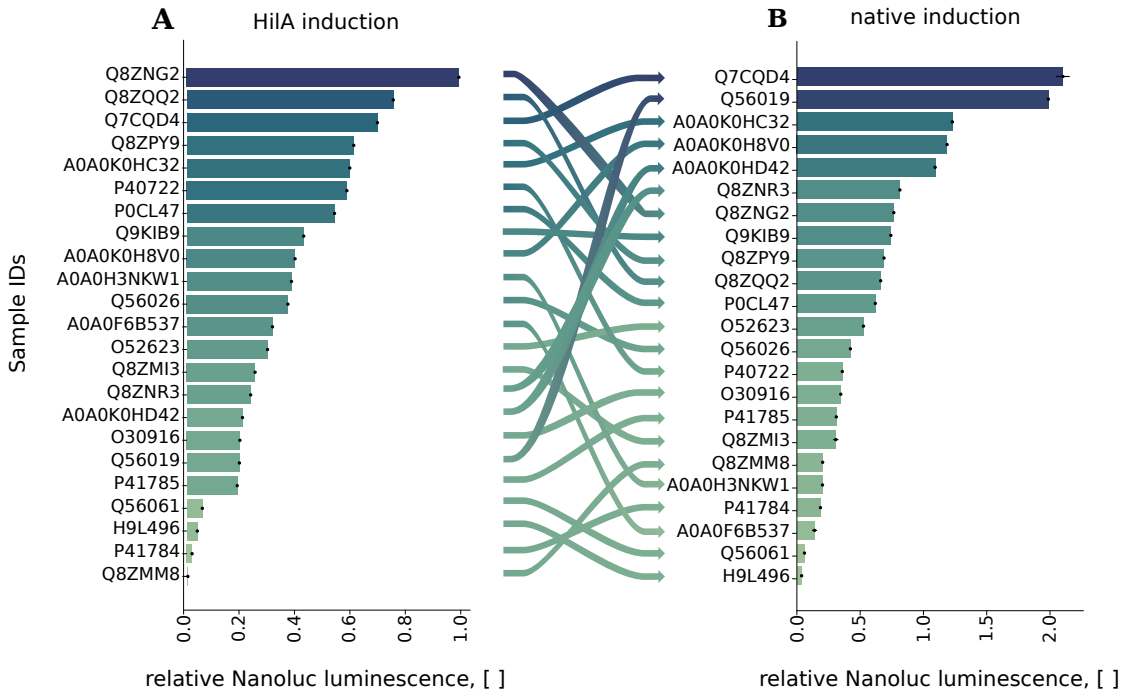


Figure 3.27.: The relative secretion efficiency of secretion signal-reporter variants is altered by employing different induction strategies. A comparison of the relative secretion quantities of secretion signal-reporter variants incubated either under HilA-induced (**A**) or native induction (**B**) conditions reveals differences in the ranking of low to high secreted reporters. **A** Nanoluc luminescence quantified in supernatants of HilA-induced cultures expressing different reporter variants visualized as a ratio against the luminescence signal of $_{FL}$SptP-NLuc, ranked from low to high. **B** Nanoluc luminescence quantified in supernatants of naturally induced cultures expressing different reporter variants visualized as a ratio against the luminescence signal of $_{FL}$SptP-NLuc. Each bar displays the median value of the respective sample.

3.1.4.5. Calculation of secretion efficiency scores

To tackle the persistent variation of the reporter luminescence signal between assays, we calculated the secretion efficiency for all controls and samples by building a ratio between secreted and total quantified protein (Equation 3.2). To achieve comparability between assays, we calculated a secretion score for each sample using the wildtype $_{\mathrm{FL}}$SptP-NLuc reporter secretion efficiency as a reference for the average secretion efficiency per assay (Equation 3.3).

$$\text{secretion efficiency} = \frac{\text{supernatant luminescence signal}}{\text{supernatant luminescence signal} + \text{pellet luminescence signal}} \tag{3.2}$$

$$\text{secretion score} = \frac{\text{sample secretion efficiency}}{\text{wildtype secretion efficiency}} \tag{3.3}$$

Secretion efficiency scores were obtained for 25 of 44 tested *Salmonella* N-termini (Figure 3.28). Secretion signals from the remaining tested N-termini were not included in the current summary, as they displayed secretion signal quantities below the negative control signal of the respective assay.

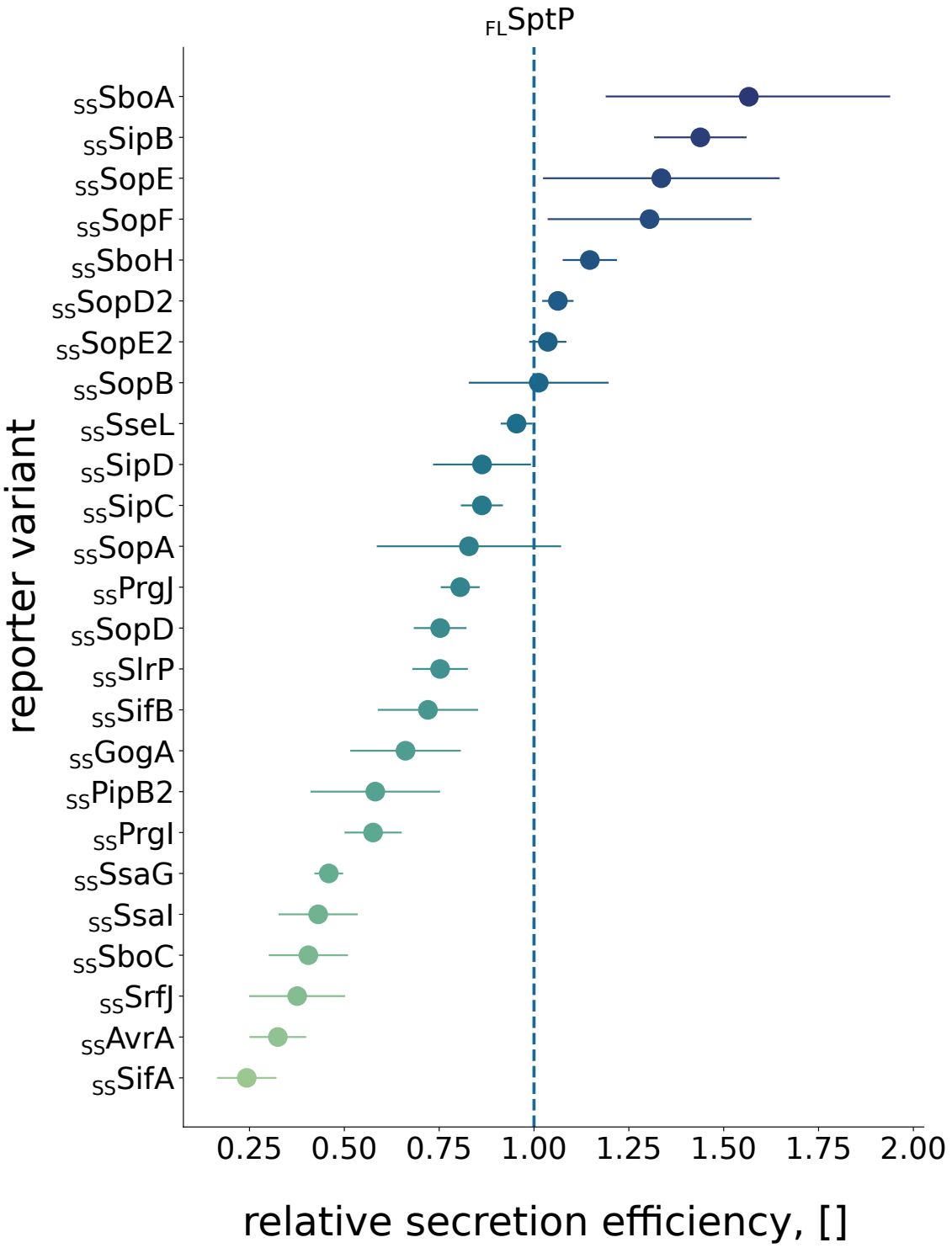


Figure 3.28.: *Salmonella* effector N-termini promote different secretion efficiencies. Secretion efficiency scores for N-termini from 25 *Salmonella* T3SS-secreted proteins fused to $SptP_{26\text{-}535}$-NLuc were calculated in relation to the native SptP reporter secretion efficiency on the respective assay day. The quantified Nanoluc-luminescence was normalized against cell growth and total reporter expression to obtain the percentage of secreted reporter protein (see Equation 3.2). Subsequently, the obtained percentage was set in relation to the wildtype reporter secretion efficiency (see Equation 3.3). The relative secretion efficiency of the $_{FL}$SptP-NLuc reporter is marked as a dashed, green line. Each datapoint displays the median value with MAD.

3.2. T3SS effector database

The quantification of secretion signal effects as well as computational prediction of T3SEs necessitates the availability of a large number of non-redundant N-terminal secretion signal sequences. For computational approaches attempting to classify protein sequences, high data quality is of utmost importance to achieve satisfactory performance [144]. This effect is more pronounced for small datasets and promotes a data-centric approach with focus on data quality [145]. Numerous efforts to identify and analyze T3SEs have led to the availability of a growing number of publically available datasets [101], [146], [147], [148], [149]. With the expanding repertoire of experimentally identified T3SEs in recent years, these datasets encompass non-redundant sequences ranging from roughly 250 to more than a thousand sequences. Common approaches to assemble such a dataset utilize a combination of automated/semi-automated sequence selection tools and subsequent curation procedures based on available literature. In this work, both a common workflow with semi-automated sequence picking as well as a bottom-up approach starting from an extensive literature review were tested and applied. The following section describes these two strategies and the resulting T3SE database.

3.2.1. Semi-automated T3SE dataset assembly

To retrieve T3SS effector sequences, an extended search and processing routine was established using the database of the *National Center for Biotechnology Information*. A text-based search on the annotated sequence data was conducted using the *Identical protein groups database*. This database contains a single non-redundant entry for each unique protein translation from various sources such as Genbank, RefSeq but also Swiss-Prot and PDB [150]. Eligible text strings used for the search are listed in table A.2A. As a result, 162237 unique sequence entries were obtained. An inspection of all words in the annotation data revealed a set of annotation tags indicating incomplete sequence entries (appendix Table A.2B). Sequences containing any of the identified tags (see Table A.2B) were removed. All remaining sequences with annotations that cast doubt on their origin as effector proteins were manually reviewed for further dataset refinement. These included proteins identified as chaperones, pseudogenes or flagellar proteins. The remaining entries were merged with the updated T3SE dataset (2015) that was used for the EffectiveT3 classification model [151]. As a final step, the CD-Hit software was utilized to perform a homology clustering with a sequence identity cut-off at 60 % [152]. Each identified cluster contains entries with a sequence identity of 60 % to each other or higher. To reduce sequence redundance, a representative sequence was manually selected from each cluster. A final dataset of 244 T3SEs was obtained, this dataset will be denoted as dataset 1 in the upcoming sections.

3.2.2. T3SE database assembly

The assembly of a T3SE dataset using a semi-automated approach is based on the annotation status of the retrieved sequences identifying them as T3SEs. Thus, the assembly of a high-quality dataset is highly reliant on actuality and completeness of the underlying databases. The review process of some of the publically available datasets revealed reoccurring problems such as lack of annotation or false positive entries derived from inapt sources. For example, our analysis of multiple published T3SE datasets revealed protein sequences obtained from crystallization studies that represent N-terminally truncated proteins. Furthermore, tracing the origin of sequences remains laborious as sufficient annotation of publication sources is often missing. The frequently outdated annotation status of sequence entries furthermore makes automated selection strategies prone to the incorrect removal of verified effector proteins undermining the completeness of the dataset. To circumvent the aforementioned problems, a bottom-up approach was pursued starting from an extensive literature review. In the 2017 survey conducted by Hu et al. on T3SS-harboring bacterial species over 30 different bacteria were identified to potentially contain T3SEs [34]. A species-specific effector search for each of the referenced species was conducted using GoogleScholar. All accessible publications were manually reviewed and examined for experiments demonstrating T3SS-mediated secretion and the original protein sequence. The retrieval of effector sequences followed a multistep procedure. All protein sequences were derived directly from referenced nucleotide or protein identifiers as specified in the reviewed publication. Initially, the https://www.uniprot.org/ was searched for the specified sequence identifier. In case no matching sequence was available, the NCBI database was searched. If this search also failed to return a sequence entry, the genomic sequence of the bacterial strain was downloaded on which the experimental validation was performed. Matching gene identifiers of the genomic annotation were used to retrieve a suitable protein sequence. For two effector entries, the Ralstonia Genomic database was employed as no other source was available (Table 3.4A). In its current state (2024-04), the dataset encompasses 735 entries that have been confirmed for T3SS-mediated secretion. As of now, a total of 2192 publications were reviewed.

Table 3.4.: Current status of effector sequences in our T3SE database. **A** Database sources for effector sequences employed in our database with the number of sequences derived from every source. **B** Annotation status of T3SS effector proteins sorted by bacterial species. Partly annotated effectors have been experimentally confirmed as T3SE yet lack definite annotation in supplementary information categories.

A Frequency of database usage for entries in our database

database	**abbreviation**	**frequency**
TrEMBL	tr	539
Swiss-Prot	sp	166
Genbank	gb	2
National library of medicine	ncbi	26
Ralsto T3E	rdb	2

B Annotation status of T3SS-secreted proteins per species.

bacterial species	**annotation status**
Acidovorax	annotated
Aeromonas	annotated
Bordetella	annotated
Burkholderia	partly annotated
Chlamydia	annotated
Chromobacterium	annotated
Citrobacter	annotated
Edwardsiella	annotated
Escherichia	annotated
Erwinia	annotated
Pantoea	annotated
Pectobacterium	annotated
Photorhabdus	annotated
Pseudomonas	partly annotated
Salmonella	annotated
Shigella	annotated
Sodalis	annotated
Ralstonia	not annotated
Rhizobium	partly annotated
Vibrio	partly annotated
Xanthomonas	partly annotated
Yersinia	annotated

3.2.2.1. Database classes

Supplementary information retreived during the literature review process was categorized into the four classes protein function, host cell targets, cellular localization and host organisms. Therefore, each completed entry in this dataset contains 6 annotation categories as displayed in Figure 3.29. Currently, 494 of 735 entries have been completely annotated, an additional 241 have been confirmed for T3SS-mediated secretion, yet lack complete annotation. For each category, a set of defining rules were established to be met.

>sp|P0CL52|SIPA_SALTY Cell invasion protein SipA OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) OX=99287 GN=sipA PE=1 SV=1

function = bundles F-actin, induces membrane ruffling, recruits syntaxin-8 to the SCVs, promotes fusion of SCVs with early endosomes
target = F-actin, Syn7, Syn8, Syn13, SNAP23, PERP
evidence = T3SS secretion assay, T3SS translocation assay
localization = host cell plasma membrane, host cell cytoskeleton
host = human
PMID = 30309979, 11331579, 8522512, 18005682, 10092234, 21902796, 10487745, 16107539, 22636784, 16869830, 20947770, 17635190, 28630067, 14512630,25486861, 15522075, 30532744

MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSGEIKGPQLEDFPALIKQASLD
ALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSALNATSDSPEAKTLLMKVGAEYTAQI
IKDGLKEKSAFGPWLPETKKAEAKLENLEKQLLDIIKNNTGGELSKLSTNLVMQEVMPYIASCI
EHNFGCTLDPLTRSNLTHLVDKAAAKAVEALDMCHQKLTQEQGTSVGREARHLEMQTLIPLLL
RNVFAQIPADKLPDPKIPEPAAGPVPDGGKKAEPTGINININIDSSNHSVDNSKHINNSRSHVD
NSQRHIDNSNHDNSRKTIDNSRTFIDNSQRNGESHHSTNSSNVSHSHSRVDSTTHQTETAHS
ASTGAIDHGIAGKIDVTAHATAEAVTNASSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIG
KPVQATVHGVDDNKQQSQTAEIVNVKPLASQLAGVENVKTDTLQSDTTVITGNKAGTTDND
NSQTDKTGPFSGLKFKQNSFLSTVPSVTNMHSMHFDARETFLGVIRKALEPDTSTPFPVRRA
FDGLRAEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQFAREAG
LTRLKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHMQR

Figure 3.29.: Exemplary fasta entry of the T3SE-database for *Salmonella* effector SipA. In addition to the original information from the fasta-file, supplemental information about function, target proteins, experimental proof for T3SS-mediated secretion, cellular localization, host organisms and literature references (Pubmed ID format) is included.

3.2.2.1.1. Evidence for T3SS-mediated secretion The range of experiments applied in T3SE characterization studies vastly exceeds the methods defined as experimental evidence for T3SS-secreted proteins. For example, it has been shown that many proteins of plant pathogens elicit hypersensitive responses (HRs) in susceptible host plants, a property indicative of T3SS effector proteins. *In-vitro* cytotoxicity assays like Yeast-to-Hybrid (Y2H) or infection assays also strengthen the presumption of certain proteins to be T3SS effectors. Finally, bioinformatic approaches identifying proteins homologous to known effectors or structural methodologies like nuclear magnetic resonance (NMR) spectroscopy, electron microscopy or crystallographic studies expand the range of potential proteins as effectors. However, none of these methods can provide a definite conclusion as to whether a protein

is secreted via the T3SS.

The key criterion for any entry in this database constitutes the experimental proof for T3SS-mediated secretion or translocation of the protein. Analogous to the Uniprot database which defines 5 types of evidence for the existence of a protein, the level of evidence of a candidate effector for T3SS-mediated secretion was classified based on the conducted experiments. The most simple type of sufficient experimental evidence is displayed by the T3SS secretion assay. Typically, a candidate effector was visualized in the supernatant of a cell culture using specific antibodies targeting the candidate protein or an enclosed tag via western blotting and compared to a T3SS-deficient strain. Experiments that lacked suitable negative controls, such as the expression in a T3SS-deficient strain, were not considered. In strains like *Chlamydiae* that are intractable to genetic modification, the proteins of interest can not be expressed in the native strain. Here, the application of a heterologous T3SS secretion assay in strains such as *Yersinia* or *Shigella* is an accepted strategy for validation. To ascertain the exclusive selection of T3SEs and secreted structural components of the nfT3SS, proteins that have been demonstrated to be secreted via the flagellar T3SS were not included in the database. This comprises several putative effectors from *Vibrio* and *Pectobacterium* [153],[104]. Furthermore, *Campylobacter* strains only possesses a flagellar T3SS, therefore 42 proteins verified for flagellar T3SS export remain categorized as putative effector proteins [154]. Considering that both systems are ancestrally related, an expansion to testing these candidates might be of interest.

A more sophisticated approach to demonstrate a proteins nature as a T3SE are translocation assays. As previously specified (subsubsection 3.1.2.2), Adenylate cyclase-based assays or TEM-1 β-Lactamase (BLA) assays that require the eukaryotic cell environment can be regarded as clear evidence for T3SS-mediated translocation. Seldomly, a unique type of assay was employed as in the case of the *Erwinia amylovora* effector DspA that induces electrolyte leakage in plants. Here, the detection of electrolyte leakage in tobacco cells infected with an *E.amylovora* wildtype strain in comparison to a T3SS-deficient mutant strain was used to prove T3SS secretion [155]. Finally, mass proteomic approaches comparing secretomes of secreting strains with T3SS-deficient strains also provide a robust classification about whether a protein can be considered a *bona fide* effector protein or not. Amongst others, this approach has been widely applied to categorize putative effector proteins from *Aeromonas*, *Citrobacter*, *Erwinia*, *Pseudomonas*, *Ralstonia*, *Escherichia* or *Shigella* [156],[157],[158],[159], [160], [161], [162], [163].

The number of approved experimental procedures applied in the current database are summarized in Table 3.5. Predicted effector proteins based on homology or experimental characterization hinting towards a role as an effector protein were excluded.

3.2.2.1.2. Effector protein functions The discovered T3SS effector proteins thus far have been characterized to vastly different extent. The attributed functions and *in vivo* roles

Table 3.5.: Experimental validation methods qualifying T3SS effector proteins for inclusion into the T3SE database, sorted by frequency.. Experimental evidence for T3SS-mediated secretion of all 735 entries of the database were summarized into the three main validation methods and sorted by frequency. Multiple effectors were validated for T3SS secretion using different experimental approaches.

experimental evidence	frequency
T3SS secretion assay	362
T3SS translocation assay	347
T3SS secretome proteomics	158

of validated effectors collected throughout the literature review were summarized in this category to obtain an overview of enzymatic activities, effects on host cells, pathways and more. Due to the different extent of functional annotation, this information class represents the broadest category in the dataset.

For the majority of T3SEs information is sparse, hence for 298 proteins no function or activity was determined. Whilst most of the collected information comprises descriptions of phenotypic alterations of host cell morphology, involvement in pathway manipulation or targeted host cell proteins, experimentally confirmed enzymatic activities have been identified in 92 effector proteins (Table 3.6). Interestingly, *Yersinia* effector YopJ is the only effector with multiple confirmed enzymatic activities (cysteine protease activity, acetyltransferase activity). The most investigated enzymatic class are E3 ubiquitin ligases with 27 confirmed effectors. Notably, the majority of these effectors belong to *Escherichia* and *Shigella* effectors with 12 and 9 entries, respectively. Other considerable protein classes encompass proteases and transferases. Expect for the few, well characterized proteins for which an enzymatic activity has experimentally been elucidated, most effectors lack a precise annotation. Therefore, a distinct categorization for these proteins is challenging. The decision to include secreted components of the T3SS lead to 79 entries comprising either structural or regulatory components of the T3SS (Table 3.6B). The secretion of these early substrates, while often highly regulated via several control mechanisms, also requires an N-terminal secretion signal.
The range of annotated function descriptions associated with the effectors in this database exceeds the scope of this thesis, a detailed overview is available in appendix C Table A.3 .

3.2.2.1.3. Target proteins Supplementary to the general function description of the annotated effectors, specific information about interaction partners of T3SEs was also included. To retain the most precise class categorization, only interactions were included for which a direct binding or enzymatic activity was experimentally determined using either *in vitro* or *in vivo* studies. The decision to also include *in vitro* reactions was made in order to achieve a general understanding of a proteins reactivity. In the future, this knowledge might help to infer potential *in vivo* host targets.

Overall, 905 unique interaction partners were identified. These include proteins, lipids,

Table 3.6.: Functional characterization of validated T3SS-secreted proteins. **A** Enzymatic activities of T3SS-secreted proteins for which an enzymatic activity has been determined, ranked by frequency of occurrence. **B** Identified T3SS-secreted components of the type 3 secretion system.

A

enzymatic class	frequency
E3 ubiquitin transferase	27
cysteine protease	12
ADP-ribosyltransferase	9
glycosyltransferase	7
zinc metalloprotease	7
acetyltransferase	6
phosphothreonine lyase	6
ADP-riboxanase	4
serine/threonine kinase	4
methyltransferase	3
tyrosine phosphatase	2
fatty acyltransferase	1
serine hydroxymethyltransferase	1
serine protease	1
SUMO-protease	1
oxidoreductase	1
phosphoglucomutase	1

B

component	frequency
T3SS translocon protein	37
T3SS needle component	27
T3SS inner rod protein	5
regulator of T3SS secretion	5

sugars or cofactors but also nucleotides as is the case for TAL effectors or the *Shigella* effector VapC [164]. Consistent with the abundance of E3 ubiquitin ligases in the dataset, components of the ubiquitin signalling pathway belong to the largest group of interaction partners, with 19 different interactions partners being targeted 54 times by 20 different T3SEs. Furthermore, small GTPases like CDC42, Rac1, RhoA and numerous Rab proteins are frequent interaction partners for effector proteins with 90 confirmed interactions. Commonly targeted functional clusters include proteins associated with inflammation or apoptosis, vesicle trafficking pathways or cytoskeletal reorganisation. A thorough investigation into the functional clusters targeted by effector proteins would certainly reveal intriguing insights into strategies adopted by different T3SS-harboring pathogens. Due to the unfinished status of the current dataset, this analysis has not yet been performed. Interestingly, not only proteins are targeted. Some effectors also manipulate the hosts on a transcriptional level or alter the cell membrane composition by targeting specific lipids (Figure 3.30). As the annotation of several effector proteins from bacteria like *Pseudomonas* or *Xanthomonas* is unfinished, it is likely that the number of effectors involved in transcriptional manipulation of plant genes will increase in the fully annotated database (Figure 3.30A).

Another group of targets comprises cell membrane lipids that are bound and modified by several effector proteins (Figure 3.30B). The *Shigella* inositol-phosphatase IpgD for example, dephosphorylates PI(4,5) P_2 into PI5P, presumably resulting in the activation of PI3K and subsequently regulation of Rac GEF activity (Figure 3.30C). Considering that similar activities have been identified for effector proteins from *Aeromonas* (Ati2) and *Salmonella* (SopB, SopF), a directed investigation could reveal how widespread the specific modulation of membrane lipids as a strategy for host cell manipulation is [165][166]. Finally, target molecules like cholesterol, bile acids (doexycholate) and sugars (glucose, N-acetylglucosamine) have been reported for isolated T3SE [167][168][169][170][171].

A T3SS effectors targeting specific DNA sequences in the host cell nucleus

effector	target nucleotide sequence	target gene
Q07061_9XANT	TATAATTAATAATCCACTT	Bs4C
AVRB3_XANEU	TATATAAACCTNNCCCTCT	UPA20
Q47867_PANAY	ACACCaAA	HsvgT
B7UJQ8_ECO27	(TC)(TC)GCCAG(ACT)	FlgL, FlgK, FlgB, FlgC, FliE, FliG, FliM, FliN

B Lipids targeted by T3SE

targeted lipids
PI(3)P
PI(4)P
PI(3,5) P_2
PI(4,5) P_2
PI(3,4,5) P_3

C

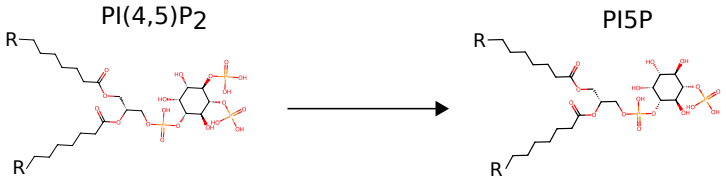


Figure 3.30.: Non-proteinaceous interaction partners targeted by T3SS-secreted proteins. A T3SS-secreted proteins binding to specific nucleotide motifs. Notably, the *E. coli* protein BolA (B7UJQ8_ECO27) while involved in biofilm formation, also represses expression of its own flagellar genes [159][172] **B** Cell membrane lipids targeted by T3SS effectors from *Bordetella*, *Salmonella* and *Shigella* specifically modifing or binding to lipids to alter the cell membrane composition or recruit target proteins. **C** The *Shigella* effector IpgD dephosporylates PI(4,5) P_2 into PI(5)P, modifying the host cell morphology [173][174][166].

3.2.2.1.4. Localizations A proteins site of activity often allows to draw inferences regarding likely interaction partners or their enzymatic activity. Experimentally determined localization patterns were identified for 315 of 735 T3SEs in the current database. All effector localizations were sorted into categories from broad to specific, thus a single effector can contain multiple localization tags. 55 effectors were broadly categorized in the cytoplasm, an additional 44 exclusively in the host cell cytosol. Specifically targeted host cell organelles are shown in Figure 3.31. Effector proteins associated with the bacte-

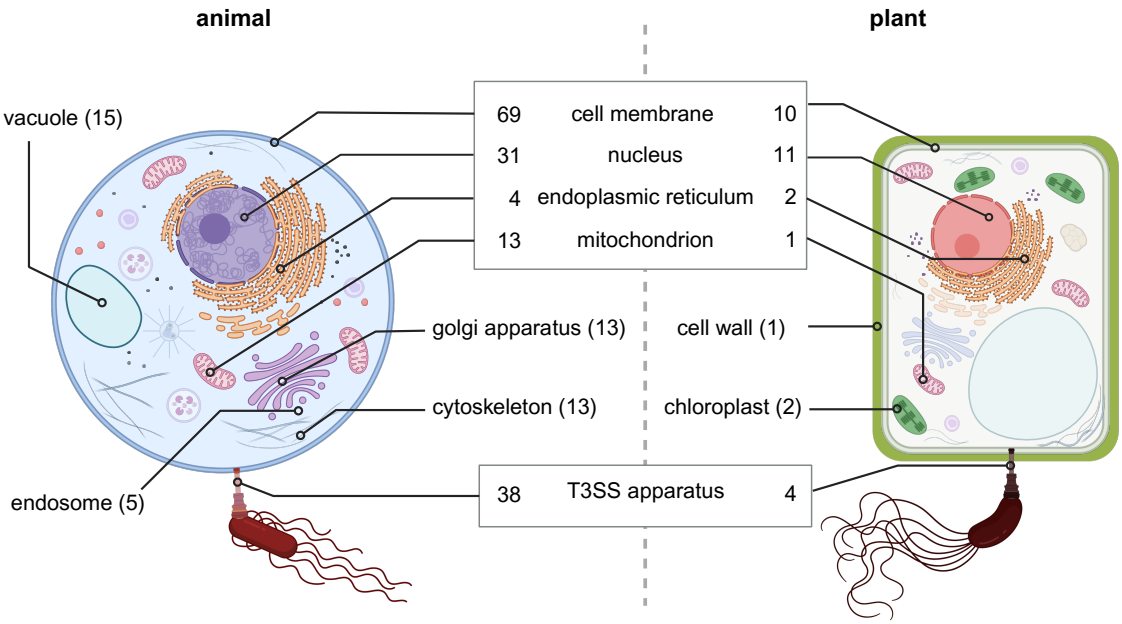


Figure 3.31.: Experimentally validated localizations of T3SS-secreted proteins identified in our database. The majority of proteins for which a cellular localization was identified, localize to the host cell membrane. Created with BioRender.com

---

rial T3SS mostly comprise structural or translocon components that are secreted during assembly. Species-specific localization patterns are present for effectors of *Salmonella*, *Shigella* or *Edwardsiella*. These pathogens invade their hosts and partly spend steps of their reproductive life cycle in a specialized phagosome-like vacuole [175]. For *Salmonella* 13 effector proteins were directly localized to the *Salmonella*-containing vacuole, whereas for *Shigella* and *Edwardsiella* a single effector is localized to the vacuole, respectively. The obligate intracellular bacterium *Chlamydia* resides in a *Chlamydia*-containing vacuole (also termed inclusion body) and the majority of its effectors are localized to the inclusion membrane (29), the inclusion lumen (16) or in close vicinity (6).

Interestingly, some effector proteins were localized exclusively to the extracellular space and thus were only shown to be secreted rather than translocated. These proteins are mainly plant effectors that are thought to act as pectate lyases on the plant cell wall.

3.2.2.1.5. Host organisms The host organism class comprises both native host organisms as well as lab strains that were utilized in the investigation of T3SE functions. Currently, animal hosts were identified in 452 instances; plant hosts in 116 instances. A detailed summary is depicted in Table 3.7

Table 3.7.: Identified host species sorted by animal and plant hosts. A Animal hosts of T3SEs sorted by descending frequency. **B** Plant hosts of T3SEs sorted by descending frequency.

A

species	frequency
Homo sapiens (human)	349
Mus musculus (house mouse)	33
Oncorhynchus mykiss (rainbow trout)	18
Salmonidae spp. (salmon)	18
Ictalurus punctatus (channel catfish)	9
Anguilla japonica (japanese eel)	6
Paralichthys olivaeus (olive flounder)	6
Cavia porcellus (guinea pig)	4
Danio rerio (zebrafish)	3
Glossina spp. (tsetse)	2
Cyprinus carpio (eurasian carp)	1
Locusta migratoria (migratory locust)	1
Spodoptera littoralis (cotton leafworm)	1
Trichopodus trichopteris (blue gourami)	1

B

species	frequency
Arabidopsis thaliana (thale cress)	28
Malus spp. (apple)	12
Capsicum annum (pepper)	11
Glycine max (soybean)	10
Beta vulgaris (beet)	10
Aeschynomene indica (curly indigo)	10
Pyrus spp. (pear)	9
Nicothiana benthamiana (benthi)	7
Solanum lycopersicum (tomato)	5
Macroptilium atropurpureum (purple bush bean)	4
Gypsophila paniculata pv. *perfecta* (common gypsophila)	2
Solanum lycopersicoides (wild tomato)	1
Vigna mungo (black gram)	1
Vigna unguiculata (cowpea)	1
Vigna radiata (mung bean)	1
Zea mays pc. *saccharata* (sweet corn)	1
Lactuca sattiva (lettuce)	1
Lablab purpureus (hyacinth bean)	1
Oryza spp. (rice)	1

3.2.2.1.6. Sequence composition Previous attempts to uncover the nature of the secretion signal have failed to elucidate a consensus pattern or motif based on the qualitative composition of the N-terminal effector sequences. The distinct enrichment of small, polar residues however, as identified by Arnold et al., and tendencies like the relative depletion of acidic and alkaline residues indicate the existence of hidden features defining the secretion signal [101]. To assess if these biases also exist in the sequences of our assembled database, we compared the composition of the N-terminal 25 residues of all 735 entries with their full-length peptide sequence. Indeed, we also identified several of the previously described propensities. In our database, residues like serine, threonine and proline are enriched compared to the full-length sequence, whereas acidic (aspartic acid, glutamic acid) and alkaline residues (arginine, lysine, leucine) are relatively depleted (Figure 3.32). This is in agreement with previous studies [101].

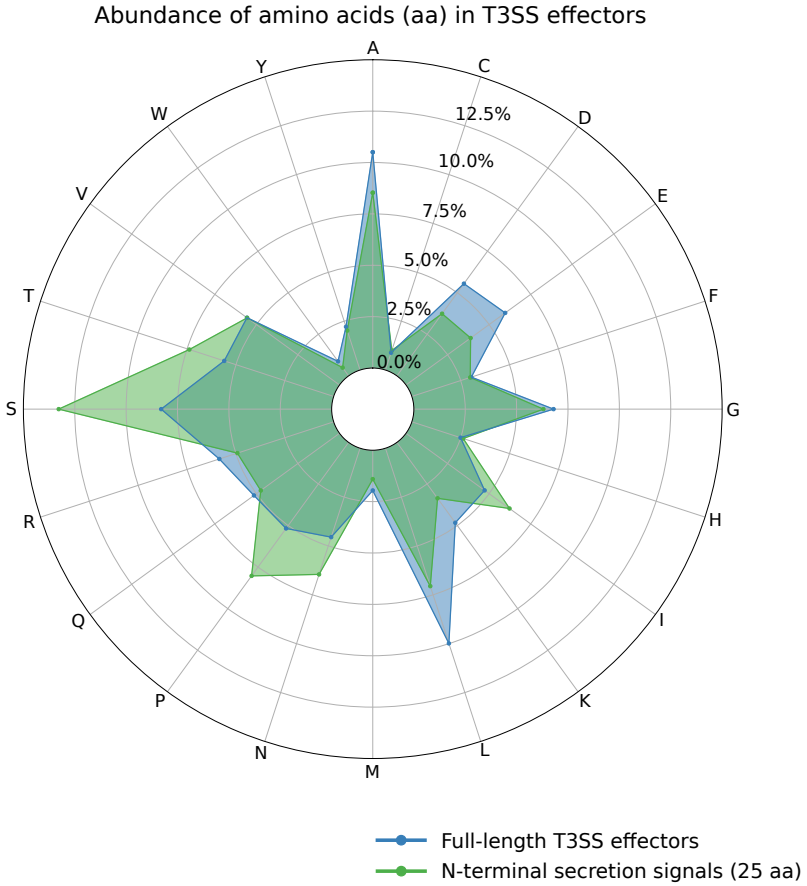


Figure 3.32.: Small, polar amino acids are enriched in the N-terminus of T3SS-secreted proteins. The relative abundance of amino acids was calculated in the N-terminal 25 residues of all entries and compared to the relative abundance of amino acids in the full length protein sequence. In agreement with previous studies, small polar residues are enriched in the N-terminus [101]. Amino acids are depicted in one-letter code, the N-terminal start codon (99 % methionine) was not included in the calculation.

3.2.3. Manually curated T3SE dataset assembly

The assembled T3SE database (subsection 3.2.2) was built with the notion of including every experimentally validated T3SS-secreted protein. Thus, it also comprises homologous protein sequences that may distort the predictional capabilities of any subsequent prediction model. For example, several *Shigella* IpaH effectors share a remarkable sequence identity within their first 10 N-terminal residues (Figure 3.33). These sequence redundancies

```
ipah4.5       ---MKPINNHSFFRSLCGLSCISRLSVEEQCTRDYHRIWDDWAREGTTTENRIQAVRLLK      57
ipah2202      ---MLPVNNPPLST---GNVSFYRTTSIDNVHNNYLSEWVEWTKNSISGENRETAFTRLQ      54
ipah7.8       ---MFSVNNTHSSV---SCSPS--INSNSTSNEHYLRILTEWEKNSSPGEERGIAFNRLS      52
ipah1880      ---MLPTNNNHRLI---SNSFSTYSIDTSRAYENYLTHWTEWKNNRIQEEQRDIAFQRLV      54
ipah3         MSIMLPINNNFSLS---QNSFYN---TISGTYADYFSAWDKWEKQALPGENRNEAVSLLK      54
ipah9.8       ---MLPINNNFSLP---QNSFYN---TISGTYADYFSAWDKWEKQALPGEERDEAVSRLK      51
                 *   **                   .    .*     .* .:    *:*  *.  *
```

Figure 3.33.: *Shigella* T3SS effector of the IpaH class display strong homology within their N-terminal 10 residues. The The first 50 residues of selected *Shigella* IpaH effectors were aligned using the ClustalOmega webserver.

---

introduce a bias that potentially skews the prediction outcome of a model trained on this data. To make our database suitable for computational prediction, the dataset preprocessing pipeline established for dataset 1 was also applied for the creation of the manually curated dataset 2(subsection 3.2.1). CD-Hit with a sequence identity cutoff of 60 % was applied to the full T3SE database to identify clusters of homologous proteins [152]. From each cluster, the sequence with the highest amount of annotated information was selected as a representative sequence, yielding a final dataset size of 623 entries.

3.2.4. Cytoplasmic proteins dataset assembly

The majority of prediction methods to identify type 3 effector proteins rely on the binary classification of candidate sequences into T3SEs or non-T3SEs. As an assessment of the performance of our assembled datasets, we also implemented a classification model to distinguish T3SEs from other proteins of T3SS-harboring bacteria. To train our model, the compilation of a second class of proteins not secreted by the T3SS machinery was necessary to allow prediction of two distinct protein classes. This class is referred to as the ***negative*** training dataset.

Previously, I employed the positive T3SE dataset used for training the EffectiveT3 program as a reference to supplement dataset 1 with sequences that were not initially included in our dataset (subsection 3.2.1). For assembly of the negative training set, the negative training set of the EffectiveT3 publication was filtered to exclude sequence entries shorter than 60 residues, followed by a random subsampling procedure (for python code see appendix B subsubsection A.2.1.2). Entries were randomly selected until the number of selected

sequences matched the size of the positive training dataset (244). Combining both datasets, we obtained a balanced 2-class dataset of 488 entries.

For dataset 2 (subsection 3.2.3), we chose to assemble our own negative training dataset to retain a better control over the selection process and origin of sequences. In consideration of the abundance of well characterized, manually reviewed cytoplasmic proteins in the Uniprot database, we chose to assemble a negative dataset of cytoplasmic, bacterial proteins.
A Swiss-Prot search was conducted for cytoplasmic proteins from all bacterial species in our T3SE database. According to the relative abundance of effectors per species, cytoplasmic proteins were selected from each species. Naturally, the amount of identified cytoplasmic proteins is much larger than for effector proteins, thus following homology reduction using CD-Hit (60 % identity cutoff), an additional subselection procedure was introduced. Using a python script, a random subselection of cytoplasmic proteins was made to adapt the size of each dataset class (see A.2.1.2). This was done to ensure that the prediction model does not preferentially learn predicting cytoplasmic proteins. In the case of an unbalanced dataset, there is always a risk that while training, the model learns to predict the majority class rather than distinguish between both classes.

Final dataset The combination of both protein classes into a single dataset suitable for performing predictive tasks led to 2 datasets with balanced class distribution of 488 and 1246 sequence entries, respectively. These are referred to as dataset 1F and dataset 2, respectively. Given the incomplete status of the T3SE database, dataset 2 will likely undergo changes in both sequence composition and dataset size.

3.2.5. LSTM prediction

In the past years a variety of model architectures have been employed to tackle the prediction of T3SEs, improving the correct classification of new effectors constantly. Although reaching prediction accuracies of up to 94 % [176], neither of the developed techniques has advanced the understanding of the biological principles underlying T3SS-mediated secretion. For sequential data, recurrent neural network (recurrent neural network (RNN)) models have become popular due to their ability to capture information between different sections of a sequence. Long-short term memory network (LSTM) models have advanced the ability of these models to capture long-term dependencies within data by introducing gate functions [177]. The introduction of the Attention mechanism by Bahdanau et al. in RNNs ultimately led to the now ubiquitous Transformer architecture [178], [179]. Attention allows to identify positions in a sequence input that are important for the correct output prediction. Therefore, it could provide useful insights into the underlying features and dependencies qualifying an N-terminal region as a T3SS secretion signal.

As a starting point, a LSTM model was assembled to assess the performance of dataset 1F (3.2.1) in a binary classification task and to experiment with feature extraction methods

and model variants. The dataset was split into training and test dataset at a ratio of 80:20. Using cost function and accuracy as metrics, a simple single-layer LSTM was constructed. The loss measures the error between a single prediction and the actual value, the cost function is the average of all losses per epoch (1 iteration over the training data). In a binary classification context, the accuracy describes the proportion of correct predictions out of all predictions.
Despite their thorough and meticulate assembly, both datasets are considerably small for machine learning and especially deep learning approaches. To circumvent the lack of available data in many biological questions, the transformation of sequence data is often accompanied by feature extraction methods that aim at enhancing the information content of an input sequence. In protein prediction tasks, protein sequences are often translated using biological features such as amino acid composition or physico-chemical properties. The human-readable SMILES representation of molecules has been widely applied in chemoinformatics and also proven beneficial for bioinformatic approaches such as protein-ligand binding prediction [180]. The usage of such feature extraction methods can help a model make better predictions but also bears the risk of introducing a bias made by the researchers choice of properties. Mathematical feature extraction methods as proposed by Bonidia et al. and others aim to bypass this risk [181]. Using a single-layer LSTM, the conversion of sequence data into numerical or one-hot encoded vectors, structural representations via SMILES formatting and mathematical feature extraction using Discrete Fourier transforms was assessed [182]. Interestingly, none of the tested feature conversions had a beneficial impact on the accuracy of the model, therefore we retained a simple conversion to numerical format. The implementation of an embedding layer had the largest effect increasing the accuracy on the test set from 70 % to 84 % (Figure 3.34). Experiments with additional layers or the usage of bidirectional LSTMs did not increase the final accuracy.

In conclusion, we established a starting point for classification of T3SEs using machine learning. Compared to published work by other groups that employ sophisticated models to obtain state-of-the-art prediction scores, our approach presents a preliminary working status [183][148][184]. A more comprehensive performance assessment using additional metrics and the application of the larger dataset 2 might help improve increasing the predictive power, furthermore switching to different model structures should be investigated.

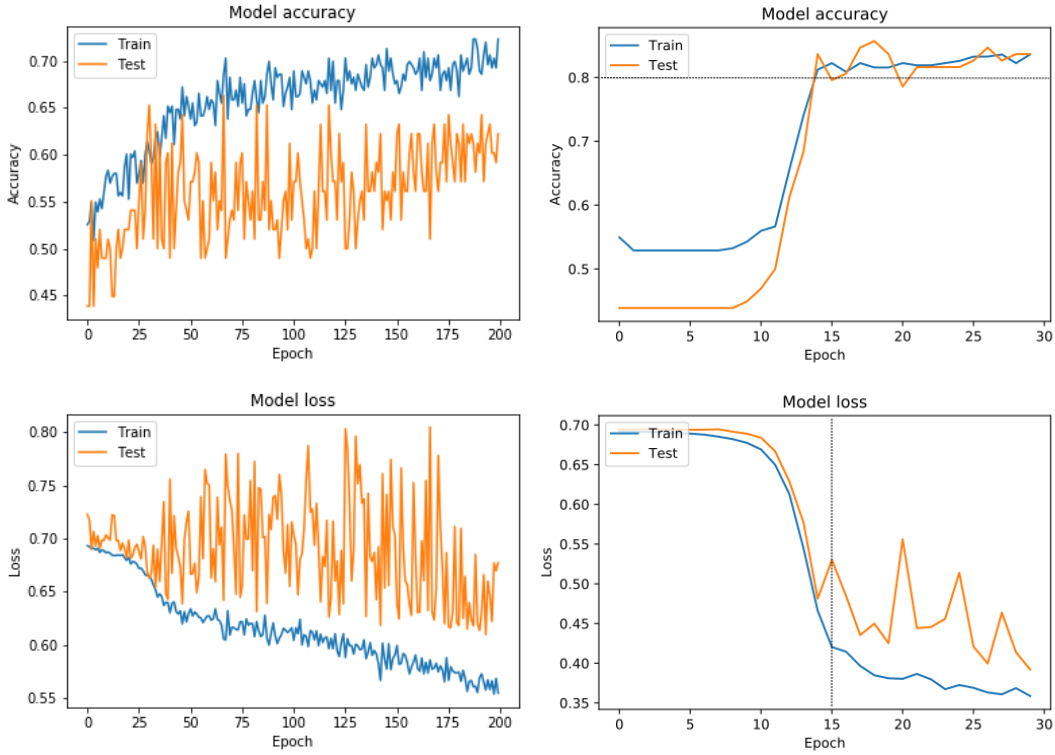


Figure 3.34.: The implementation of an embedding layer increases the models accuracy. A A single-layer LSTM was trained for 200 epochs on dataset 1F. **B** A single-layer LSTM with an embedding layer was trained for 30 epochs on dataset 1F. The embedding layer increases the final test accuracy and the speed or learning drastically.

4. Discussion

Secretion and translocation of substrates via the T3SS is a highly regulated process that is modulated on multiple transcriptional, translational and posttranslational levels [185][186][187][188]. The necessity of the extreme N-terminal secretion signal for T3SS-mediated secretion has been the focus of previous studies, but an explanation for its role in secretion and an understanding of the intrinsic properties that qualify a sequence to function as as secretion signal have yet to be resolved [101]. Despite extensive knowledge of the structural assembly of the T3SS, the scarcity of experimental data on secretion quantities and kinetics of T3SS substrates have impaired painting a more comprehensive picture of the T3SS effector recognition process and in particular of the role of the secretion signal during passage through the T3SS.

The few instances that went beyond the qualitative characterization of the effector N-termini, have implied an impact of the N-terminal amino acid composition on the secretion efficiency of a T3SE, prompting us to develop a methodology that can monitor the isolated effect of the N-terminus on the amount of secreted protein [106]. In this work, a quantitative high-throughput T3SS secretion assay was established that enables an assessment of the subtle differences within the peptide sequence of secretion signals that impair or promote efficient T3SS-mediated secretion. By assembling the largest and most comprehensive database on T3SS-secreted proteins, the foundation has been laid to experimentally determine the effects of over 700 secretion signals on the efficiency of secretion and to elucidate the role of the secretion signal in a mechanistic context. Moreover, the extended annotation of effectors in our database provides an excellent source of information to foster the investigation of novel and uncharacterized effectors.

4.1. Assay design

In its current version, the developed quantitative T3SS secretion assay relies on the quantification of luminescence signals emitted from the *Salmonella* effector SptP tagged C-terminally with the engineered Nanoluc luciferase from *Oplophorus gracilirostris* [142]. By replacing the N-terminus of SptP with secretion signals from other T3SEs and quantifying its impact on secreted reporter protein in the supernatant, an assessment of the secretion efficiency is made.

For simplicity, we adapted a T3SS secretion assay that is routinely applied to the qualitative detection of T3SE secretion. To serve the demands of a high-throughput setup, we chose to complement a suitable *Salmonella* carrier strain with a vector carrying the SptP reporter effector. Whilst this approach is not capable of fully preserving the native induction conditions, it eliminates the need to laboriously manipulate the genome for each secretion signal variant, drastically reducing the workload. In comparison to expression from the chromosome, the use of a vector-based reporter strategy results in a higher reporter expression due to the elevated copy number of the plasmid. Since the intrabacterial effector concentration influences the secretion hierarchy, we aimed to limit the extent of additional expression by incorporating the nucleotide region upstream of the *sicP-sptP* operon containing the natural promoter region and utilizing the low-copy plasmid pWSK29 as backbone [117]. Alternative induction methods using an arabinose-inducible plasmid expressing the master regulator HilA were also assessed, but given the potential disruption of the secretion hierarchy and efficiency, this approach was abandoned in the final version of the assay (Figure A.1).

Secretion signals have been identified in substrates from many T3SS-harboring bacterial species [34]. While there is a consensus on its essential role in guiding effectors towards the T3SS, definitions in regard to its length are diverging [92]. Secretion signal aided transport of substrates via the T3SS has been demonstrated for secretion signals as short as 5 residues, while other substrates require at minimum 10 to 15 residues to promote export [92]. Importantly, the N-terminal fusion of shorter secretion signal peptides to a reporter generally mediates weaker secretion efficiencies compared to longer segments that often facilitate secretion comparable to wildtype levels. In some instances, the residues downstream of the secretion signal contain additional translocation signals or chaperone binding domains [74]. A search through 77 identified chaperones and their targets revealed that none of these signals appear before residue 27 in any validated effector protein. Hence, to exclude the possibility of hampering with these domains when replacing the N-terminal secretion signal, we chose to study the effect of the first N-terminal 25 residues on secretion quantity and allow a clear separation between secretion signal and CBDs or translocation domains.

4.1.1. Reporter choice

The *Salmonella* effector SptP was selected as a scaffold for quantifying the effect of individual secretion signals. In contrast to other effectors like SopE, SopE2, SopA and SipA that share the multi-cargo chaperone InvB, SptP is chaperoned by the single-cargo chaperone SicP conferring secretion specificity towards the nfT3SS [87]. Considering that differential binding affinities of chaperone-effector pairs to the ATPase SctN might also contribute to the secretion hierarchy, the use of a reporter that competes with other effector proteins for available chaperones might therefore prove to be detrimental [82]. Using a reporter protein like SptP that relies on a single specific chaperone, eliminates the risk of a secretion rate limiting competition.

The qualitative assessment of SptP expression and T3SS secretion provided satisfactory results, but the introduction of N-termini from other effectors in place of the native SptP N-terminus initially abolished expression and secretion. Despite the rich annotation status of SptP (Uniprot entry: SPTP_SALTY (2024-04)), a review of available literature revealed the erroneous deposition of the *sptP* coding sequence. In 2011, Button and Galán investigated the translational coupling of the SicP-SptP operon [136]. They also experimentally identified an alternative starting site for the *sptP* CDS that does not concur with the Uniprot-annotated sequence. Located 24 bp downstream of the Uniprot-annotated start codon, an unusual TTG codon marks the beginning of the coding sequence of SptP, truncating its actual length by 8 amino acids.

In our work, shifting the secretion signal exchange site to the experimentally verified start codon restored expression and secretion of the first secretion-signal reporter fusion protein ($SipA_{1\text{-}25}SptP_{26\text{-}535}$), confirming the experimental work conducted by Button and Galán. The translational coupling of SicP and SptP is mediated via two stem loop structures forming in the messenger RNA (mRNA). Whereas the first stem loops plays a critical role in regulation of SptP expression, disruption of the second stem loop containing the TTG start codon does not alter translation of SptP [136]. Hence, replacing the N-terminus of SptP with N-termini from other effectors should not conflict with translation of the reporter protein.

In summary, SptP has proven to be an appropriate choice as a reporter for our quantification assay, despite its highly regulated translation process. This work corroborated evidence for a shorter CDS of SptP, thus an update of the outdated Uniprot entry should be submitted.

4.1.2. Reporter quantification

We investigated two methodologies to quantify the reporter protein. Using a mass spectrometric approach to quantify the complete T3SS secretome was quickly abandoned due to the workload per sample, higher costs and insufficient detection of effectors resulting from an inadequate growth medium masking the presence of most *Salmonella* proteins. Theoretically, the holistic detection of all T3SS-secreted proteins might provide an intriguing

insight how the secretome composition is influenced by intrabacterial effector expression levels and whether different secretion signals alter the secretion hierarchy and magnitude of secretion. In principle, switching to alternative growth and sample preparation protocols as conducted by Auweter et al., using LPM medium instead of LB medium might have improved the detection of *Salmonella* proteins. The higher workload per secretion signal variant however impaired an application in a high-throughput setup.

In order to accelerate and simplify the simultaneous quantification of multiple secretion signal variants, we adopted a luminescence-based quantification method by C-terminally tagging the SptP reporter protein with the Nanoluc luciferase. To investigate effector proteins and elucidate their role *in vivo*, researchers often label candidate effectors with suitable protein tags. For instance, to monitor Type III effector translocation into host cells, the creation of translational fusion proteins of an effector with the adenylate-cyclase domain of the *Bordetella pertussis* toxin adenylate cyclase (CyaA) or TEM-1 BLA has been widely established [190][191]. The CyaA assay exploits the calmodulin-dependence of the adenylate cyclase domain of CyaA to convert adenosine triphosphate (ATP) to cyclic Adenosinemonophospate (cAMP) requiring the eukaryotic host cell environment.

For detection of TEM-1 activity, Foerster resonance energy transfer (FRET)-based substrates such as CCF2/AM or CCF4/AM are trapped inside the host cell, cleavage of a β-lactam linker within the substrate results in a detectable fluorescent shift upon translocation of TEM-1 tagged effectors [192] [190]. As a tool for quantifying T3SS-mediated secretion, these type of tags are not applicable as they require the infection of host cells.

As shown by Westerhausen et al., the Nanoluc luciferase is well suited for performing large-scale reporter assays in automated fashion [141]. Known as one of the brightest and smallest luciferases to date, Nanoluc provides an excellent signal-to-noise ratio and a strong signal strength [193]. To prevent a competition of our reporter with the native chromosomally-encoded SptP protein, a SB905 Δ*sptP* mutant strain was created using the pORTMAGE methodology [125][194]. Notably, while we did not notice changes in the phenotype of the created strain, a recent study reports off-target mutations induced via the pORTMAGE mutagenesis protocol [195]. To rule out additional and potentially deleterious mutations, a comparative genome sequencing of our carrier strain with the original wildtype strain could resolve whether mutations were introduced.

In our cell culture setup, we chose to implement a 96-well format as a compromise between the simultaneous measurement of a large amount of samples and a cell culture volume sufficient to obtain stable, detectable secretion (per-well). In the current setup, this allows to sample six different strains with eight biological replicates per plate (V=1.5 mL), practically the processing of 18 different strain variants plus control strains is manageable per assay. To attain a satisfactory statistical certainty, we chose to grow 8 cultures per tested secretion signal variant, each culture was quantified in triplicate.

We optimized all steps of the quantification protocol, starting with the cell culture conditions, sample handling as well as device settings. The remaining issue of elevated cell lysis

during incubation could not be eliminated by changing the culture conditions. Presumably, the geometry of the 96-well deep well plates acts as an adverse factor for viable cell growth. Considering that low-oxygen conditions similar to environmental conditions in the small intestine have been shown to induce the activation of T3SS SPI-1 regulatory genes, other factors such as shear stress might be the leading cause for the observed cell death [196][197]. To account for Nanoluc signals originating from cell lysis, every assay included a negative control of a T3SS-deficient strain (SB905 Δ*invA* Δ*sptP*) to define a baseline signal unrelated to T3SS secretion. As a tradeoff between the optimal signal-to-noise ratio and a sufficiently high throughput of samples, the endpoint of incubation at 4 h was selected at the cost of a decreased signal resolution. Differences in growth rate were taken into accoung by measuring the optical density of each individual culture and normalizing the cultures' luminescence signal against the optical density.

4.1.3. Expression control

To reliably evaluate the secretion efficiency of our reporter, we chose to monitor the intrabacterial reporter concentration. Common techniques to study the secretion of translocation of effectors and their kinetics often neglect the total amount of protein produced and focus on the amount of secreted/translocated protein in the host cells. For some effectors an increase in intrabacterial effector concentration promotes higher translocation rates into host cells as shown for *Escherichia* effectors EspF, EspG, EspH and Map [117]. Accordingly, an estimate whether elevated secretion quantities are the result of increased effector expression or can be ascribed to the introduced N-terminal secretion signal variant is of central essence [117]. In this work, we compared two approaches to assess expression.
First, a second reporter effector immediately downstream of *sicP-sptP* was incorporated to set the secretion quantities of each SptP reporter variant in relation to a stable, unaltered second reporter effector. We utilized the *Salmonella* SPI-1 effector SipA that has been shown to be secreted in high abundance in T3SS-mediated manner [198]. By including the 18 bp 5'-UTR region bridging the CDSs of *sipD* and *sipA* on the *sipBCDA* operon, an artifical *sicP-sptP-sipA* operon was created. InvB, the chaperone for SipA, was not included as it also primes the effectors SopE, SopE2 and SopA for secretion. Hence, a plasmid-mediated increase of InvB could have farther-reaching effects on secretion hierarchy affecting multiple effectors and thus complicate the interpretability of our assay.
SipA was C-terminally tagged with the red firefly luciferase. Despite its large size of 60 kDa and a significantly weaker signal strength, the firefly luciferase is one of the few luciferases compatible with Nanoluc in dual-use systems as both require different substrates for activity. Using a Δ*sptP* Δ*sipA* strain complemented with plasmid variants carrying a secretion-signal SptP-NLuc fusion and the native SipA-RFLuc, the quantitative detection of both luciferases was conducted. Unfortunately, the firefly luciferase displayed extremely

weak luminescence signals with average signal strengths below ranging from 50 RLU to 10 000 RLU. For comparison, Nanoluc luminescence signals displayed up to 1000-fold higher signal intensities. In relation to the absolute signal strength this resulted in high variance between samples, making the firefly luminescence signal unreliable as a reporter secretion control. The cause for the weak and highly variable signals emitted from the firefly luciferase is unclear. A relationship between size and secretion quantity has been indicated by the differential secretion rates of SopE2 and SptP, with SopE2 (26 kDA) being secreted at a roughly 2-fold rate to SptP (59 kDA)[84]. Assuming a constant rate of secretion, it is therefore conceivable that the large size of the SipA-RFLuc construct (134 kDA) attenuates secretion of higher quantities of SipA-RFLuc and prevents a reliable detection. This infers that the luminescence signal variance in is within normal bounds, but disproportionally elevated by the low average signal intensity. In retrospective, choosing a smaller and brighter luciferase like the *Gaussia* (*Gaussia* luciferase (GLuc)) or *Renilla* luciferases (*Renilla* luciferase (RLuc)) for detection would have probably been advantageous despite cross-reactivity of their substrates with the Nanoluc luciferase [199]. As performed by Sarrion-Perdigones et al., individual signals of some coelenterazine-consuming luciferases can be distinguished using band-pass filters on the respective emission maxima of the luciferases, thus separately facilitating quantitative detection of NLuc, GLuc and RLuc luminescence [199].
Instead, we included an additional sample preparation step by measuring luminescence signal intensities in the cell pellet. Because the NLuc-substrate furimazine is membrane-permeable, we were able to obtain a quantitative measurement of non-secreted reporter protein and derive a relative secretion efficiency for every secretion signal variant tested.

Statistically, the variance in our assay is comparable to other studies. As described by Westerhausen et al. that proclaim a coefficient of variation (CV) of 7 %, the luminescence signals of the supernatant samples currently display a CV of 7.3 % in our assay. The degree of variation in the pellet samples is considerable higher with 27.1 % thus further optimization of the sample preparation procedure might prove beneficial. A potential cause of the displayed sample heterogeneity could be that the Nanoluc substrate furimazine has to diffuse through the cell membrane into the bacteria. According to the manufacturer, membrane permeability was tested on eukaryotic cells, thus the diderm architecture of *Salmonella* might complicate a homogenous accessibility throughout the pellet sample. By adapting the pellet sample preparation procedure and solubilise the cell pellet with a detergent of choice instead of resuspending it in growth medium, a more homogeneous sample could be obtained. This remains to be tested. The overall variability between assays diplays another factor to be considered. To monitor general differences we took the positive control sample SB905 Δ*sptP* expressing and secreting $_{\text{FL}}$SptP-NLuc (pMP059) and compared its signal strength over the course of all assays. Astonishingly, the signal strength varied to a significant degree between assays despite identical assay conditions and handling (2.23×10^6 RLU to 1.84×10^7 RLU for $_{\text{FL}}$SptP-NLuc, see Figure 3.25). Consid-

ering the careful optimization of all assay parameters, the elimination of external influence factors and the fact that only a fraction of the bacterial population expresses virulence genes, we attribute this degree of variation to biological factors [83][84]. Thus, as a means to enable comparison of secretion efficiencies across different assays, all secretion efficiencies calculated per assay were normalized against the positive/wildtype control of the respective assay. This assumes that during assay cell culture all cultures behave similar to the wildtype positive control, a conjecture that is difficult to ascertain. Therefore, to narrow the degree of variation it is advisable to quantify each secretion signal variant across a minimum of 2 assays with a large enough sample size (n=4-8) to provide a more comprehensive picture of the factual secretion efficiency and reflect on the observed biological variation.

4.1.4. Effect of the secretion signal on secretion quantity

4.1.4.1. Secretion quantities

To date, secretion quantities for secretion signals of the majority of *Salmonella* effector proteins (42 of 54) have been tested. With our automated primer generation script we streamlined the generation of new plasmids via backbone-PCR. The codon usage of the generated primers followed the *Salmonella* LT2 codon table, thus potential effects of RNA regulatory elements as proposed by Anderson and Schneewind can not be directly studied [93]. The remaining 12 of 54 constructs failed to assemble despite manual examination of all primer pairs.

As previously reported, T3SS secretion signals contain seemingly universal characteristics facilitating transport even in heterologous bacterial species [158][200]. To assess the general applicability of our approach, we corroborated this by testing secretion signals from the *Yersinia* effectors YopH and YopO, *Shigella* effector OspB, *Escherichia* Map and NleE, as well as *Chlamydia* effectors CT_115 and CT_223 in our *Salmonella*-based secretion quantification assay (Figure 3.11).

Initially, secretion quantities for 42 secretion signal variants could be obtained using an arabinose-inducible HilA plasmid for enhanced expression, whereas the subsequent methodology relying on the natural induction of the T3SS thus far yielded secretion quantities for 25 N-terminal variants. As we excluded all samples displaying secretion quantities below or in the range of the negative control, a fraction of the naturally induced samples were not included in the subsequent analysis. Several of the excluded samples also exhibited low expression levels, a potential cause for the weak secretion level leading to their exclusion. Notably, the reporter fusion $SipA_{1\text{-}25}SptP_{26\text{-}535}$ was expressed and secreted poorly, a finding that stands in contrast to our initial proof-of-principle and the high abundance of SipA in other secretome studies [198]. While for the initial proof-of-principle the native nucleotide sequence was used, the later approach in our high-throughput setup followed the automated primer generation procedure that results in an altered nucleotide usage possibly diminishing SipA expression. Especially rare codons can have a major impact on protein expression, it

would be therefore of interest to investigate a potential link between the codon usage of the created reporter fusions and the degree of expression from our plasmid [201]. Weakly expressed secretion signal reporter fusions could thus be easily identified and optimized for enhanced expression. Furthermore, it might be a valid strategy to investigate different codon compositions for poorly expressed N-terminal variants.

Intriguingly, a notable number of sequences revealed only minor differences of their secretion quantities relative to the wildtype signal for natively and HilA-induced samples (Figure 3.27). Although the natural induction method produces higher secretion quantities relative to the wildtype signal, both methods are comparable to a limited extent. It is therefore reasonable to speculate that the secretion efficiency is in parts determined by the secretion signal, irrespective of the expression level of the reporter.
A ranking of the secretion quantities revealed that the wildtype reporter sequence makes up the majority of the highest secretion quantities. This indicates good robustness of the assay as similar secretion quantities are observed between different assays. However, it also implies that the native SptP signal sequence is highly optimized towards expression and secretion of SptP. The replacement of the native N-terminal sequence with different secretion signals might therefore entail the disruption of this delicate regulation process, resulting in an alteration of expression and secretion levels of all secretion signal variants. Nonetheless, considering that both expression and secretion of the reporter is quantified in our setup, the derivation of relative secretion efficiencies allows to extract valuable information.

4.1.4.2. Secretion efficiencies

To evaluate the measured secretion quantities, we also quantified the amount of expressed reporter protein. As the first attempt using a second reporter protein failed, the current setup includes quantification of luminescence signals from pellet samples. The variation of the luminescence signals measured in the pellets was significantly higher than in the supernatant samples indicating a heterogenous sample composition. Despite this, the estimate of the total expressed reporter protein is a crucial parameter for evaluating the effects of the tested N-termini on expression and secretion. Considering the apparent variability between positive/wildtype controls in different assays, calculating relative ratios of secretion efficiencies and normalizing them to the wildtype secretion efficiency of the respective assay provided the means for a comparability between assays. Furthermore, it also allows to trace whether weak luminescence signals in the supernatant are the cause of low expression or low secretion efficiency.
The limited number of tested N-termini thus far prevented identification of a correlation between high secretion efficiencies and the secretion signals composition. Expanding the amount of tested N-termini will be necessary to gain insights to what drives efficient secretion. It will be crucial to further expand the generation of quantitative secretion data,

thus eliminating the time-consuming generation of reporter variant plasmids could be accelerated by the application of commercially available combinatorial libraries .

4.2. T3SE database

The assembly of the T3SE database, whilst not envisioned to such a scope in the initial project idea, constitutes the foundation for all subsequent work. Data of the highest possible quality are required both for the large-scale experimental analysis of secretion quantities and for the development of a prediction pipeline for the identification of T3SEs. In light of published type 3 effector databases and datasets often lacking traceable sources for included effector sequences, containing erroneous sequences and the outdated or unavailable status of many datasets, we shifted our focus towards assembly of our own T3SE database [202][101]. Despite the fact that many of these datasets have successfully been used for T3SE prediction tasks, the experimental status of its entries often remains unclear and a rationale behind the separation of verified versus putative effector proteins is lacking.
To achieve the highest possible quality for our database and guarantee the exclusive incorporation of T3SS-secreted proteins, we established criteria defining what type of experimental evidence qualifies a protein to be categorized as a validated T3SS-secreted protein. Experiments lacking suitable controls were disregarded and the respective effectors listed as putative T3SEs. In comparison to the largest published T3SE databases we have expanded the number of verified T3SS effector proteins significantly already, despite its currently unfinished status. Datasets of recent prediction models such as DeepT3_4, DeepT3 2.0, EP3, Bastion3 contain 379, 302, 379 and 379 T3SEs respectively based on published datasets [184][203][183][148]. The T3SE databases BastionHub, EffectiveDB and T3SEdb comprise 1194, 504 and 504 experimentally verified effectors respectively, compared to the 735 entries of our database [204][151] [202]. Considering that the BastionHub database includes homologous protein sequences of the same effector, a juxtaposition of our database with Bastion3 T3SS substrates might reveal the fraction of unique effectors within BastionHub [204].
The assembly of most of these databases relied on retrieval of sequences and annotation data from both automated approaches and literature reviews, whereas our approach was a purely manual inspection of literature. While clearly associated with a much higher workload, the high quality of our database justifies the undertaken efforts. Considering the outdated status of many sequence annotations in publically available sequence databases as exemplified on the Uniprot annotation of the *Salmonella* effector SptP, the reliance of several published approaches on the available annotation must be seen as a major source of error introducing false positive entries. In comparison, we can provide additional information regarding effector functions, enzymatic activities, host cell targets and interaction partners, as well as host cell localizations. Critically, all this information is directly derived from experimental data such as *in-vivo* or *in-vitro* binding studies, enzymatic assays or fluorescence microscopy. Albeit still in a preliminary, unfinished status, our database thus provides the most comprehensive collection of information for T3SS effector proteins available to date. Once the assembly process is completed, a thorough analysis of the database will certainly promote the identification of new promising research targets and

questions.

4.2.1. Dataset preprocessing

For any computational prediction/classification task, the prediction outcome is highly dependent on the underlying data used for training. Ideally, a dataset should be of high quality, quantity and variability. With the T3SE database a foundation for a high-quality dataset comprising the largest number of T3SS effector proteins was available. The presence of highly similar sequences in a dataset can introduce biases that lead to overfitting and low generalizability of the prediction model to unseen data. A common procedure to increase variability in a protein dataset is to remove all sequences displaying homology above a certain threshold. In contrast to other studies that used thresholds of 70 %, we opted for a homology cutoff at 60 % to obtain a variable dataset without losing larger fractions of data [184][148]. In principle, even lower thresholds (30 %) as applied by Li et al. result in higher diversity data, but also drastically reduce the amount of retained sequences. We tried to establish a balance between dataset size and variability, nonetheless an investigation into what tradeoff between quantity and variability provides optimal results could be beneficial.

4.2.2. Prediction of secretion efficiencies

While the focus of this work lied on establishing the experimental quantification setup and assembly of the T3SE database, the preliminary results obtained from the LSTM prediction model with a test accuracy above 80 % mark a solid starting point for subsequent work. Whilst lacking a comprehensive assessment of the models performance using additional metrics such as the Matthews correlation coefficient (MCC), F1-score, receiver-operator curve, sensitivity and specificity, the principal application of our datasets to a binary classification problem has been shown. Using different feature extraction methods such as SMILES or physico-chemical features did not improve the models performance, thus the focus of future efforts should lie on optimization of the model architecture.
With the increasing availability of quantitative secretion data linked to a specific peptide sequence from our experimental setup, the application of state-of-the-art model architectures offer an interesting perspective to gain a deeper understanding of the secretion signal properties that promote efficient secretion. Although the usage of an LSTM is an eligible method for predictions on sequential data that has been utilized on T3SEs datasets in the past [184], newer model architectures have superseded the wide application of LSTMs. Specifically, transformer models that rely on the Attention mechanism introduced by Bahdanau et al. offer an intriguing perspective as they allow to identify key positions within a sequence important for making the right prediction [179]. In principle, this might facilitate the deduction of specific sequence features within the secretion signal. A first application

of an attention-based prediction model on T3SE predication was published by Li et al., however without capitalizing on identifying features of potential biological relevance [183]. As a long-term prospect we envision adapting such a model architecture to continously predict secretion quantities/efficiencies based on the input sequence as a regression model rather than a binary classification task.

The rich annotation of our T3SE database with information about enzymatic activity, host cell targets or host cell localization for a large number of effectors might also prove beneficial to assemble multiple specific prediction models and integrate them into a final ensemble pipeline. As shown by the EP3 ensemble prediction pipeline that combines 6 different classifiers, such a methodology can help to achieve a more robust and generalizable prediction outcome [149]. Furthermore, by predicting putative host cell localizations and likely host cell targets, a subsequent experimental characterization of a predicted effector protein would be narrowed, thus streamlining the experimental characterization process.

5. Conclusion

5.1. Conclusion

To conclude, I have established a quantitative secretion assay suitable for high-throughput analysis of secretion signals and provided a platform for testing candidate effectors based on their N-termini. The assay accounts for both the cell growth of individual cultures as well as for the overall expression level of our reporter. Expression differences induced by variant N-termini are thus accounted for. Having quantified the impact of N-termini from *Salmonella* effectors on the secretion efficiency, we indeed identify different secretion efficiencies that can directly be associated with the introduced secretion signal. To gain an understanding which features within the secretion signal is required for efficient secretion, the number of tested N-termini thus far is not sufficient, therefore the expansion of the quantification process to a larger number of effectors is essential. To do so, I have furthermore assembled the largest and most comprehensive T3SE database to date, providing a direct trace of experimentally verified effector sequences to the original publication in which the experiment was conducted. By not only providing sequence data and T3SE evidence, but advanced annotation of experimentally determined host cell targets, enzymatic activities, interaction partners and host cell localization the present database can be used to narrow the experimental characterization of effectors of interest.
Finally, I have conducted preliminary tests that confirm the general usability of the underlying data by applying the assembled dataset to a binary classification task predicting whether candidate proteins are T3SEs.

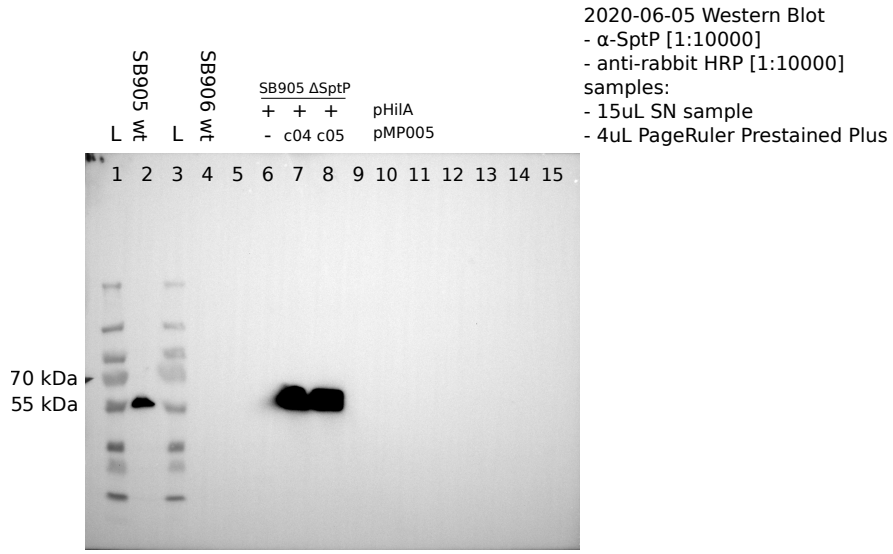
A. Appendices

A.1. Appendix A

A.1.1. Figures

A.1.1.1. Immunoblotting

A



B

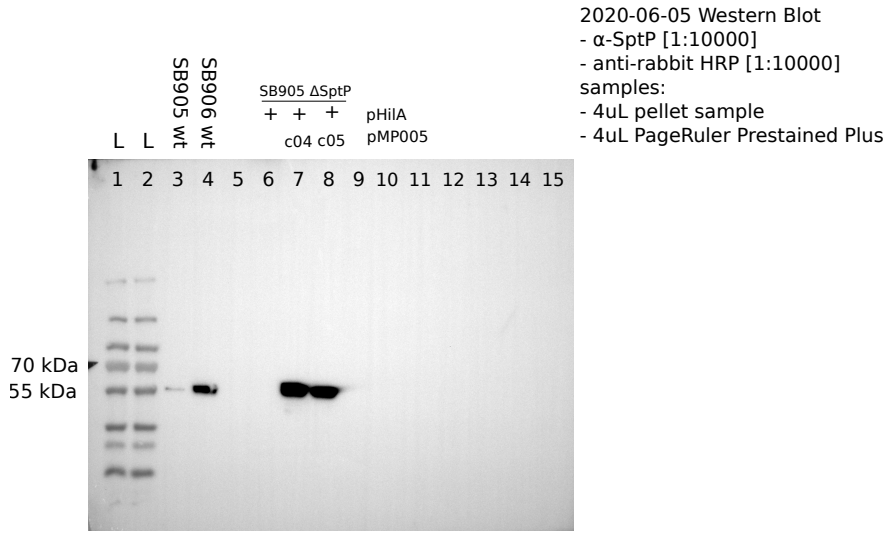
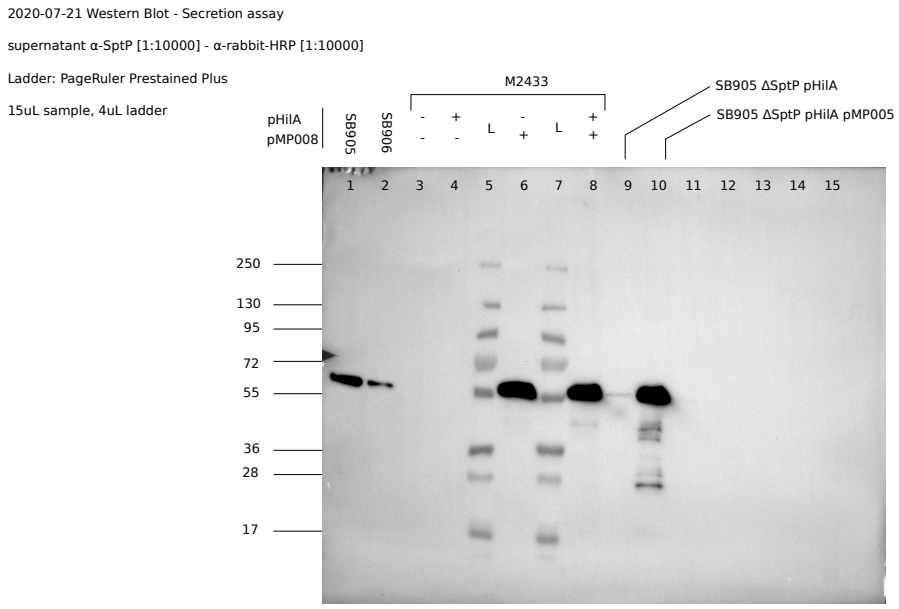


Figure A.1.: Expression and secretion of SptP from plasmid pMP005 as visualized by immunoblotting. *S. typhimurium* strain SB905 ΔsptP complemented with plasmid pMP005 and pHila pacyc184 expresses and secretes SptP. Cultures of wildtype strain *S. typhimurium* SB905 and T3SS-deficient SB905 (Δ*prgH*, negative control) were grown as controls, as well as an uncomplemented strain SB905 ΔsptP. All cultures were grown according to section 2.2.1. Plasmid tags co4 and c05 denote different clones of plasmid pMP005, L=ladder. Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SptP. **A** supernatant samples **B** pellet samples



A Secretion assay confirming expression and secretion of plasmid pMP008

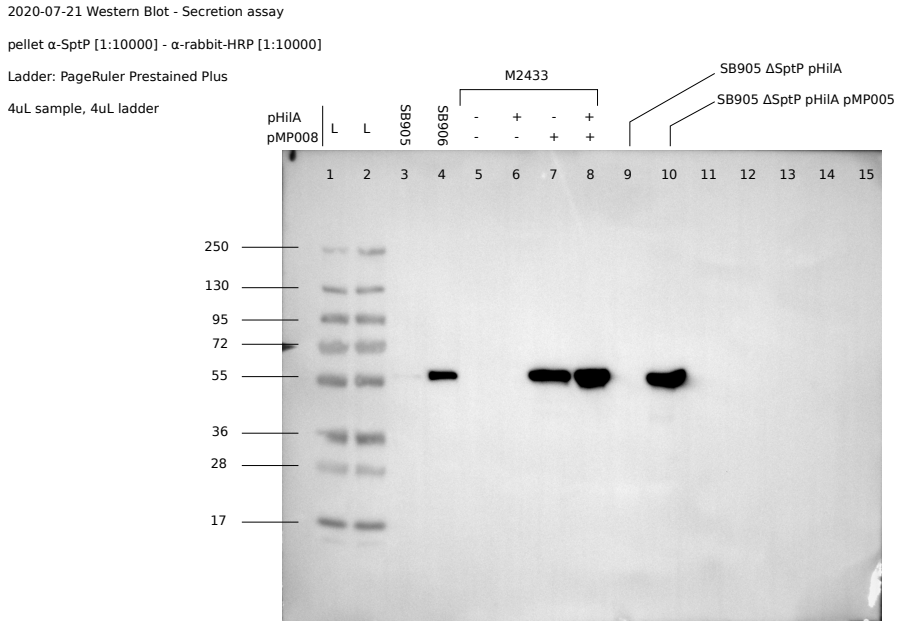
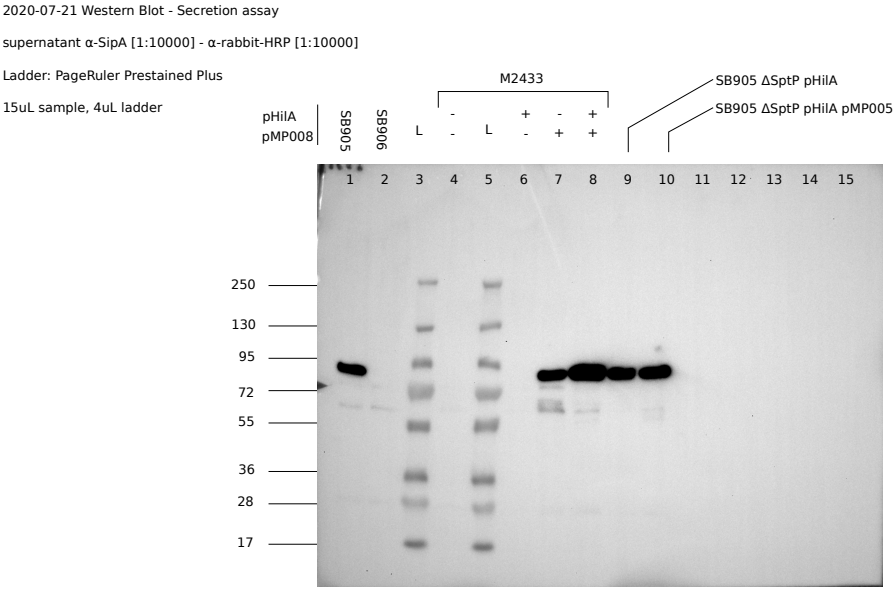


Figure A.2.: Expression and secretion of SptP from plasmid pMP008 as visualized by immunoblotting. Cultures of wildtype strain *S. typhimurium* SB905 and T3SS-deficient SB905 (Δ*prgH*, negative control) were grown as controls, the M4233 strain was grown either uncomplemented or complemented with combinations of pHilA and pMP008. As an additional control strain SB905 complemented with pHilA and pMP005 was grown. All cultures were grown according to section 2.2.1. L=ladder. Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SptP. **A** supernatant samples **B** pellet samples

A



B

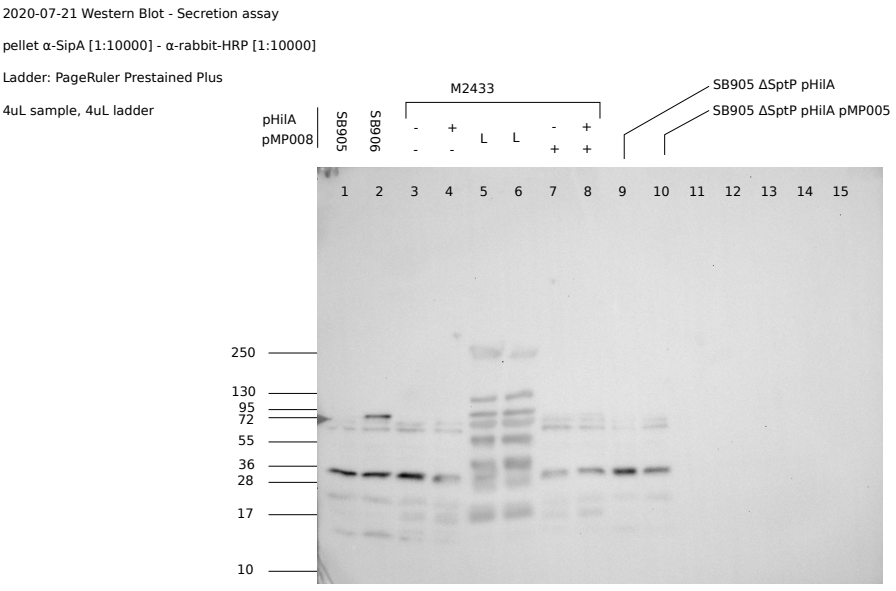
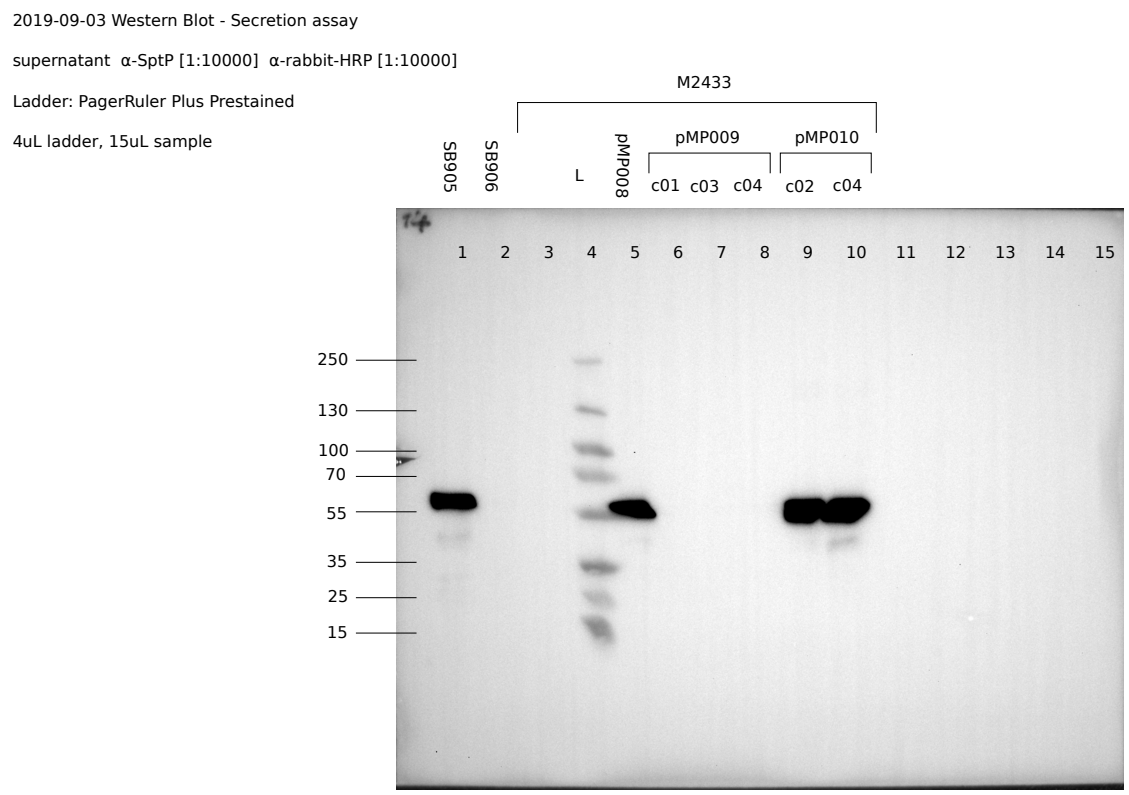


Figure A.3.: Expression and secretion of SipA from plasmid pMP008 as visualized by immunoblotting. Cultures of wildtype strain *S. typhimurium* SB905 and T3SS-deficient SB905 (Δ*prgH*, negative control) were grown as controls, the M4233 strain was grown either uncomplemented or complemented with combinations of pHilA and pMP008. As an additional control strain SB905 complemented with pHilA and pMP005 was grown. All cultures were grown according to section 2.2.1. L=ladder. Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SipA. **A** supernatant samples **B** pellet samples

A



B

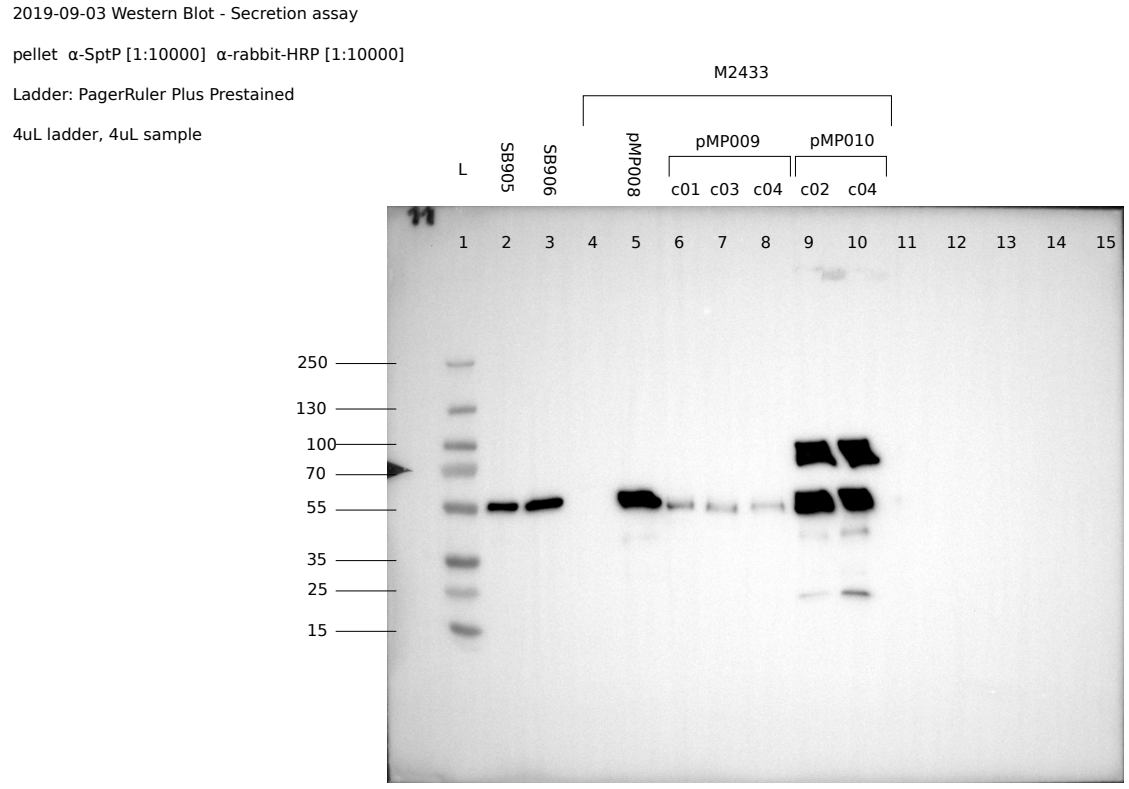
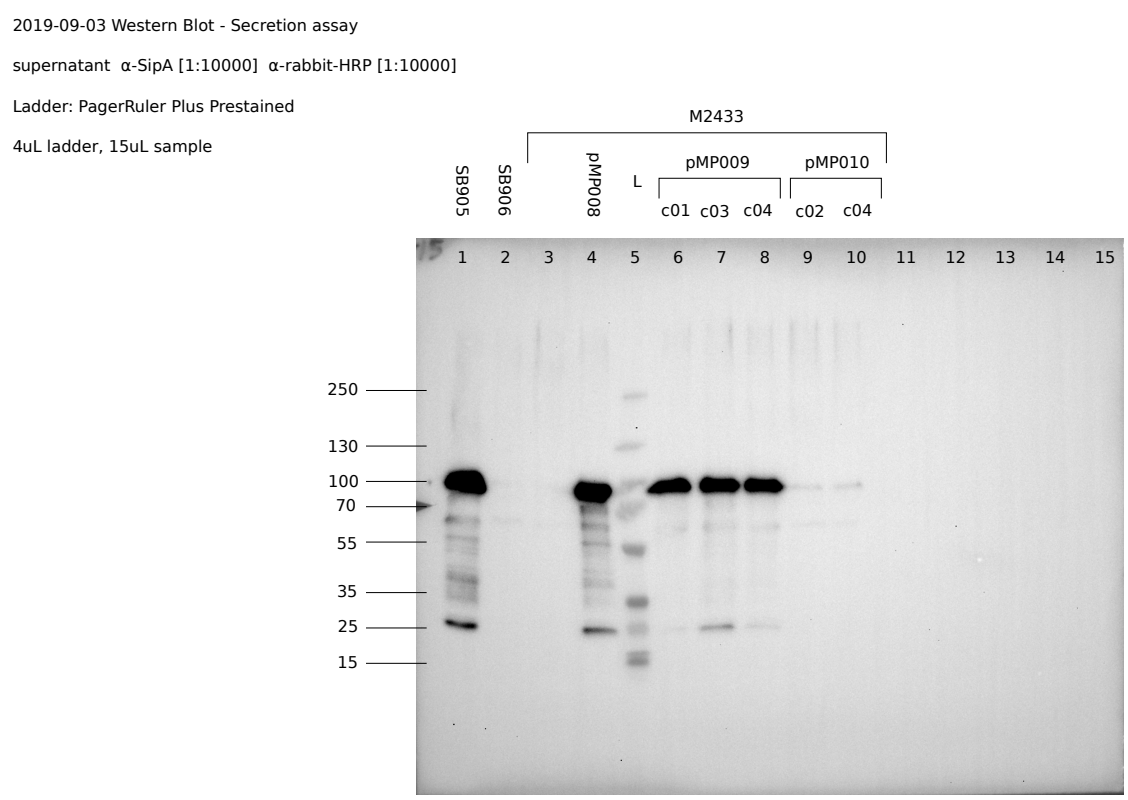


Figure A.4.: Expression and secretion of $SipA_{1\text{-}25}$-$SptP_{26\text{-}543}$ and SptP from plasmids pMP009 and pMP010, respectively as visualized by immunoblotting. Cultures of wildtype strain *S. typhimurium* SB905 and T3SS-deficient SB905 (Δ*prgH*, negative control) were grown as controls, the M4233 strain was grown either uncomplemented or complemented with pMP008 ($_{FL}$SptP, $_{FL}$SipA), pMP009 ($SipA_{1\text{-}25}$-$SptP_{26\text{-}543}$ and SipA) or pMP010 ($SptP_{1\text{-}25}$-$SipA_{26\text{-}685}$ and $_{FL}$SptP). All cultures were grown according to section 2.2.1 Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SptP. **A** supernatant samples **B** pellet samples

A



B

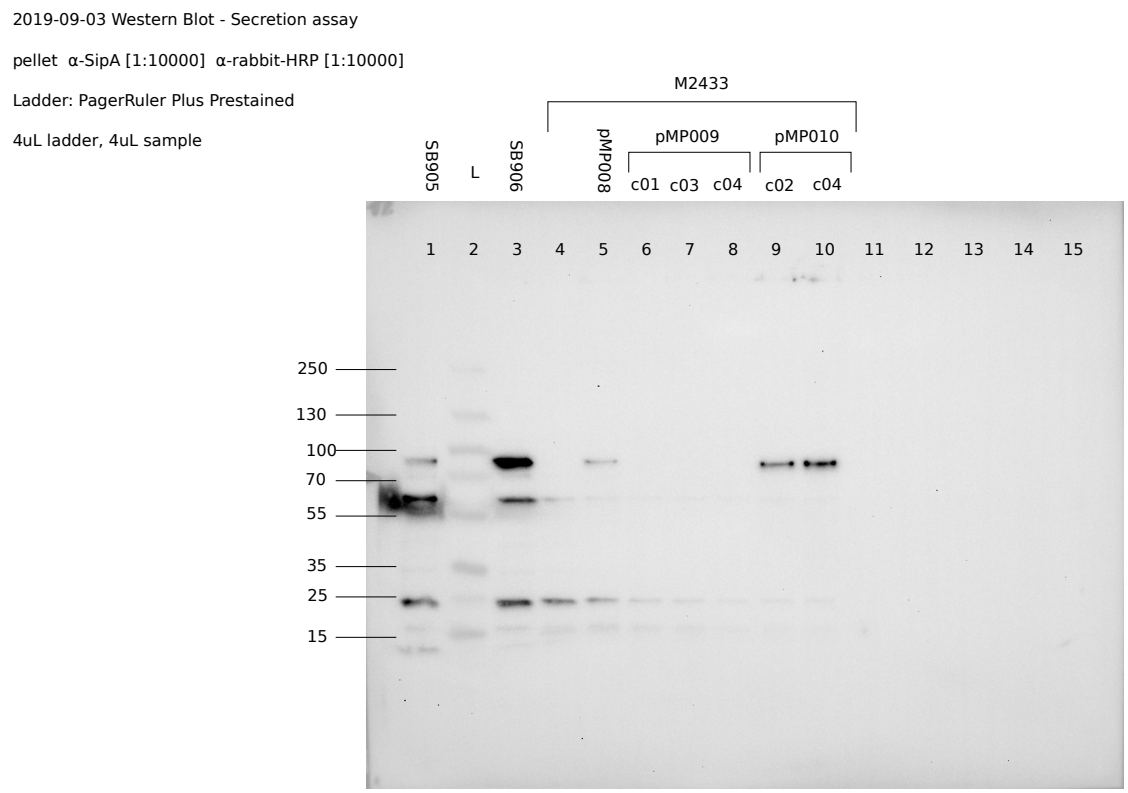
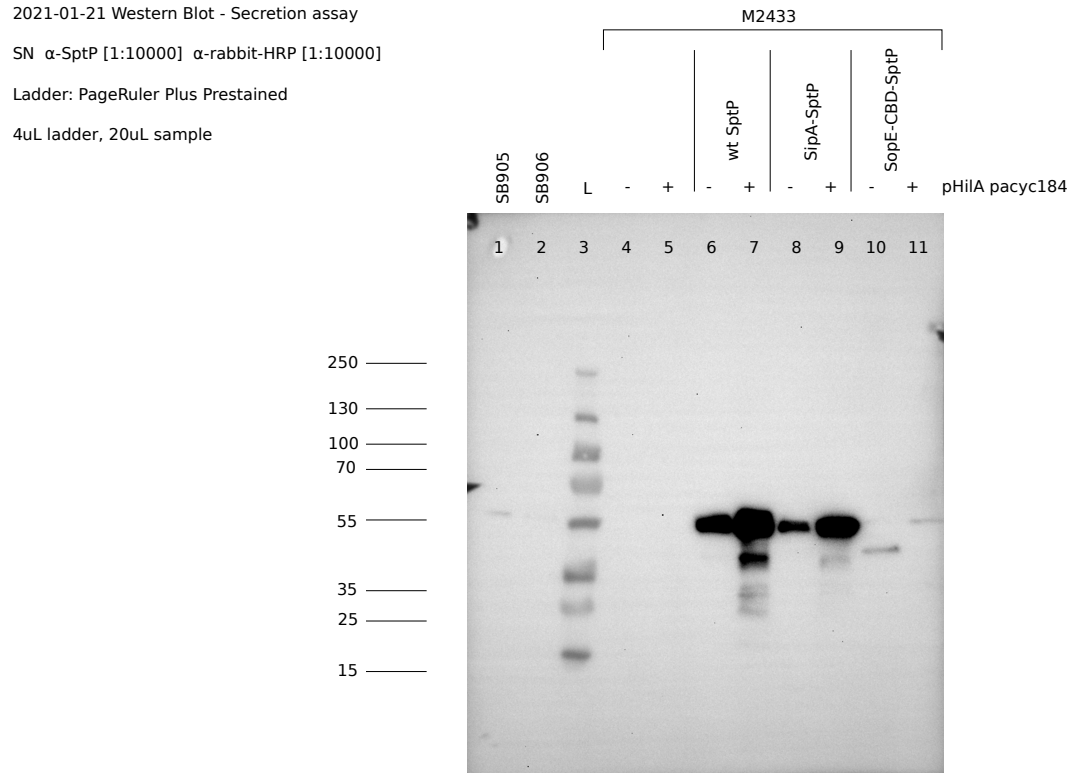


Figure A.5.: Expression and secretion of $SptP_{1\text{-}25}$-$SipA_{26\text{-}685}$ and SipA from plasmids pMP009 and pMP010, respectively as visualized by immunoblotting. Cultures of wildtype strain *S. typhimurium* SB905 and T3SS-deficient SB905 (Δ*prgH*, negative control) were grown as controls, the M4233 strain was grown either uncomplemented or complemented with pMP008 ($_{FL}$SptP, $_{FL}$SipA), pMP009 ($SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$ and SipA) or pMP010 ($SptP_{1\text{-}25}$-$SipA_{26\text{-}685}$ and $_{FL}$SptP). All cultures were grown according to section 2.2.1. Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SipA. **A** supernatant samples **B** pellet samples

A



B

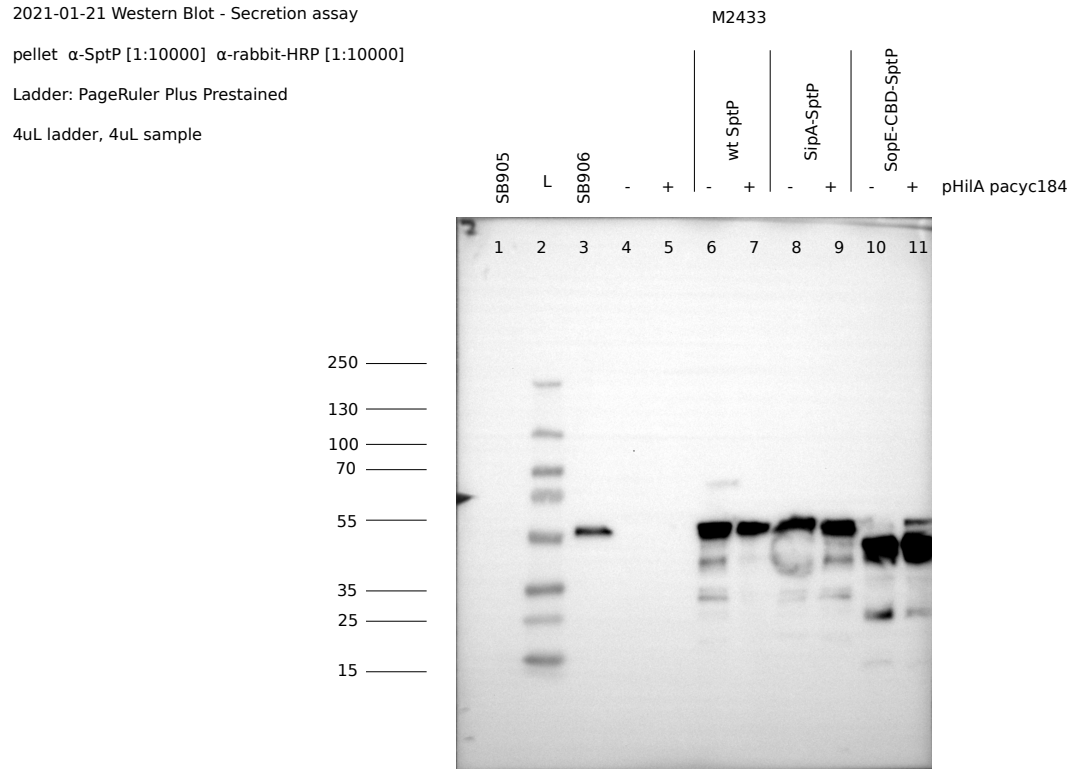
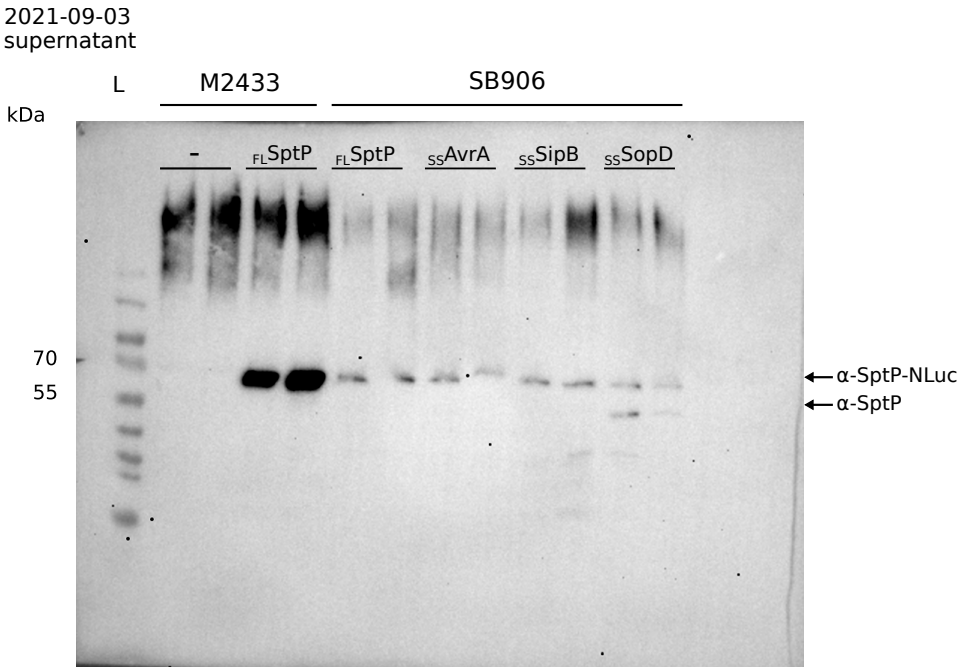


Figure A.6.: Expression and secretion of $SipA_{1\text{-}25}$-$SptP_{26\text{-}535}$ and SptP from plasmids pMP017 as visualized by immunoblotting. Cultures of wildtype strain *S. typhimurium* SB905 and T3SS-deficient SB905 (Δ*prgH*, negative control) were grown as controls, the M4233 strain was grown either uncomplemented or complemented with pMP008 (wt SptP), pMP017 (SipA-SptP), pMP018 (SopE-CBD-SptP) and pHilA. All cultures were grown according to section 2.2.1. Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SptP. **A** supernatant samples **B** pellet samples

A



B

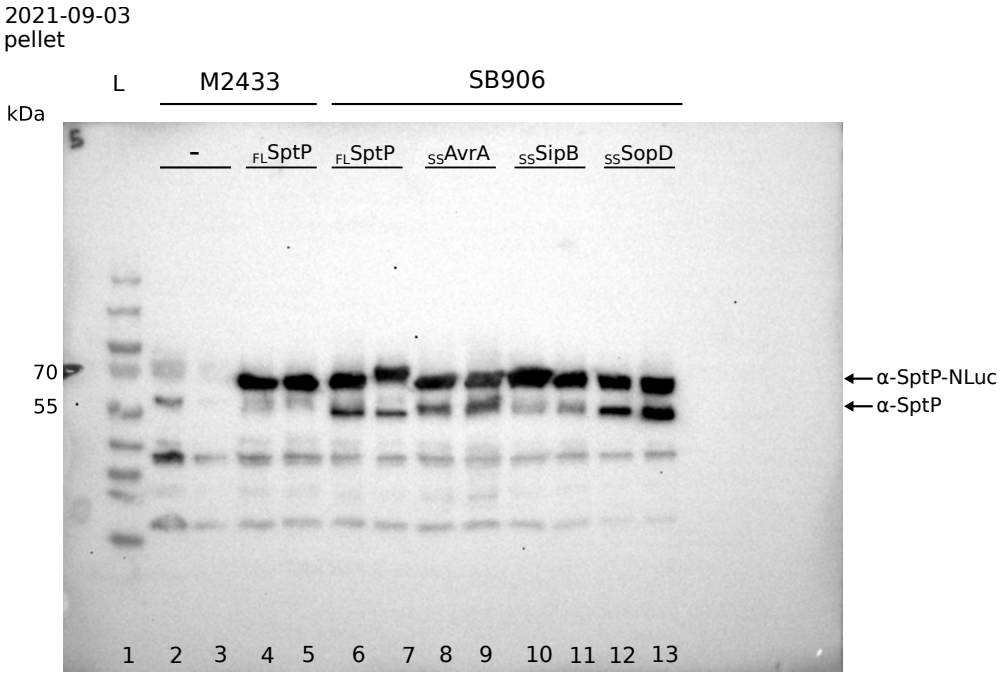


Figure A.7.: Original western blots for expression of reporter variants in T3SS-deficient strain SB906. Expression and secretion of reporter variants for $_{FL}$SptP-NLuc and reporter variants Avra, SipB and SopD was tested in the T3SS-deficient strain SB906 (Δ*prgH*). As a positive control strain M2433 expressing $_{FL}$SptP was grown. All cultures were grown according to section 2.2.1. Immunoblotting was performed as described in section 2.2.4 using a polyclonal antibody against SptP. **A** supernatant samples **B** pellet samples

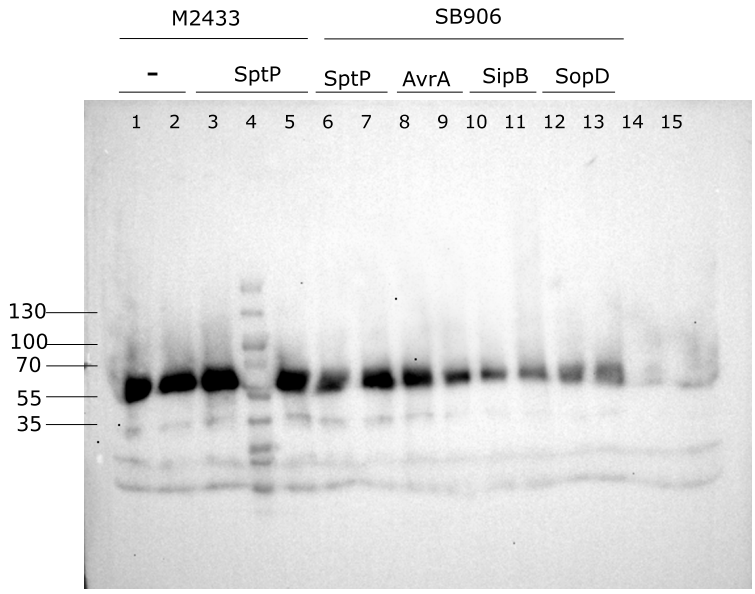
A

2021-09-03 Western blot - secretion assay

SN α-GroEL [1:10000] α-rabbit HRP [1:10000]

ladder: PAgeRuler Prestained

4uL ladder, 20uL sample



B

2021-09-03 Western blot - secretion assay

pellet α-GroEL [1:10000], α-rabbit-HRP [1:10000]

ladder: PageRuler Prestained

4uL ladder, 4uL sample

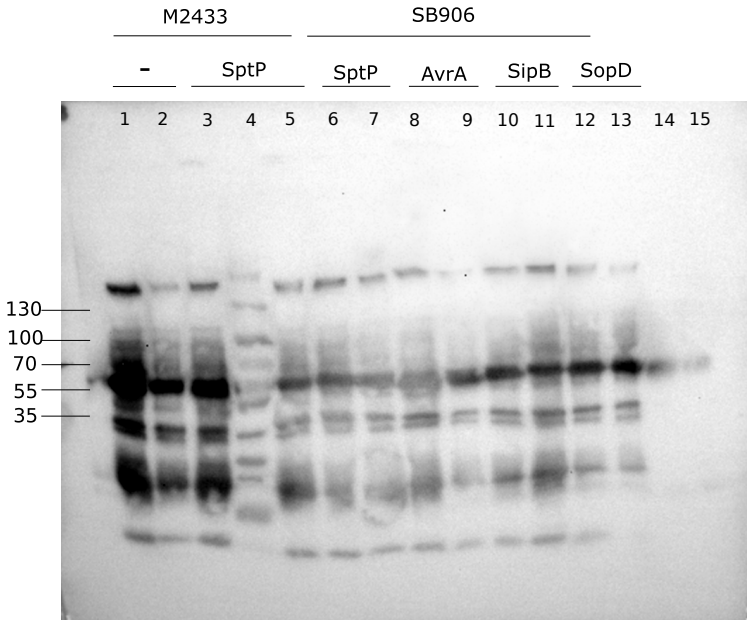


Figure A.8.: Original blot of lysis control western blot. Immunoblotting against the cytoplasmic marker groEL was performed according to section 2.2.4 using a commercial antibody against groEL. All samples are identical to the cultures displayed in Figure A.7. **A** supernatant samples **B** pellet samples

A.1.2. Plots

A.1.2.1. Secretion quantification assays

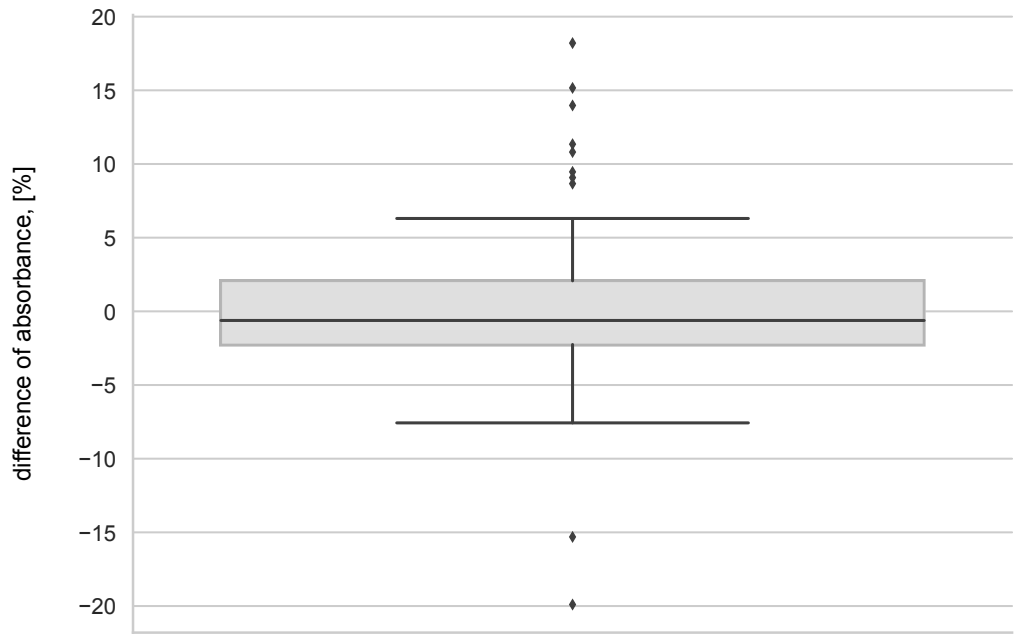


Figure A.9.: The effect of shaking on absorbance measurements is negligible. Using the previously generated cell culture dilutions, absorbance$_{600\,\mathrm{nm}}$ was remeasured in the platereader setup including a shaking step of the 96-well plate (30 s, 300 rpm) prior to the measurement. On average, the difference in absorbance between both measurements stays below 5 %.

A.1.3. Tables

Table A.1.: *Salmonella enterica* subsp. *enterica* Typhimurium LT2 codon table, frequency per thousand Kazusa database. * denotes the stop codons.

residue	codon	frequency [$\frac{1}{1000}$]	residue	codon	frequency [$\frac{1}{1000}$]
F	UUU	23.3	Y	UAU	17.1
	UUC	15.3		UAC	11.6
L	UUA	13.2	H	CAU	13.3
	UUG	12.4		CAC	9.6
	CUU	11.8	Q	CAA	12.7
	CUC	10.4		CAG	31.0
	CUA	4.9	N	AAU	17.8
	CUG	53.6		AAC	20.1
I	AUU	29.3	K	AAA	31.7
	AUC	24.4		CAA	12.7
	AUA	5.3		AAG	11.3
M	AUG	27.4	D	GAU	31.6
V	GUU	15.5		GAC	20.3
	GUC	18.2	E	GAA	35.4
	GUA	11.4		GAG	20.7
	GUG	25.2	C	UGU	4.8
S	UCU	7.2		UGC	6.6
	UCC	10.1	W	UGG	15.2
	UCA	6.2	R	CGU	18.8
	UCA	9.5		CGC	23.3
	AGU	7.4		CGA	3.6
	AGC	17.4		CGG	6.9
P	CCU	7.2		AGA	2.3
	CCC	6.9		AGG	1.6
	CCA	5.8	G	GGU	17.4
	CCG	24.7		GGC	35.3
T	ACU	6.7		GGA	8.7
	ACC	23.3		GGG	12.0
	ACA	5.8	*	UAA	1.9
	ACG	18.8		UAG	0.3
A	GCU	12.8		UGA	1.0
	GCC	29.1			
	GCA	13.0			
	GCG	42.5			

A.2. Appendix B

All code was written using Python 3.

A.2.1. Python code

A.2.1.1. Primer designer

The following code snippet generates primers for the backbone-PCR protocol 2.2.2 to introduce new N-termini in place of the natural secretion signal of sptP. This code was written in collaboration with J.Ahrendt.

```
from Bio.Seq import Seq
from Bio import SeqIO
from Bio.SeqRecord import SeqRecord
from Bio.SeqFeature import SeqFeature, FeatureLocation
from Bio.SeqUtils import GC
import primer3 as p3
import datetime


class PlasmidBuilder():


    def __init__(self,  codon_table = None):


        if codon_table == None:

            self.codon_usage_table =
                                                {'Salmonella␣
                                                    typhimurium␣LT2':
                                                                        {
                                                                            ,
                                                                            F
                                                                            ,
                                                                            :{
                                                                            '
                                                                            UUU
                                                                            '
                                                                            :23.3,

                                                        'UUC':15.3},
                                                'L':{'UUA':13.2,
                                                        'UUG':12.4,
                                                        'CUU':11.8,
                                                        'CUC':10.4,
                                                        'CUA':4.9,
```

```
        'CUG': 53.6},
    'I': {'AUU': 29.3,
          'AUC': 24.4,
          'AUA': 5.3},
    'M': {'AUG': 27.4},
    'V': {'GUU': 15.5,
          'GUC': 18.2,
          'GUA': 11.4,
          'GUG': 25.2},
    'S': {'UCU': 7.2,
          'UCC': 10.1,
          'UCA': 6.2,
          'UCG': 9.5,
          'AGU': 7.3,
          'AGC': 17.4},
    'P': {'CCU': 7.2,
          'CCC': 6.9,
          'CCA': 5.8,
          'CCG': 24.7},
    'T': {'ACU': 6.7,
          'ACC': 23.3,
          'ACA': 5.8,
          'ACG': 18.8},
    'A': {'GCU': 12.8,
          'GCC': 29.1,
          'GCA': 13.0,
          'GCG': 42.5},
    'Y': {'UAU': 17.1,
          'UAC': 11.6},
    'H': {'CAU': 13.3,
          'CAC': 9.6},
    'Q': {'CAA': 12.7,
          'CAG': 31.0},
    'N': {'AAU': 17.8,
          'AAC': 20.1},
    'K': {'AAA': 31.7,
          'CAA': 12.7,
          'AAG': 11.3},
    'D': {'GAU': 31.6,
          'GAC': 20.3},
    'E': {'GAA': 35.4,
          'GAG': 20.7},
    'C': {'UGU': 4.8,
          'UGC': 6.6},
    'W': {'UGG': 15.2},
    'R': {'CGU': 18.8,
          'CGC': 23.3,
          'CGA': 3.6,
          'CGG': 6.9,
```

```
                                        'AGA':2.3,
                                        'AGG':1.6},
                                    'G':{'GGU':17.4,
                                        'GGC':35.3,
                                        'GGA':8.7,
                                        'GGG':12.0},
                                    '*':{'UAA':1.9,
                                        'UAG':0.3,
                                        'UGA':1.0}
                                            }
                                        }
        else:
            self.codon_usage_table = codon_table



def get_codon_probabilities(self):
    '''
    Calculates the probabilities for a residue to be encoded by a
        certain codon.
    This calculation is done by summing up all occurrences of the
        codons per 1000 residues as given
    by the input codon usage table and subsequently calculating the
         percentual occurrence.

    E.g. In S.typhimurium LT2, phenylalanine occurrs 23.3/1000
        residues encoded as UUU and 15.3/1000 residues as UUC.
    Therefore, phenylalaline is encoded in 60.3% of occurrences by
        UUU.

    Input parameters:
    ----------

        codon_table : dict of dicts
            Dictionnary containing the codon frequency per 1000
                residues for the given organism.


    Returns:
    ----------
        codon_probability_dict : dict of dicts
            Dictionnary containing the percentual codon usage per
                residue for the fiven organism.

    '''


    codon_probability_dict = {}
```

```
        for strain in self.codon_usage_table.keys():
            res_codon_probability = []
            for residue in self.codon_usage_table[strain].keys():
                residue_codons = self.codon_usage_table[strain][residue
                    ]
                codon_no = len(residue_codons.values())
                total_res_usage = sum(residue_codons.values())
                codon_dict = {}
                for key, value in residue_codons.items():
                    probability = value/(total_res_usage)
                    codon_dict[key] = probability

                codon_probability_dict[residue] = codon_dict

        return codon_probability_dict




def protein_to_DNA_sequence(self, protein_sequence, most_common=
    True):
    '''

    Converts a protein sequence into a DNA sequence.
    If most_common = False (default=True), a weighted random choice
        based on the codon occurrence probability is used to
       create a DNA sequence. Otherwise,
    the most common codons for each residue will be used.

    Input parameters :
    ------------

    protein_sequence : str
        String of a peptide sequence
    most_common : bool (default = True)
        Whether the most common codon or a weighted random choice
           based on codon occurrence probabilities is used to
           generate the DNA sequence.


    Returns :
    ----------

    dna_sequence : str
        String of the encoding DNA sequence for a given protein/
           peptide.
```

```
        '''

        self.probabilities = self.get_codon_probabilities()

        if most_common == True:

            most_common_codon_sequence = ''

            for residue in protein_sequence:

                codons = list(self.probabilities[residue].keys())

                probability = list(self.probabilities[residue].values()
                    )

                most_common_codon = [codons[probability.index(max(
                    probability))]]

                most_common_codon_sequence += ''.join(most_common_codon
                    )


            self.insert_sequence = most_common_codon_sequence

            return most_common_codon_sequence


        elif most_common == False:

            shuffled_codon_sequence = ''

            for residue in protein_sequence:
                codon_choice =  random.choices(list(self.probabilities[
                    residue].keys()) ,list(self.probabilities[residue].
                    values()))
                shuffled_codon_sequence += ''.join(codon_choice)

            self.insert_sequence = shuffled_codon_sequence

            return shuffled_codon_sequence



class Primer():


    def __init__(self, amplification_sequence):

        self.amp_sequence = Seq(amplification_sequence)
```

```
        self.amp_sequence_rv = self.amp_sequence.complement()

        self.amp_sequence = str(self.amp_sequence)
        self.amp_sequence_rv = str(self.amp_sequence_rv)


    def calculate_primer_tm(self, sequence):

        melting_t = p3.calcTm(sequence)

        return melting_t




    def calculate_hairpin_formation(self, primer):

        hairpin = p3.calcHairpin(primer)
        hairpin_tm = p3.calcHairpinTm(primer)

        return hairpin, hairpin_tm



    def create_primer(self, fw=True, tm_range : list = [54,72],
        default_tm : int = 62, primer_length_range : list = [20,35]) ->
        str:
        '''
        Creates the forward primer for the sequence region to be
            amplified. When creating the primer, the optimal melting
            temperature is prioritized over optimal primer length.

        Input parameters :
        ----------
        fw : bool (default=True)
            Whether the forward or reverse primer is generated

        tm_range : list (default = [54, 72])
            Minimal and maximal acceptable melting temperature

        default_tm : int (default = 62)
            Optimal melting temperature.

        primer_length_range : list (default = [20,35])
            Minimal and maximal primer length


        Returns :
```

```
----------

[primer, primer_tm] : list
    Primer sequence fitting to the specified instructions (Tm,
        length) and corresponding Tm

'''

primer_length = min(primer_length_range)
primer_tm = 0

if fw == True:


    primer = self.amp_sequence[:primer_length]
    primer_tm = self.calculate_primer_tm(primer)
    rounded_tm = round(primer_tm)

    while primer_length in range(min(primer_length_range), max(
        primer_length_range)) and rounded_tm <= default_tm:

        primer = self.amp_sequence[:primer_length]
        primer_tm = self.calculate_primer_tm(primer)
        rounded_tm = round(primer_tm)

        primer_length += 1


        if (primer_tm < min(tm_range)) and (primer_length ==
            max(primer_length_range)):
            print('The␣longest␣possible␣oligo␣(length:␣{})␣
                created␣is␣outside␣of␣the␣specified␣melting␣
                temperature␣range␣([{}{}]).␣Please␣adapt␣either
                ␣tm_range␣or␣allowed␣oligo_length.'.format(max(
                primer_length_range), min(tm_range), max(
                tm_range)))



    while primer_length in range(min(primer_length_range), max(
        primer_length_range)) and primer_tm >= default_tm:

        primer = self.amp_sequence[:primer_length]
        primer_tm = self.calculate_primer_tm(primer)
        rounded_tm = round(primer_tm)
        primer_length -= 1
```

```
            if (primer_tm > max(tm_range)) and (primer_length ==
                min(primer_length_range)):
                print('The shortest possible oligo (length: {}) 
                    created is outside of the specified melting 
                    temperature range ([{}{}]). Please adapt either 
                    tm_range or allowed oligo_length.'.format(min(
                    primer_length_range), min(tm_range), max(
                    tm_range)))



        return [primer, primer_tm]


    elif fw == False:


        primer = self.amp_sequence[:primer_length]
        primer_tm = self.calculate_primer_tm(primer)
        rounded_tm = round(primer_tm)

        while (primer_length in range(min(primer_length_range), max
            (primer_length_range))) and (rounded_tm <= default_tm):

            primer = self.amp_sequence_rv[-primer_length:]
            primer_tm = self.calculate_primer_tm(primer)
            rounded_tm = round(primer_tm)
            primer_length += 1

            if (primer_tm < min(tm_range)) and (primer_length ==
                max(primer_length_range)):
                print('The longest possible oligo (length: {}) 
                    created is below the specified melting 
                    temperature range ([{} °C, {} °C]). Please 
                    adapt either tm_range or allowed oligo_length.'
                    .format(max(primer_length_range), min(tm_range)
                    , max(tm_range)))


        while (primer_length in range(min(primer_length_range), max
            (primer_length_range))) and (primer_tm >= default_tm):

            primer = self.amp_sequence_rv[-primer_length:]
            primer_tm = self.calculate_primer_tm(primer)
            rounded_tm = round(primer_tm)
            primer_length -= 1
```

```
                if (primer_tm > max(tm_range)) and (primer_length ==
                    min(primer_length_range)):
                     print('The shortest possible oligo (length: {}) 
                         created is above the specified melting 
                         temperature range ([{} °C, {} °C]).  Please 
                         adapt either tm_range or allowed oligo_length.'
                         .format(min(primer_length_range), min(tm_range)
                         , max(tm_range)))



            return [primer, primer_tm]



    def create_primer_pair(self, default_tm : int = 62,
        max_tm_difference : int = 3, max_length_difference : int = 8,
        create_overlaps = False):

        fw_primer = self.create_primer(fw=True, default_tm=default_tm)
        rv_primer = self.create_primer(fw=False, default_tm=default_tm)

        fw_len = len(fw_primer[0])
        fw_tm = int(fw_primer[1])

        rv_len = len(rv_primer[0])
        rv_tm = int(rv_primer[1])


        if (fw_tm in range(rv_tm - max_tm_difference, rv_tm +
            max_tm_difference)) and (fw_len in range(rv_len -
            max_length_difference, rv_len + max_length_difference)):
             return fw_primer, rv_primer
```

A.2.1.2. Dataset random subselection

```
from Bio import SeqIO
from Bio.SeqRecord import SeqRecord
from Bio.Alphabet import IUPAC
from Bio.Seq import Seq

import pandas as pd

''' Loading the EffectiveT3 negative training dataset. '''
''' ------------------------------------------------- '''
```

```
fasta_file = 'TTSS_negative_training.faa'

records = list(SeqIO.parse(fasta_file, "fasta"))

neg_seqlist = []

for record in records:
    seqentry = [record.id, record.name, str(record.seq), record.
        description]
    neg_seqlist.append(seqentry)

columns= ['Protein_ID', 'Name', 'Sequence', 'Description']
df_neg = pd.DataFrame(neg_seqlist, columns=columns)


''' Random subselection of entries from the negative training dataset.

        Sequences shorter than 60 residues are excluded.        '''


rand_list = []
neg_rand_list = []

for i in range(0, 348):

    while len(neg_rand_list) != 244:

        randint = np.random.randint(0,348)
        seq = df_neg.iloc[randint]

        if len(seq[2]) >= 60 and randint not in rand_list:
            neg_rand_list.append(seq)

        else:
            pass

''' Writing the resulting dataset to file. Both csv- and fasta-files
    are created.

        The naming of the resulting file follows this nomenclature:

        date_initials_EffectiveT3_negative_dataset_Dataset_(number)
            _status(Finished).csv/fasta

        (YYYYMMDD)_(MP)_EffectiveT3_negative_dataset_D_X_F    '''

columns= ['Protein_ID', 'Name', 'Sequence', 'Description']
df_neg_subsampled = pd.DataFrame(neg_rand_list, columns=columns)
```

```
df_neg_subsampled.to_csv('YYYYMMDD_MP_EffectiveT3_negative_dataset_DXF.
    csv')



df_neg_fasta_sequences = []

df_neg_fasta = df_neg_subsampled.reset_index()

for i in range(0, len(df_neg_fasta)):

    sequence = SeqRecord(Seq(df_neg_fasta['Sequence'][i], IUPAC.protein
        ),
         id=df_neg_fasta['Protein_ID'][i], name=df_neg_fasta['
             Description'][i],
         description=df_neg_fasta['Description'][i])
    df_neg_fasta_sequences.append(sequence)

SeqIO.write(df_neg_fasta_sequences, '
    YYYYMMDD_MP_EffectiveT3_negative_dataset_DXF.fasta', 'fasta')
```

A.3. Appendix C

A.3.1. T3SE database

Appendix C contains addditional information used for assembly of the T3SE database, as well as information about host cell functions, targets and localizations.

Table A.2.: Search terms and annotations used for collecting and cleaning dataset 1. A Search terms used for collecting T3SS-secreted proteins from the NCBI database for assembly of dataset 1 (subsection 3.2.1). The following key strings were used to select protein entries from the NCBI IPG database. **B** Protein entries retreived from the IPG database contain annotations that indicates the presence of non-T3SE sequences. All strings displayed in this table were isolated from fasta-files collected using the search terms in Table A.2A. Entries containing any of the tags were manually reviewed to determine inclusion or discarding the respective entry.

A	**B**
search term	**annotation tag**
T3SS secreted effector	fragment
T3SS effector protein	truncated
type 3 secreted effector	putative
type 3 effector protein	partial
type III effector	probable
type III secreted effector	uncharacterized
type III effector protein	hypothetical
type III secretion system effector	nonfunctional
type III secretion system effector protein	chaperone/co-chaperone
type 3 secretion system effector	candidate
type 3 secretion system effector protein	subunit
	unknown
	pseudogene
	unassigned
	flagellar
	unnamed
	crystallin

Table A.3.: Functions associated with T3SE-secreted proteins.

Uniprot_ID	gene_name	function
A4SUG8	acrH	secreted chaperone for AopB/AopD
A4SUG9	acrV	tip translocon protein, needle tip protein
Q93Q17	aexT	GTPase activity, actin ADP-ribosylation
D5LUP3	aexU	ADP-ribosylation, GAP activity towards Rac1, Cdc42 and RhoA, delays degradation of IkkB in host cells, reduces IL-6 and IL-8 secretion, disrupts the host cell skeleton
A4SUG7	aopB	translocon protein, needle tip protein
A4SUG6	aopD	translocon protein, needle tip protein
Q27RI1	aopH	putative tyrosine phosphatase activity, might induce cytoskeletal damage
A4SUH7	aopN	SctW family T3SS gatekeeper subunit, putatively involved in regulation of secretion
A0A1Z3MNJ8	aopO	putative serine/threonine kinase
Q14SK3	aopP	acetyltransferase, inhibits the NF-kappa-B signaling pathway, induces apoptosis
A4SUF6	ascF	needle subunit, early substrate
A4SUF4	ascH	regulator needle assembly
A4SUF3	ascI	needle component
A4SUI0	ascP	needle length control, secretion regulation
A4SUH4	ascX	needle subunit
A4SUE8	ati1	chaperone for Ati2
A4SUE7	ati2	inositol polyphosphatase 5-phosphatase, dephosphorylates PtdIns(4,5)P2 and PtdIns(3,4,5)P3
A4SUG4	exsE	negative transcriptional regulator
Q9REZ5	bopB	translocon protein, formation of the translocon pore
A0A6N3S484	bopD	translocon protein, formation of translocon pore
Q84CS9	bopN	putative gatekeeper subunit, induces IL-10 in host cells, involved in downregulation of MAPK kinase
Q84CT6	bscF	needle filament protein
A0A5P2MRE5	bsp22	needle tip protein, tip complex formation
A0A6N3S465	bspR/btrA	negatively regulates T3SS secretion, anti-sigma factor
A0A6N3S5J5	bteA	induces caspase-1 independent necrotic cell death, targets PI(4,5)P2, indirectly involved in dephosphorylation of tyrosine-phosphorylated proteins
A0A2U3QAH8	BRAD3257_7792	contains a ULP domain
A0A0K0WSG7	bel2-5	NF-independent symbiosis, nodulation restriction, contains a ubiquitin-like protease domain
A0A2U3QA97	ernA	required for nodulation, might regulate host gene expression
A0A562K8J2	gunA2	cellulase activity
A0A384VCK8	innB	regulates nodulation formation with different symbionts
A0A6F8NWU7	nopA	T3SS needle/pilus protein
SPP98367.1	nopAB	promotes nodulation
A0A2U3Q975	nopAO	increase in nodulation and nitrogenase activity
C4ALD1	nopC	influences nodule formation
A0A2U9K2V6	nopD	SUMO-protease

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
G7DET7	nopE1	modulates gene expression of plant hormones
G7DGM5	nopE2	modulates gene expression of plant hormones
A0A2U8GEX4	nopI	beneficial for nodule formation
P55704	nopL	modulates MAPK kinase pathways
P55456	nopM	NEL/E3 ubiquitin ligase activity
A0A2U3Q8X5	NopM,NopM1	putative novel E3 ubiquitin ligase
P55724	nopP	involved in nodule formation
A0A2U3Q8S9	nopP,nopP1	positively influences symbiosis
A0A2U3QAQ0	nopP,nopP2	cause of incompability between ORS3257 and V.radiata
P55730	nopT	cysteine autoprotease
Q89T99	blr2140,nopT1	cysteine protease, induces HR-like cell death
Q9AMW4	blr2058,nopT2	cysteine protease
P55711	nopX	translocon protein
Q63K38	BPSS1528,bapA	involved in early invasion stages, involved in adherence and cell-to-cell spread
Q63K40	bapC	putative transglycosylase
Q63K34	bipB	translocon protein, involved in MNGC formation, induces apoptosis, involved in phagosomal escape
Q63K35	bipC	translocon protein, F-actin binding capability
Q63K37	bipD	translocon protein
Q63K42	bopA	contains a Rho GTPase inactivation domain, contains a cholesterol-binding domain
Q63K41	bopE	guanine nucleotide exchange factor, induces activation of caspase-1 and caspase-7, induces cytoskeleton rearrangements
Q63K18	bsaL	needle protein
Q63KH5	cif	deamidase activity, papain-like hydrolytic activity, putative cysteine protease
Q9Z9F3	CP_0748	Inc protein, contains DUF648-domain
Q9Z9D2	CP_0725	Inc protein
Q9Z9B4	CP_0707	Inc protein
Q9Z949	CPn_0132	Inc protein
Q9Z937	CP_0627	Inc protein
A0A0F7WL04	CPn_0174	Inc protein
Q9Z903	CP_0587	Inc protein
Q9Z8X3	CP_0554	Inc protein
Q9Z8W9	CP_0550	Inc protein
Q9Z8W1	CP_0542	Inc protein
A0A0F7WDZ9	CPn_0354	contains a DUF1389 domain
A0A0F7WNQ2	CPn_0355	contains a DUF1389 domain
Q9Z8I6	CP_0401	contains a DUF1389 domain
Q9Z8A0	IncV;(CT005)	interacts with VAPs, promotes the formation of ER-inclusion MCS
Q9Z868	CPn_0483	deubiquitiniase activity, targets NDP52
Q9Z7W9	CP_0163	interacts with Rab GTPases 1,10 and 11
A0A0F7WRH6	CPn_1027	interacts with Caprin2 and GSK-beta, enhances anti-apoptopic activity

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
Q9Z8A1	CT006	interacts with VAMP3
O84046	glgX	glycogen hydrolase, glycogen degradation
O84089	malQ	4-alpha-glucanotransferase
O84091	lcrE/copN	putative secretion regulator
O84107	cteG	localizes to Golgi
P0DJI3	incD	binds to the ceramide transfer protein CERT
P0DJI4	incE	modulation of retromer-dependent trafficking, interacts with SNX5/6
P0DPS6	incG	phosphorylated by phosphoserine binding protein 14-3-3-beta, binds VAMP3
P0CI27	incA	induces homotypic fusion of inclusions, targets VAMP3/7/8
O84155	CT_153	mimics MACPF domain
O84226	ipaM	targets CEP170, modulation of the microtubule network, inhibition of host cell cytokinesis
O84229	CT_226	interaction with LRRF1 and FLII
O84231	CT_228	recruits MYPT1, regulates host cell egress
K0GA27	cpoS	binds to multiple Rab GTPases
Q9Z8P7	incB	interacts with host protein Snapin
O84236	incC	control of inclusion membrane stability
O84250	glgP	glycogen phosphorylase
O84251	CT_249	localizes to inclusion membranes
O84290	CT_288	interacts with centrosomal host protein CCDC146
O84297	mrsA_1	phosphoglucomutase
O84363	CT_358	localizes to the inclusion membrane
O84388	CT_383	localizes to the inclusion membrane
O84432	CT_425	immunogenic in humans
O84439	glyA	serine hydroxymethyltransferase
O84449	srp/crpA	stimulates CD8+ T cell response
O84462	tarP	targets actin,ABI1,VAV2,PI3K,SHC1
O84466	CT_460	localizes to the host cell nucleus, putatively involved in chromatin remodeling
O84582	copB	translocon protein, localizes to inclusion membrane
Q9Z798	copD	translocon protein
O84587	CT_583	target of human T cells, similar to virulence factor pGp6D
O84616	cadD	oxidoreductase, binds to death domains, binds to Fas,DR4 and DR5, induces apoptosis in host cells
O84624	CT_619	interacts with Tsg101 and Hrs (ESCRT-I complex)
A0A0F7X663	CT620	interacts with Tsg101
O84626	CT_621	interacts with Hrs (ESCRT-I complex)
O84627	CT_622	beneficial for efficient bacterial entry
O84673	CdsF	needle protein
O84700	tmeA	targets AHNAK and N-WASP, modulation of actin cytoskeleton
Q9Z754	CT_711	localizes to the host cell nucleus
O84718	CT_712	DUF582-domain protein

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
O84742	CT_737	histone methyltransferase, methylates host cell core histones H2B, H3 and H4
O84743	yycJ	metallo-hydrolase
O84804	glgA	glycogen synthase, synthesizes alpha-1,4-glucan
O84854	CT_847	targets host cell GCIP, modulation of host cell growth
Q822C7	CT848	contains a DUF720-domain
O84856	CT_849	contains a DUF720-domain
O84858	CT_850	targets DYNLT1
A0A0H3MBG9	copD2	translocon protein
K0G8L2	copB2	translocon protein
O84874	glgB	glycogen branching
O84883	tepP	interacts with host scaffolding proteins Crk-1 and Crk-2
Q7NUT1	cipA	presumably F-actin binding
Q7NUS8	cipB	translocon protein, pore formation in host cell membrane
Q7NUS9	cipC	translocon protein
Q7NUT0	cipD	translocon protein
Q7NWF2	copC	ADP-riboaxanation, inhibits host cell apoptosis, necroptosis and pyroptosis, preferably modifies apoptotic caspases
Q7P1B7	copE	guanine exchange factor, activates Rac1 and Cdc42, induces actin rearrangement
Q7NZE8	copH	propable tyrosine-phosphatase
A0A2T5NVZ6	cprI	T3SS needle protein
Q7NY09	cteC	ADP-ribosyltransferase, modifies ubiquitin on residue T66, disrupts synthesis of polyubiquitin
Q7NYG6	spvC/virA	phosphothreonine lyase, most likely targets MAPKs
D2TKD9	escF	needle protein
D2TKE3	espA	needle filament protein
D2TKE1	espB	translocon protein
D2TKE2	espD	translocon protein
D2TKD7	espF	induces actin polymerization
D2TKD5	espG	induces microtubule depolymerization, blocks WRC-dependent actin-polymerization
D2TKF1	espH	putatively inactivates Rho-GTPases, induces cytoskeletal disruption
D2TJZ4	espI	disrupts COPII cargo packaging and protein trafficking from golgi to ER, targets PDZ-domain proteins
D2TRY1	espJ	ADP-ribosyltransferase activity, inhibits Csk kinase
D2TT36	espL2	cysteine protease, inhibits necroptosis and nf-kappa-B-signalling, induces actin aggregation/bundling, targets RHIM-domain proteins
D2TI21	espM2	activates RhoA, triggers phosphorylation of cofilin (ROCK-LIMK-cofilin pathway), triggers formation of actin stress fibers
D2TM85	espM3	activates RhoA, triggers phosphorylation of cofilin (ROCK-LIMK-cofilin pathway), triggers formation of actin stress fibers
D2TI15	espT	induces formation of lamellipodia and membrane ruffling, activation of Rac1 and Cdc42, requires Wave2 and Abi1
D2TMV7	espV	induces cell rounding

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
D2TTX7	espX7	HECT-like E3 ubiquitin ligase
D2TKF8	espZ	promotes host cell survival, upregulates phosphorylation of AKT and focal adhesion kinase (FAK) via CD98
D2TKE9	map	disrupts mitochondrial membrane potential, involved in formation of actin-rich filopodia, guanine-nucleotide exchange factor for Cdc42, activates ERK an p38
A0A482PDI9	nleB	N-acetylglucosamine-transferase, disrupts pro-inflammatory NF-kappa-B signalling, protection from oxidative stress, targets death-domain proteins
D2TK70	nleC	zinc protease, disrupts the NF-kappa-B inflammatory pathway
D2TML3	nleD-1	zinc protease, cleaves MAPKs, disrupts the NF-kappa-B inflammatory pathway
D2TT38	nleE	SAM-dependent cystein methyltransferase, targets ubiquitin-chain binding activity of TAB2 and TAB3, methylates cysteine in NpI4-zinc-finger-domain, inhibits phosphorylation of NF-kappa-B inhibitor I-kappa-B
D2TRX8	nleF	caspase inhibitor, dampens apoptosis and inflammasome activation, disrupts intracellular protein trafficking
D2TK72	nleG1	E3-ubiquitin ligase, targets many mitochondrial and some nucleus proteins, capable of autoubiquitination
D2TRY0	nleG7	E3 ubiquitin ligase
D2TI20	nleG8	E3 ubiquitin ligase, contains a PDZ-domain binding motif
D2TRX7	nleH	promotes colonization, inhibits phosporylation of RPS3, disrupts NF-kappa-B pathway
A0A7W3BQ23	sctI	T3SS inner rod subunit
D2TKE8	tir	activates the N-WASP-Arp2/3 actin assembly, induces actin pedestal formation
B7UJQ8	bolA	transcription factor, negatively regulates flagella transcription, induces biofilm formation
P0DUW5	cif	glutamine deamidase, abolishes activity of CRL complexes by deamidation of Q40 of host NEDD8
B7UM94	espA	T3SS needle filament protein
Q05129	espB	translocon protein, recruits alpha-catenin to the adherence site, inhibits interaction of myosin and actin
B7UM93	espD	translocon protein, involved in pore formation
Q7DB85	espF	induces N-WASP mediated actin nucleation, slows epithelial cell turnover, modifies tight junctions
B7UMC8	espG	interferes with the WAVE regulatory complex, blocks ARNO signaling, triggers actin stress fibre formation, disrupts host cell microtubules, GAP activity towards Rab1
B7UH72	espG2	disrupts host cell microtubules, alters epithelial paracellular permeability, activates RhoA in MDCK monolayer cells, GAP activity towards Rab1
B7UMA2	espH	inactivates RhoGEFs, perturbs epithelial desmosomal junctions, involved in depletion of esmoglein-2

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
Q8XAJ5	nleA	inhibits caspase-1 activation, inhibits function of the COPII complex, prevents deubiquitination of NLRP3, inhibits secretion of IL-1-beta, binds to PDZ-domain proteins
A0A6M7GZB8	espJ	ADP-ribosyltransferase activity, amidates and ADP-ribosylates kinase Src on on residue E310, suppresses phagocytosis by inhibition of SFK-mediated phosphorylation of Fc-gamma-RIIa, ADP-ribosylates multiple kinases
A0A0H3JP21	espL2	cysteine protease activity, enhances activity of host cell annexin A2, increases bundling activity of F-actin, cleaves RHIM-domain containing proteins
Q8X4Q6	espM1	activates the RhoA pathway, represses formation of actin-pedestals, induces actin stress fibre formation
Q8X4W3	espM2	induces mislocalization of tight junctions, activates the RhoA pathway, represses formation of actin-pedestals
A0A0H3JFN8	espO1-1	inhibits STS-induced apoptosis, regulates IL-22 secretion, impacts neutrophil chemotaxis
Q5K5L9	espS	involved in actin pedestal modulation, interacts with IQGAP1
WP_001119657.1	espT	activates Rac1 and Cdc42, induces membrane ruffling
Q8X9A5	espW	triggers formation of membrane ruffles and flower-shaped structures, activates Rac1, targets the motor protein Kif15
A0A0H3JC80	espX2	SopA-like effector
A0A0H3JDV8	espX7	HECT-like E3 ubiquitin ligase, ubiquitylation of JNK kinases, disruption of the NF-kabba-B pathway
Q8XA11	espY1	putatively involved in apoptosis and cell cycle regulation
Q8XE85	espY3	induces elongation of actin pedestals
Q7DB68	espZ	promotes host cell survival, upregulates phosphorylation of AKT and focal adhesion kinase (FAK) via CD98
B7USU2	fliC	flagella filament protein, induces TLR5-dependent inflammatory response
B7UJB3	ivy	lysozyme C inhibitor
B7UI23	lifA	putative glycosyltransferase, putative cysteine protease, inhibits T-cell proliferation
B7UMA0	map	disrupts mitochondrial membrane potential, involved in formation of actin-rich filopodia, guanine-nucleotide exchange factor for Cdc42, activates ERK and p38, involved in hijacking of host endosomes
Q8XBX8	nleB1	N-actylglucosaminyltransferase activity, modifies death domains of target proteins FADD, TRADD, FAS, TNFR1, DR3 and RIPK1, intracellularly enhances GshB activity
Q8X837	nleB2	Glycosyltransferase activity, modifies RIPK1 and TNFR1, interacts with ensconsin
Q8X834	nleC	zinc metalloprotease, cleaves NF-kappa-B subunits, inhibits IL-8 secretion
A0A0H3JGR6	nleD	zinc metalloprotease, cleaves JNK and p38 kinases, inhibits secretion of proinflammatory cytokines

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
B7UI22	nleE	SAM-dependent cysteine methyltransferase, targets ubiquitin-chain binding activity of TAB2 and TAB3, methylates cysteine in NpI4-zinc-finger-domain, inhibits phosphorylation of NF-kappa-B inhibitor I-kappa-B p65/RelA
Q8XAL7	nleF	inhibits maturation and secretion of IL-18, binds and inhibits cleavage of several caspase, blocks FasL-induced cell death
B7UNX2	nleG/nleI	E3 ubiquitin ligase
ECs_1811	nleG2-1	E3 ubiquitin ligase
Q8X4X1	nleG2-2	E3 ubiquitin ligase
Q8X509	nleG2-3	E3 ubiquitin ligase, triggers degradation of hexokinase-2 and SNAP29, directly binds to hexokinase-2, autoubiquitinase activity
Q8X4X3	nleG5-1	E3 ubiquitin ligase, targets the MED15 subunit of the mediator complex, autoubiquitinase activity
Q8X507	nleG5-2	E3 ubiquitin ligase
Q8X4X2	nleG6-1	putative E3 ubiquitin ligase
A0A0H3JE38	nleG6-2	E3 ubiquitin ligase, autoubiquitinase activity
ECs_3488	nleG6-3	putative E3 ubiquitin ligase
B6DZZ5	nleG7	E3 ubiquitin ligase, recognizes PDZ-domains
Q8XAN6	nleG/nleG8-1	E3 ubiquitin ligase, binds to the cell cycle regulator CDC20, interrupts cel cycle progression
Q8X9A7	nleG8-2	E3 ubiquitin ligase, binds to the cell cycle regulator CDC20
Q8X831	nleH1-1/nleH1	protein kinase, autophosphorylase activity, inhibits phosphorylation of RPS3 by IKK-beta, attenuates nuclear localization of RPS3, inhibits caspase-dependent apoptosis, phosphorylates CRKL, suppresses activation of ERK1/2 and p38
Q8XAL6	nleH1_2/nleH2	protein kinase, autophosphorylase activity, inhibits caspase-dependent apoptosis, suppresses caspase-3 and p38
Q8X482	espF(U)/tccP	couples Tir to the host cell actin cytoskeleton, activates and binds to N-WASP
B7UM99	tir	induces host actin pedestal formation
Q7DB77	tir	induces actin pedestal formation, inhibits phosphorylation of TAK1
D4I1J6	avrRpt2	C70 cystein protease
O54581	dspE/dspA	induces necrosis, induces electrolyte leakage, interacts with host serine/threonine kinases
Q9LAW7	eop1	putative cystein protease, involved in host-specific virulence
A0A830ZZY8	eop2	contains a pectate lyase domain
E5B7T4	flgL	flagellar hook-associated protein
A2I5X7	eop3/hopX1	putative cysteine protease, elicits HR response in N.tabacum
Q46618	hrpA	forms the T3SS needle/pilus
D4HVM8	hrpJ	involved in regulation of harpin secretion, required for harpin secretion, elicits a HR response
D4HVL2	hrpK	putative translocator
Q01099	hrpN	forms pores, induces oxidate stress, elicits HR response, induces inhibition of ATP synthesis in chloroplasts, involved in host cell necrosis, promotes translocation of DspA

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
D4HVP9	hrpW	putative pectate lyase, contributes to callose accumulation, minor effect on pathogenicity
D4I2R4	traF	putatively involved in plasmid transfer and pilus formation
WP_196766517.1	hopAY1	putative cysteine protease
Q8VQ16	avrPphD/hopD1	involved in gall formation, elicits HR response
Q9FCZ8	hrpA	T3SS needle/pilus protein, forms the T3SS pilus
PEI06249.1	hrpK	involved in gall formation
Q9FCY8	hrpN	harpin, elicits HR response in host cells
Q9KH45	hrpN	harpin, elicits HR response in host cells
Q2LDQ5	hsvB	putative transcription factor, involved in host specificity determination
Q47867	hsvG	putative transcription factor, targets the consensus sequence ACACC/aAA, binds to the promoter region of host protein HsvgT
O85666	pthG	induces PIN2 expression, triggers HR response in host cells
Q9FCY7	wtsE	perturbs the phenylpropanoid metabolism in maize, induces cell death
Q84H14	C6H65_01640	serine protease, targets GTPase RhoA and Rac1
P13835	avrB	induces phosphorylation of RIN4 proteins, phosphorylates RIN4b
Q48B66	avrB2/avrB2-3	elicits HR response
Q5D157	avrB3/avrB4-1	triggers RPM1-dependent signaling, phosphorylates RIPK and RIN4
Q887C9	avrE1	involved in cell lysis and necrosis, linked to downregulation of NHL13
Q7BE94	avrRpm1	ADP-ribosyltransferase
Q48B92	avrRps4	triggers hypersensitive response in lettuce
Q6LAD6	avrRpt2	cysteine protease
Q9I1S4	exoY	nucleotidyl cyclase, binds to F-actin, induces increase of permeability between endothelial monolayers
Q886L1	hopAF1	targets methylthioadenosine nucleosidases MTN1 and MTN2
Q888W0	hopAI1	phospho-threonine lyase, dephosphorylates MPK3,MPK4 and MPK6
Q52430	hopAR1	cysteine protease
E5G0U3	hopAZ1	involved in host cell cytoskeleton modulation
Q888Y8	hopD1/avrD1	elicits HR response, involved in syringolide production, suppresses ETI response
Q87W42	hopG1	induces actin bundling, induces host chlorosis, interacts with mitochondrial kinesin KIN7D
Q8RP09	hopI1	induces chloroplast thylakoid remodeling, suppresses SA accumulation
Q88BH0	hopK1	triggers HR
Q4ZX47	hopZ3	acetyltransferase, disrupts the PTO defense pathway, targets RIN4-proteins, targets other effectors
Q52473	hrpA1	pilus-forming protein
Q87W38	hrpB	AraC-family transcriptional regulator
G3XDD1	hrpJ	controls secretion of translocon proteins

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
Q9F0B0	hrpZ/hrpZ1	harpin
G3XD49	pcrV	translocon protein
Q9I324	popB	translocon protein
Q9I323	popD	translocon protein
G3XCX6	popN	gatekeeper protein
P95434	pscF	T3SS needle component
Q7ALE9	hrpY	needle/pilus protein
Q8XPQ6	ripAZ1/rip71	induces cell death
RCFBP_11525	ripBM	putative serine/threonine kinase
Q8ZMI3	avrA	acetyltransferase, inhibits c-Jun, JNK, Ap1 and NF-kappa-B signalling, stabilizes tight junction protein ZO-1, deubiquitinates I-kappa-B-alpha and beta-catenin
A0A0F6B537	gogA	zinc metalloprotease, inhibits NF-kappa-B activation
Q8ZN18	gogB	targets the SCF ubiquitin ligase complex, dampens host inflammatory response
A0A0F6AZI6	gtgA	zinc metalloprotease, inhibits NF-kappa-B activation
A0A0H3N9Y3	gtgE	cysteine protease, proteolytically cleaves Rab29, Rab32 and Rab38
A0A0H3NF83	orgC	accelerates polymerization of PrgI, assists in needle filament assembly
A0A0F6AZQ0	pipA	zinc metalloprotease, inhibits NF-kappa-B activation
Q8ZMM8	pipB2	targets and activates kinesin-1
P41784	prgI	T3SS needle filament protein
P41785	prgJ	T3SS inner rod protein
A0A0K0H8V0	sboA	putative ubiquitin ligase
A0A0K0HD42	sboC/seoC	ADP-ribosylates Src and Csk kinases
A0A0K0HC32	sboH	prevents caspase-3 activation
A0A0K0H9B7	sboI	putative E3 ubiquitin transferase
Q56061	sifA	induces aggregation of LAMP-positive compartments, targets the BLOC-2 complex, interacts with Rab7
P0CL52	sipA	bundles F-actin, induces membrane ruffling, recruits syntaxin-8 to the SCVs, promotes fusion of SCVs with early endosomes
Q56019	sipB	translocon protein, forms a translocon pore with SipC, binds and activates caspase-1, mediates rapid pyroptosis and delayed apoptosis in macrophages, reduces p65 translocation into the nucleus
P0CL47	sipC	forms a translocon pore with SipB, binds,bundles and polymerizes F-actin, involved in PERP accumulation at the host membrane, recruits syntaxin-6
Q56026	sipD	needle tip protein, induces apoptosis via caspase-3 activation, reduces p65 translocation into the nucleus
Q8ZQQ2	slrP	E3 ubiquitin ligase, targets and ubiquitinates Trx1, targets the ER chapereone ERdj3
Q8ZNR3	sopA	HECT-like E3 ubiquitin ligase, disrupts tight junctions, involved in PMN migration

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
O30916	sopB	phosphoinositide phosphatase, dephosphorylates PI(4,5)P3, perturbs endosome to lysosome trafficking, alters membrange charge of SCVs via reduction of negatively charged lipids, acts as a guanine nucleotide dissociation inhibitor for Cdc42
P40722	sopD	can both activate and inactivate Rab8a, GDI activity towards Rab8a, GAP activity towards Rab8a
Q8ZQC8	sopD2	GAP activity, binds to annexin A2, targets Rab proteins
O52623	sopE	GEF activity, disrupts cell polarity, disrupts tight junctions, activates caspase-1, induces formation of lamellipodia and membrane ruffling
Q7CQD4	sopE2	GEF activity, induces a proinflammatory response, disrupts tight junctions, induces formation of filopodia, represses serine synthesis
Q8ZPY9	sopF	ADP-ribsyltransferase, inhibits autophagy, blocks association of ATG16L1 with the vacuolar ATPase subunit ATP6V0C via ADP-ribosylation, promotes the integrity of the SCV
P74873	sptP	tyrosine phospatase, GTPase-activating (GAP) domain, reverts cytoskeletal changes induced by SopE/SopE2, dephosphorylates the VCP/p97 AAA+ ATPase
P0A2M9	spvC	phosphothreonine lyase, dephosporylates ERK1/2, p38 and JNK kinases
P0A2N2	spvD	cystein protease, inhibits NF-kappa-B signaling, prevents nuclear accumulation of p65
A0A0H3NPQ1	srfJ	putative glycoside hydrolase
P0CZ04	spiC	inhibits endosome-endosome fusion, interferes with intracellular trafficking
A0A0H3NKW1	ssaG	needle filament protein
H9L496	ssaI	inner rod protein
Q7BVH7	sseB	translocon protein
O84947	sseC	translocon protein
Q9R803	sseD	translocon protein
H9L407	sseF	required for juxtanuclear positioning of the SCV, recruits dynein to the SCV, interferes with autophagosome formation, inhibits small GTPase Rab1A
H9L486	sseG	required for juxtanuclear positioning of the SCV, interferes with autophagosome formation, inhibits small GTPase Rab1A
A0A0F6AZL3	sseI	deamidase activiy, deamidates trimeric G-proteins, binds to IQ-GAP1, inhibits normal cell migration of macrophages and dendritic cells
Q9FD10	sseJ	deacylase, phospholipase A and GCAT activity, esterifies cholesterol to cholesterolesters, induces cholesterol accumulation
Q9L9J3	sseK1	glycosyltransferase, inhibits NF-kappa-B activation pathways
Q8ZNP4	sseK2	glycosyltransferase
P0DUJ7	sseK3	glycosyltransferase, inhibits NF-kappa-B activation
Q8ZNG2	sseL	deubiquitinase activity, alters host lipid metabolism
D0ZVG2	sspH1	E3 ubiquitin ligase, inhibits NF-kappa-B activation

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
P0CE12	sspH2	E3 ubiquitin ligase
Q8ZPD7	steA	interferes with the NF-kappa-B pathway, prevents dissociation of Cand-1 from cullin-1
Q8ZP57	steC	induces ROCK-like F-actin reorganization, kinase activity, phosphorylates multiple target proteins, controls intracellular replication
Q8ZNP2	steD	induces ubiquitination of MHCII, targets CD97 for degradation
A0A0F6B506	sarA	mimics cytokine receptor signaling (gp130), alters substrate specificity of GSK3, inhibits NF-kappa-B activation
Q8Z7T2	sboD/stoD	U-box E3 ubiquitin ligase
P18010	ipaA	induces F-actin depolymerization via binding to vinculin, harbors 3 vinculin-binding sites, activates vinculin
P18011	ipaB	translocon protein, involved in secretion control, induces cell cycle arrest, promotes unscheduled APC activation, induces apoptosis
P18012	ipaC	translocon protein, forms a complex with IpaB, nucleates actin, triggers src-dependent actin polymerization
P18013	ipaD	needle tip protein, essential for effector secretion, environmental sensor of bile salts, triggers loss of mitochondrial membrane potential, induces apoptotic cell death via caspase activation
A0A0H2UY03	ipaH_1	E3 ubiquitin ligase, inhibits PKC-mediated NF-kappa-B activation, targets TRAF2 for degradation
Q83R64	ipaH_4	putative E3 ubiquitin ligase
D2A6P4	ipaH4/H7, putative NEL E3 ubiquitin ligase	
A0A0H2V170	ipaH_7	NEL E3 ubiquitin ligase
A0A0H2USG1	ipaH1.4	NEL E3 ubiquitin ligase, inhibits LUBAC signalling
Q83RJ4	ipaH3	NEL E3 ubiquitin ligase
P18009	ipaH4.5	NEL E3 ubiquitin ligase, induces NLRP3-mediated inflammasome activation, inhibits the transcriptional activity of NF-kappa-B, disrupts MPR trafficking and lysosomal function, inhibits IRF3 signalling
P18014	ipaH7.8	NEL E3 ubiquitin ligase, induces inflammasome activation and pyroptosis, inhibits GSDMD-dependent pyroptosis
Q8VSC3	ipaH9.8	NEL E3 ubiquitin ligase, prevents Nod1-dependent NF-kappa-B activation, targets NEMO/IKBKG for ubiquitination, downregulates proinflammatory gene expression, modulates the acute innate imune response
Q54150	ipaJ	cysteine protease, cleaves N-myristoylated GTPases, inhibits protein trafficking from Golgi membranes
P33548	ipbB1	induces membrane ruffling, activates Rac1 via the ELMO-DOCK180 pathway, Rho GEF for Rac1 and Cdc42, involved in actin cocoon regulation, promotes efficient DMV escape
Q9AJW7	ipgB2	induces stress fibres, activates RhoA signalling, GEF for RhoA, promotes invasion into epithelial cell monolayers

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
Q07566	ipgD	inosotol phosphatase activity, dephosphorylates PtdIns(4,5)P2 into PtdIns(5)P, increases cocoon formation, promotes membrane blebbing and cell rounding, recruits Rab11 to the membrane, inhibits migration of activated CD4+ T cells
P33546	icsB	18-carbon fatty acyltransferase activity, binds cholesterol, represses early recruitment of LC3, inhibits autophagy, modifies a number of actin regulating proteins
Q04640	mxiC	negative regulator of T3SS secretion
P0A223	mxiH	secreted needle filament protein, induce NLRC4 inflammasome-mediated pyroptosis
P0A225	mxiI	inner rod protein, induce NLRC4 inflammasome-mediated pyroptosis
CAC05784.1	orf48	putative TA component
CAC05854.1	orf176	putative antitoxin component
A0A822PPP2	ERS574920_04342	cysteine protease, induces TOR inhibition hypersensitivy, downregulates IL-8 production, promotes PMN migration, increases host cell proliferation via activation of mTORC1, involved in activation of the MEK/ERK inflammatory pathway
Q8VSJ7	ospC1	ADP-riboxanation activity, prevents apoptosis, inhibits caspase-8 activation, promotes PMN transepithelial migration, inhibits IFN signaling, abolishes STAT1 phosphorylation
Q8VSL8	ospC2	ADP-riboxanation activity, required for inflammation, inhibits IFN-gamma 1 signaling
A0A0H2US87	ospC3	ADP-riboxanation activity, inhibits caspase-4 dependent inflammatory cell death
A0A822PRD6	ERS574920_04294	anti-activator role of virulence gene expression
Q6XW09	ospD2	putative cysteine protease, limits translocation of VirA, controls timing of host cell death
Q99Q01	ospD3	cysteine protease, prevents pyroptosis and necroptosis, inhibits IL-8 secretion, specifically targets RHIM-domains
D2AJY3	ospE1	inhibits apoptosis, binds to the integrin-linked kinase, interferes with focal adhesion disassembly, promotes bacterial adhesion to host cells
Q6BBS0	ospE2	binds to the integrin-linked kinase, interferes with focal adhesion disassembly, promotes bacterial adhesion to host cells
Q8VSP9	ospF	inactivates MAPK pathways, dephosphorylates ERK1/2, JNK and p38 kinases, required for efficient PMN migration, reprograms host cell gene expression
D2AJU3	ospG	serine/threonine kinase, inhibits the host inflammatory response, autophosphorylation activity, inhibits NF-kappa-B activation via inhibition of phospho-I-kappa-B-alpha ubiquitination, preferentially binds to ubiquitinated E2 enzymes
Q8VSD5	ospI	glutamine deamidase activity, deamidates UBC13, blocks TRAF6-dependent NF-kappa-B signaling

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
A0A3T2V133	ospZ	inactivates the TAK/Tab kinase complex, blocks nuclear translocation of p65, methylates TAB3
Q7BU69	virA	phosphothreonine lyase, interacts with alpha- and beta-tubulins, triggers necrosis, induces membrane ruffling, induces calpain activation, GAP activity towards multiple Rab proteins
O06662	vapC/MvpA	antitoxin, cleaves fMet-RNA
D0ZDK2	escE	regulates T3SS effector secretion
Q4G4C8	eseB	translocon protein, involved in autoaggregation and biofilm formation, builds filamentous appendages
D0ZDL0	eseC	transolcon protein
D0ZDK9	eseD	translocon protein
Q4G4D4	eseG	disassembles microtubules
D0Z825	eseH/eseN	phosphotreonine lyase activity, inhibits phosphorylation of ERK1/2, p38-alpha and JNK
Q9KI30	eseH	contains a E3 ubiquitin ligase domain and 5 LRR repeats
D0ZDM8	eseJ	negatively regulates adherence of bacteria, downregulates bacterial type 1 fimbriae, suppresses host cell apoptosis, disrupts endosomal maturation and lysosome fusion
C5B8F6	eseJ	contains a E3 ubiquitin domain and 12 LRR repeats
D0Z7K8	eseK	inhibits phosphorylation of p38-alpha, JNK and ERK1/2, inhibits expression of TNF-alpha, promoters bacterial colonization
WP_107775268.1	eseK	binds CD74, contains a E3 ubiquitin ligase domain and 12 LRR repeats
C5BD28	eseL	contains a E3 ubiquitin ligase domain and 6 LRR repeats
C5BD30	eseM	contains a E3 ubiquitin ligase domain and 11 LRR repeats
C5BA03	eseN	inactivates/dephosphorylates ERK1/2, phosphothreonine lyase domain, binds to th emajor vault protein
C5B8P0	eseO	contains a Shigella enterotoxin 2 domain and an ankyrin repreat domain
Q87GH0	vopA/vopP	acetyltransferase activity
Q87P61	vopB1	T3SS1 translocon protein
B9A7Z8	vopB2	T3SS2 translocon protein
Q87GF5	vopD2	T3SS2 translocon protein
Q87GF5	VPA1361	T3SS2 translocon protein
A0A6B3LEM4	vopE	modulates the CWI-MAPK pathway, GTPase-activating domain(GAP)
WP_000920496.1	vopK	putative acetyltransferase
Q87GI7	vopO	activates the RhoA-ROCK pathway, induces stress fibre formation
Q87P35	vopR	contributes to cell rounding, binds to PIP2
Q87GI9	vopT	ADP-ribosylase activity, targets Ras
Q87GF9	vopV	bundles F-actin
Q87GH1	vopW	hydrophilic translocon protein
Q87GH1	VPA1345	hydrophilic translocon protein
Q87GI0	vopZ	inhibits autophosphorylation of kinase TAK1, prevents activation of MAPK- and NF-kappa-B signaling

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
Q3C000	avrBs1	elicits HR response in plants, suppresses activation of the HOG MAPK pathway, induces cell enlargement and ion leakage
Q3BZN0	avrBs2	contains a glycelophodiesterase domain, involved in suppression of the PTI immunity answer
P14727	avrBs3	TAL effector, induces chlorosis, induces cell hypertrophy via regulation of cell enlargement, induces ion leakage, binds upa-motif containing DNA
Q07061	avrBs4/avrBs3-2	TAL effector, activates expression of Bs4C in host plants, induces chlorosis, induces catalase accummulation in peroxisomes, involved in suppression of plant defense responses
G0T341	avrBst/xopJ2	elicits HR response in host plants, acetyltransferase activity, acetylates ACIP1, binds to CaSGT1, reduces phosphorylation of SGT1
Q8P4H6	avrXccB	contains an acetyltransferase domain
Q4UWF4	xopAC	uridylates PBL2 and BIK1
P69979	yscM/lcrQ	negatively regulates T3SS effector secretion, binds LcrH
P23994	lcrV	involved in pore formation, regulates expression of YopB and YopD, induces apoptosis of human T-cells, upregulates IL-10, represses TNF-alpha signalling, binds TLR2 and receptor-bound hIFN-gamma
Q8D1P5	yipA	forms a complex with YitA,YitB,YitC and YipB
A0A0H2W9Z2	tccC2;yipB	forms a complex with YitA,YitB,YitC and YipA
Q8D1P8	yitA	forms a complex with YitB,YitC,YipA and YipB
A0A2U2H2J5	yitB	forms a complex with YitA,YitC,YipA and YipB
Q8D1P6	yitC	forms a complex with YitA,YitB,YipA and YipB
A0A3N4B6U4	ylrA	contains E3 ubiquitin ligase domain and an LRR-domain, inhibits growth when expressed in yeast
A0A3G5L8K9	ylrB	contains an LRR domain
A0A380PKG6	ylrC	contains an E3 ubiquitin ligase domain and an LRR-domain, inhibits growth when expressed in yeast
Q06114	yopB	translocon protein, pore forming protein, triggers Ras-dependent activation of NF-kappa-B
P37132	yopD	translocon protein, pore forming protein, negatively regulate yop translation prior to secretion
P31492	yopE	induces disruption of cytoskeleton, catalyzes GTP hydrolysis to inhibit Rho GTPase activity, induces inflammasome activation, prevents ROS formation
P08538	yopH	tyrosine phosphatase, interrupts activating signals for GEFs
O68718	yopJ	cysteine protease activity, acetyltransferase activity, inhibits caspase-1 dependent response, inhibits NF-kappa-B and JNK signaling
P17778	yopM	inhibits pyrin inflammasome activation, phosphorylates 14-3-3 binding sites of PRK kinases
Q663K1	yopN	blocks secretion of Yops prior to host cell contact, involved in virulence

Functions associated with T3SE-secreted proteins. (continued)

Uniprot_ID	gene_name	function
Q05608	ypkA/yopO	inhibits multiple G-alpha-1 signalling pathways, binds RhoA and Rac1, binds to actin, recruits and phosphorylates actin polymerization regulators
P27474	yopQ/yopK	inhibits bacterial adherence to the host cell, prevents inflammasome activation, interacts with T3SS translocon pore, controls effector secretion, prevents caspase-1 dependent pyroptosis
P68590	yopR, involved in needle filament polymerization, contributes to virulence	
Q93RN4	yopT	induces cytoskeleton disruption, triggers dephosphorylation of pyrin, cleaves the C-terminus of Rho GTPases
O85477	yplA	phospholipase PLA2 activity, induces acute inflammation
Q01247	yscF	needle filament protein, induces proinflammatory cytokines
P69971	yscI	inner rod protein
P68587	yscP	needle length regulator, substrate specifity switch regulator
P69986	yscU	dissocation of YscUcc induces Yop secretion upon Ca2+-depletion, substrate switch regulator, C-terminal secretion signal
P61416	yscX	required for Yop secretion, builds a ternary complex with YscY and YscV
Q8KQ84	yspB	translocon protein
A1JQ85	yspC	translocon protein
A0A8B6KWG7	yspD	translocon protein
A1JI19	yspE	contains an ADP-ribosyltransferase domain
A1JRY1	yspI	directly binds to FAK, inhibits cellular migration, induces cell paralysis
A1JRY5	yspK	serine/threonine kinase, interacts with E2 proteins
A0A485DGC3	sseJ/yspM	contains a GDSL lipase domain
A1JTB0	yspP	phosphatase activity
Q6RK53	dspE	elicits HR response in host cells, induces host cell death
C4ALD0	nopA	major T3SS needle pilus subunit
M4PUR5	gunA	endo-glycoside hydrolase/cellulase activity

Table A.5.: Identified target proteins of T3SS-secreted proteins This table contains all proteins that have been experimentally validated to be directly interacting, binding or modified by T3SS-secreted proteins.

target protein	effector (Uniprot/NCBI ID)	effector gene name
UBE2L3	D2AJU3	ospG
14-3-3-beta	P0DPS6	incG
14-3-3-gamma	P0CE12	sspH2

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
1700055N04Rik	A0A0F6AZI6	gtgA
ABCF2	Q7DB85	espF
ABI1	O84462	tarP/CT
ACADM	D2TK72	nleG1
ACBD3	H9L407, H9L486	sseF, sseG
ACIP1	G0T341	avrBst/xopJ2
AHNAK	O84700	tmeA
AIP	P0CE12	sspH2
ANXA6	Q7DB85	espF
ARF1	Q54150	ipaJ
ARF3	Q54150	ipaJ
ARF4	Q54150	ipaJ
ARF5	Q54150	ipaJ
ARL1	Q54150	ipaJ
ARL4C	Q54150	ipaJ
ATG1a/AT3G61960	Q48B92	avrRps4
ATP6V0C	Q8ZPY9	sopF
Abhd12	A0A0F6AZI6	gtgA
Abl	D2TKE8, B7UM99	tir, tir
AcsI4	A0A0F6AZI6	gtgA
Aldh3b1	A0A0F6AZI6	gtgA
AnxA2	Q8ZMM8, Q8ZQC8	pipB2, sopD2
AopB	A4SUG8, A4SUG6	acrH, aopD
AopD	A4SUG8, A4SUG7	acrH, aopB
Arf1	B7UMC8	espG
Arf5	B7UMC8	espG
Arf6	B7UMC8	espG
Arg	D2TKE8, B7UM99	tir
AtMPK6	P0A2M9	spvC
AtRIN4	Q5D157	avrB4-1
AtRIPK	Q5D157	avrB4-1
Ati2	A4SUE8	ati1
Atp5o	A0A0F6AZI6	gtgA
AvrB3	Q4ZX47	hopZ3
AvrPto1	Q4ZX47	hopZ3
AvrRpm1	Q4ZX47	hopZ3
B4gaint1	A0A0F6AZI6	gtgA
BAG regulator 2	P0CE12	sspH2
BASP1	P33546	icsB
BI1	Q8X831, Q8XAL6	nleH1-1/nleH1, nleH1_2/nleH2
BIK1	Q52430, Q4UWF4	avrPph3, xopAC
BopB	A0A6N3S484	bopD
BopD	Q9REZ5, A0A5P2MRE5	bopB, bsp22

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
Btk	A0A6M7GZB8	espJ
BtrS	A0A6N3S465	bspR
Bub3	P0CE12	sspH2
C8orf71	Q8X834	nleC
CAPN3	Q8XA11	espY1
CCDC146	O84290	CT_288
CD44	P18011	ipaB
CD74	WP_107775268.1	eseK
CD97	Q8ZNP2	STM2139/steD
CD98	D2TKF8, Q7DB68	espZ, espZ
CDC20	Q8XAN6, Q8X9A7	nleG/nleG8-1, nleG8-2
CDKN2AIPNL	Q8XA11	espY1
CENHP	A0A6M7GZB8	espJ
CEP170	O84226	ipaM
CERT	P0DJI3	incD
CHMP5	P33546	icsB
CLK1	Q8XA11	espY1
CNP	P33546	icsB
CRIK	Q9AJW7	ipgB2
CRKL	Q8X831	nleH1-1/nleH1
CTDSPL2	Q8X834	nleC
CaADC1	G0T341	xopJ2
CaALDH1	G0T341	xopJ2
CaHSP70A	G0T341	xopJ2
CaSGT1	G0T341	xopJ2
Cdc42	Q93Q17, D5LUP3, Q63K41, Q7P1B7, D2TKE9, B7UMA0, O30916, O52623, Q7CQD4, P74873, P33548, Q9AJW7, P31492, Q93RN4	aexT, aexU, bopE, copE, map, sopB, sopE, sopE2, sptP, ipbB1, ipgB2, yopE, yopT
CdsF	O84582, Q9Z798	copB, copD
CdsN	Q9Z798	copD
CopD	O84673	CdsF
CopN	Q9Z798	copD
Crk1	O84883	tepP
Crk2	O84883	tepP
Csk	D2TRY1, A0A6M7GZB8, A0A0K0HD42	espJ, espJ, sboC,seoC
Cugbp2	D0ZDK2	escE
Cyc1	A0A0F6AZI6	gtgA
DHFR	Q8XAL7	nleF

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
DIPM1	O54581	dspE/dspA
DIPM2	O54581	dspE/dspA
DIPM3	O54581	dspE/dspA
DIPM4	O54581	dspE/dspA
DNA	O06662	vapC/MvpA
DNA((TC)(TC)GCCAG(ACT))	B7UJQ8	bolA
DNA(TATAATTAATAATC-CACTT)	Q07061	avrBs4
DNA(TATATAAAC-CTNNCCCTCT)	P14727	avrBs3
DNA/ACACCaAA	Q47867	hsvG
DNAJC14	Q8XA11	espY1
DR3	A0A482PDI9	nleB
DR4	O84616	cadD
DR5	O84616	cadD
DRG2	Q8XBX8	nleB1
DSCR4	Q8XAJ5	nleA
DYNLT1	O84858	CT_850
Dbl/MCF2	B7UMA2	espH
ELMO1	P33548	ipbB1
ELMO2	P33548	ipbB1
ERI3	Q8XAL7	nleF
ERK1/2	C5BA03	eseN
ERK2	A0A0F6AZL3	sseI
ERdj3	Q8ZQQ2	slrP
EVL	Q05608	ypkA/yopO
Epb50/NHERF1	B7UMA0	map
EscF	B7UM94	espA
EseB	Q4G4C8, D0ZDK9	eseB, eseD
EseB,EseD	D0ZDL0	eseC
EseC	Q4G4C8, D0ZDK9	eseB, eseD
EseD	Q4G4C8	eseB
EspA	B7UM93	espD
EspD	B7UM94, Q05129, B7UM93	espA, eaeB/espB, espD
Etk	D2TKE8	tir
Etk/BMX	B7UM99	tir
ExsC	A4SUG4	exsE
F-actin	Q63K35, O84462, Q7NUT1, Q9I1S4, P0CL52, P0CL47, Q87GF9, Q05608	bipC, tarP, cipA, exoY, sipA, sipC, vopV, ypkA/yopO
FADD	A0A482PDI9, Q8XBX8, Q8ZNP4	nleB, nleB1, sseK2

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
FAK	P08538, A1JRY1	yopH, yspI
FAS	Q8XBX8	nleB1
FEM1B	A0A0H3JFN8	espO1-1
FLII	O84229	CT_226
FMNL1	Q8ZP57	steC
FMNL2	Q8ZP57	steC
FMNL3	Q8ZP57	steC
FRMD3	Q8XAJ5	nleA
Fas	O84616	cadD
Fyb	P08538	yopH
G-actin	P18012, Q05608	ipaC, ypkA/yopO
G-alpha-i2/Gnai2	A0A0F6AZL3	sseI
G-alpha-i3/Gnai3	A0A0F6AZL3	sseI
G-alpha-q	Q05608	ypkA/yopO
GAPDH	A0A482PDI9, D2TK72, Q8XBX8, Q9L9J3	nleB, nleG1, nleB1, sseK1
GCIP	O84854	CT_847
GEF-H1	Q87GI7	vopO/traA
GLoB	Q9L9J3	sseK1
GLoC	Q9L9J3	sseK1
GM130	B7UMC8, B7UH72	espG, espG2
GNG12	P33546	icsB
GSDMB	P18014	ipaH7.8
GSDMD	P18014	ipaH7.8
GSK-beta	A0A0F7WRH6	CPn_1027
GSK3-alpha	A0A0F6B506	sarA
GSK3-beta	A0A0F6B506	sarA
Gab1	P08538	yopH
Gab2	P08538	yopH
Gaint7	A0A0F6AZI6	gtgA
Glc1P	O84297	mrsA_1
GloA	Q9L9J3	sseK1
Gm10250	A0A0F6AZI6	gtgA
Gpnmb	A0A0F6AZI6	gtgA
GshB	A0A482PDI9, Q8XBX8	nleB, nleB1
H-Ras	P33546	icsB
HAX-1	A0A0H3JFN8, D2AJY3	espO1-1, ospE1
HIF-1-alpha	Q8XBX8	nleB1
HK2	Q8X509	nleG2-3
HMGN2	Q8XAL7	nleF
HNRNPM	D2TRA0	nleK
HOIL-1L	A0A0H2USG1	ipaH1.4
HOIP	A0A0H2USG1	ipaH1.4
HP1-gamma/CBX3	Q8VSP9	ospF

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
HPCAL1	Q7DB77	tir
HPS3	Q56061	sifA
HPS5	Q56061	sifA
HSP27	Q8ZP57	steC
HSPD1	D2TI15	espT
Hadha	A0A0F6AZI6	gtgA
Hck	A0A6M7GZB8	espJ
HprW	D4HVM8	hrpJ
HrpJ	Q01099, D4HVP9	hrpN, hrpW
HrpN	D4HVM8	hrpJ
Hrs	O84624, O84626, Q9Z754, O84718	CT_619, CT_621, CT_711, CT_712
Hsd17b12	A0A0F6AZI6	gtgA
Hsp70	Q8RP09	hopI1/hopPmaI
IFT20	A0A6M7GZB8	espJ
IKBA/NFKBIA	Q8ZMI3, Q8ZNG2, O68718	avrA, sseL, yopJ
IKK-alpha	A0A0H3JDV8	espX7,nleL
IKK-beta	A0A0H3JDV8, Q8X831, Q8XAL6	espX7,nleL, nleH1-1/nleH1, nleH1_2/nleH2
ILK	A0A0H3JFN8, D2AJY3	espO1-1, ospE1
INF2	Q05608	ypkA/yopO
IQGAP1	Q5K5L9, A0A0F6AZL3, A0A822PPP2, P17778	espS/ibE, sseI, ERS574920_04342, yopM
IRSp53	D2TKE8	tir
IRSp53/BAIAP2	Q8X482, B7UM99, Q7DB77	espF(U)/tccP, tir, tir
IRTKS	D2TKE8	tir
IRTKS/BAIAP2L1	Q8X482, B7UM99, Q7DB77	espF(U)/tccP, tir, tir
IcsA/VirG	P33546	icsB
IpaA	Q07566	ipgD
IpaB	P18012, P18013	ipaC, ipaD
IpaC	P18011	ipaB
Iqgap1	A0A0F6AZI6	gtgA
JNK	Q8VSP9	ospF
JNK1/MAPK8	A0A0H3JDV8, A0A0H3JGR6	espX7,nleL, nleD
JNK2/MAPK9	D2TML3, A0A0H3JDV8, A0A0H3JGR6	nleD-1, espX7, nleL, nleD
JNK3/MAPK10	A0A0H3JDV8	espX7, nleL
K-Ras	P33546	icsB
KIN7D	Q87W42	hopG1

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
Kif15	Q8X9A5	espW
LARG/ARHGEF12	B7UMA2	espH
LAT	P08538	yopH
LDHB	D2TJZ4	espI
LM04	Q8XAL7	nleF
LRRC18	Q8XBX8	nleB1
LRRF1	O84229	CT_226
LcrG	P23994	lcrV
LcrH	P69979	yscM/lcrQ
LcrQ	P37132	yopD
LcrV	Q06114, P37132	yopB, yopD
Lrrc59	A0A0F6AZI6	gtgA
Lyn	A0A6M7GZB8	espJ
MAD2B/Mad2L2	P18011	ipaB
MAD2L2	Q7DB85	espF
MAGI-2	Q8XAJ5	nleA
MAGI-3	Q8XAJ5	nleA
MALS3	Q8XAJ5	nleA
MAP2K4	O68718	yopJ
MAP2K7	O68718	yopJ
MAP7	D2TT36	espL2
MAPK1/ERK2	Q888W0, Q8VSP9	hopAI1, ospF
MAPK11	Q8VSP9	ospF
MAPK12	Q8VSP9	ospF
MAPK13	Q8VSP9	ospF
MAPK14	Q8VSP9	ospF
MAPK2/ERK1	P0A2M9	spvC
MAPK3/ERK1	P0A2M9, Q8VSP9	spvC, ospF
MARCH8	Q8ZNP2	STM2139/steD
MARCKS	P33546	icsB
MARCKSL1	P33546	icsB
MATN2	P27474	yopQ/yopK
MC7	D2TK72	nleG1
MED15	Q8X4X3	nleG5-1
MEK1	Q8ZP57	steC
METTL2A	A0A0H3JGR6	nleD
MHCII	Q8ZNP2	STM2139/steD
MKK4	Q8ZMI3	avrA
MKK7	Q8ZMI3	avrA
MPK3	Q888W0	hopAI1
MPK4	Q888W0	hopAI1
MPK6	Q888W0	hopAI1
MRFAP1L1	A0A6M7GZB8	espJ
MTN1	Q886L1	hopAF1

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
MTN2	Q886L1	hopAF1
MVP	C5BA03	eseN
MYPT1	O84231	CT_228
Mena	Q8X482	espF(U)/tccP
Mogs	A0A0F6AZI6	gtgA
MxiA	Q04640	mxiC
MxiC	P0A225	mxiI
MxiE	A0A822PRD6	ERS574920_04294
MxiH	P18013	ipaD
MxiI	Q04640	mxiC
N-WASP	O84700, D2TKD7, Q7DB85, Q8X482	tmeA, espF, espF, espF(U)/tccP
NCALD	Q7DB77	tir
NDP52	Q9Z868	CPn_0483/ChlaOTU
NEDD8	Q63KH5, P0DUW5	cif, cif
NEMO	Q8VSC3	ipaH9.8
NHERF1	Q8XAJ5	nleA
NHERF2	Q8XAJ5, Q8X831	nleA, nleH1-1/nleH1
NLRP3	Q8XAJ5, P18009	nleA, ipaH4.5
NOD1	P0CE12	sspH2
NPC1	Q9FD10	sseJ
NTL9	Q888Y8	hopD1
Nck1	D2TKE8, B7UM99	tir, tir
Nck2	D2TKE8, B7UM99	tir, tir
NleH1	Q8XAL6	nleH1_2/nleH2
NleH2	Q8X831	nleH1-1/nleH1
NopB	C4ALD0	nopA
OSPB	Q8ZNG2	sseL
OSPB1	Q9FD10	sseJ
OrgC	P41784	prgI
P(3,5)P2	Q8ZPY9	sopF
PAK1	B7UMC8	espG
PAK2	B7UMC8	espG
PBL1	Q52430	hopAR1
PBL11	Q52430	hopAR1
PBL2	Q52430, Q4UWF4	hopAR1, xopAC/AvrAC
PBL3	Q52430	hopAR1
PBL5	Q52430	hopAR1
PBL7	Q52430	hopAR1
PBL9	Q52430	hopAR1
PBS1	Q52430	hopAR1
PCID2	Q8XA11	espY1
PDE6D	Q7DB77	tir
PDLIM7	D2AJY3, Q6BBS0	ospE1, ospE2

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
PDZ-GEF/ARHGEF11	B7UMA2	espH
Epb50/NHERF1	D2TKE9	map
syntrophin	D2TJZ4	espI
MALS3	D2TJZ4	espI
PDZK1	D2TJZ4	espI
PDZK2	D2TJZ4	espI
PDZK11	D2TJZ4	espI
SNX27	D2TJZ4	espI
MAGI-2	D2TJZ4	espI
MAGI-3	D2TJZ4	espI
PSD-95	D2TJZ4	espI
SAP97	D2TJZ4	espI
SAP102	D2TJZ4	espI
Sec23A	D2TJZ4	espI
Sec24B	D2TJZ4	espI
Sec24C	D2TJZ4	espI
SLC3A2	D2TJZ4	espI
PDZ8	Q8ZQ59	pipB
PDZK1	Q8XAJ5	nleA
PDZK11	Q8XAJ5	nleA
PDZK2	Q8XAJ5	nleA
PENK	Q8XAJ5	nleA
PERP	P0CL52	sipA
PGGT-1	Q56061	sifA
PHGDH	D2TJZ4, D2TK72	espI, nleG1
PI(3)P	Q8ZPY9	sopF
PI(3,4,5)P3	Q8ZPY9	sopF
PI(4)P	Q8ZPD7	steA
PI(4,5)P2	Q07566	ipgD
PI(4,5)P3	O30916	sopB
PI3K	O84462, B7UM99	tarP, tir
PI3K/VPS34	D2TKE8	tir
PIH1D1	Q8XA11	espY1
PIP2	Q87P35	vopR
PKM	D2TK72	nleG1
PKN1	D0ZVG2	sspH1
PLC-gamma-2	P08538	yopH
PLEKHM1	Q56061	sifA
POLR2E	Q8XBX8	nleB1
PRAM-1	P08538	yopH
PRK2	P17778	yopM
PRKDC	D2TK72	nleG1
PSD-95	Q8XAJ5	nleA
PSMC1	Q8XA11	espY1

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
PSMD10	B7UI22	nleE
PTP4A1	Q8XAJ5	nleA
PipB2	Q56061	sifA
PrgI	Q56026	sipD
Prpf31	A0A0F6B537	gogA
PtdIns(3,4,5)P3	A4SUE7	ati2
PtdIns(4,5)P2	A4SUE7	ati2
PtdIns(4,5)P2/PIP2	A0A6N3S5J5	bteA
Ptges2	A0A0F6AZI6	gtgA
Pyk	P08538	yopH
RACK1	P27474	yopQ/yopK
RB1	Q8VSP9	ospF
RBCK1	Q05129	eaeB/espB
RHoG	P33546	icsB
RIC8A	A0A6M7GZB8	espJ
RICK	O68718	yopJ
RIN4	D4I1J6, Q6LAD6	avrRpt2, avrRpt2
AtRIN4	P13835	avrB
GmRIN4A	P13835	avrB
GMRIN4B	P13835	avrB
RIPK1	D2TT36, A0A482PDI9, A0A0H3JP21, Q8XBX8, Q8X837, Q99Q01	espL2, nleB, espL2, nleB1, nleB2, ospD3
RIPK3	D2TT36, A0A0H3JP21, Q99Q01	espL2, espL2, ospD3
ROCK1	Q9AJW7	ipgB2
ROCK2	Q9AJW7	ipgB2
RPN13	P18009	ipaH4.5
RPN8	G0T341	xopJ2
RPS3	D2TRX7, Q8X831, Q8XAL6, Q8ZNG2	nleH, nleH1-1/nleH1, nleH1_2/nleH2, sseL
RRS1/AT5G45260	Q48B92	avrRps4
RSK1	P17778	yopM
Rab1	Q9Z7W9, K0GA27, B7UMC8, B7UH72, Q7BU69	CP_0163, cpoS, espG, spG2, virA
Rab10	Q9Z7W9, K0GA27	CP_0163, cpoS
Rab11	Q9Z7W9	CP_0163
Rab11A	P33546	icsB
Rab11B	P0DUJ7, P33546	sseK3, icsB
Rab13	P33546	icsB
Rab14	K0GA27	cpoS
Rab1A	H9L407, H9L486, P0DUJ7	sseF, sseG, sseK3

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
Rab22A	P33546	icsB
Rab23	P33546	icsB
Rab29	A0A0H3N9Y3, Q8ZQC8	gtgE, sopD2
Rab31	P18009	ipaH4.5
Rab32	A0A0H3N9Y3, Q8ZQC8	gtgE, sopD2
Rab33	Q7BU69	virA
Rab34	K0GA27, P33546	cpoS, icsB
Rab35	K0GA27, P33546, Q7BU69	cpoS, icsB, virA
Rab37	Q7BU69	virA
Rab38	A0A0H3N9Y3, Q8ZQC8	gtgE, sopD2
Rab4	K0GA27	CT229/cpoS
Rab5	O52623	sopE
Rab5A	P0DUJ7	sseK3
Rab5B	P0DUJ7	sseK3
Rab5C	P0DUJ7	sseK3
Rab6	K0GA27, Q7BU69	cpoS, virA
Rab8	K0GA27, Q8ZQC8	cpoS, sopD2
Rab8A	P40722, P33546	sopD, icsB
Rac	Q93RN4	yopT
Rac1	Q93Q17, D5LUP3, Q63K41, Q7P1B7, Q84H14, O52623, P74873, P33548, Q9AJW7, P33546, P31492, Q05608	aexT, aexU, bopE, copE, C6H65_01640, sopE, sptP, ipbB1, ipgB2, icsB, yopE, ypkA/yopO
Rac2	P31492, Q05608	yopE, ypkA/yopO
RalA	P33546	icsB
RalB	P33546	icsB
Ras	Q87GI9	VPA1327/vopT
RelB	Q8X834, A0A0F6B537, A0A0F6AZI6, A0A0F6AZQ0	nleC, gogA, gtgA, pipA
RhoA	Q93Q17, D5LUP3, Q8X4W3, Q84H14, Q9AJW7, P33546, P31492, Q05608, Q93RN4	aexT, aexU, espM2, C6H65_01640, ipgB2, icsB, yopE, ypkA/yopO, yopT
RhoA(GDP)	Q56061	sifA
RhoA(GTP)	Q9FD10	sseJ
RhoC	Q9FD10, P33546	sseJ, icsB

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
RhoG	P31492, Q93RN4	yopE, yopT
Rpn2	A0A0F6AZI6	gtgA
S100A6	Q8ZNG2	sseL
SAP102	Q8XAJ5	nleA
SAP97	Q8XAJ5	nleA
SEPT7	P33546	icsB
SERPINH1	D2TK72	nleG1
SHC1	O84462	tarP
SHP-1	B7UM99, Q7DB77	tir
SHP-2	B7UM99	tir
SKAP-HOM	P08538	yopH
SKIP/PLEKHM2	Q56061	sifA
SLC25A5	D2TK72	nleG1
SLC25A6	D2TK72	nleG1
SLP-76	P08538	yopH
SNAP23	P0CL52	sipA
SNX18	D2TKD7	espF
SNX27	Q8XAJ5	nleA
SNX33	D2TKD7, Q7DB85	espF, espF
SNX5	P0DJI4	incE
SNX6	P0DJI4	incE
SNX9	D2TKD7, Q7DB85	espF, espF
STAT3	A0A0F6B506	sarA
STING	O68718	yopJ
STK16	Q05129, Q7DB77	espB, tir
Sec24A	Q8XAJ5	nleA
Sec24B	Q8XAJ5	nleA
Sec24C	Q8XAJ5	nleA
Sec24D	Q8XAJ5	nleA
SepL	Q7DB77	tir
SgpI1	A0A0F6AZI6	gtgA
Shp-2	D2TKE8	tir
SipB	P0CL47, Q56026	sipC, sipD
SipC	Q56019	sipB
SipD	P41784, Q56019	prgI, sipB
Skp1	Q8ZN18	gogB
SlPTO	Q4ZX47	hopZ3
SlRIN4-1	Q4ZX47	hopZ3
SlRIN4-2	Q4ZX47	hopZ3
SlRIN4-3	Q4ZX47	hopZ3
SlRIPK	Q4ZX47	hopZ3
Slc25a11	A0A0F6AZI6	gtgA
Slc25a12	A0A0F6AZI6	gtgA
Slc25a13	A0A0F6AZI6	gtgA

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
SnRK1	G0T341	xopJ2
Spa47	Q04640	mxiC
Src	A0A6M7GZB8, A0A0K0HD42	espJ, sboC,seoC
SsaE	Q7BVH7	sseB
SsaM	P0CZ04	spiC
SseF	H9L486	sseG
SseG	H9L407	sseF
Ssr1	A0A0F6AZI6	gtgA
Ste7	Q8VSC3	ipaH9.8
SycN/YscB	Q663K1	yopN
Syk	A0A6M7GZB8	espJ
Syn13	P0CL52	sipA
Syn6	P0CL47	sipC
Syn7	P0CL52	sipA
Syn8	P0CL52	sipA
TAB2	D2TT38, B7UI22, A0A3T2V133	nleE, nleE, ospZ
TAB3	D2TT38, B7UI22, A0A3T2V133	nleE, nleE, ospZ
TAK1	O68718	yopJ
TBCB	Q9L9J3, P0DUJ7	sseK1, sseK3
TBK1	P18009	ipaH4.5
TC0F1	Q8XAJ5	nleA
TIM17B	Q7DB68	espZ
TIM17b	D2TKF8	espZ
TLR2	P23994	lcrV
TMEM127	Q8ZNP2	steD
TNFR1	A0A482PDI9, Q8X837	nleB, nleB2
TNFRSF1A/TNFR1	Q8XBX8	nleB1
TNFRSF25/DR3	Q8XBX8	nleB1
TNIP1	Q8VSC3	ipaH9.8
TRADD	A0A482PDI9, Q8XBX8, P0DUJ7	nleB, nleB1, sseK3
TRAF2	A0A0H3JDV8, A0A0H2UY03, O68718	espX7,nleL, ipaH_1/ipaH0722, yopJ
TRAF5	A0A0H3JDV8	espX7,nleL
TRAF6	A0A0H3JDV8, O68718	espX7,nleL, yopJ
TRIF	D2TT36, A0A0H3JP21	espL2, espL2
TRIM32	P0DUJ7	sseK3
TRIM56	Q8ZNR3	sopA
TRIM65	Q8ZNR3	sopA
TRIP6	A0A0F6AZL3	sseI
TRNT1	Q8XAL7	nleF

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
TUFM	D2TK72	nleG1
TassC	P0CZ04	spiC
Tco98p	A0A822PPP2	ERS574920_04342
Tec	A0A6M7GZB8	espJ
Tmp21	D2TRX8, Q8XAL7	nleF, nleF
Toca/1	P33546	icsB
Trx1	Q8ZQQ2	slrP
Tsg101	O84624, A0A0F7X663	CT_619, CT620
TyeA	Q663K1	yopN
U2AF35	Q8VSC3	ipaH9.8
UBE2D1	Q8X509, D2AJU3	nleG2-3, ospG
UBE2D1/UBCH5A	Q8ZNR3, Q83R64, D2A6P4, A0A0H2V170, Q83RJ4, P18009	sopA, ipaH_4/ipah1880, ipaH2022/ipaH2202/ipaH4/H7, ipaH_7, ipaH3, ipaH4.5
UBE2D2	Q8X509, Q8X4X3, A0A0H3JE38, A0A0H2USG1, Q8VSC3, D2AJU3	nleG2-3, nleG5-1, nleG6-2, ipaH1.4, ipaH9.8, ospG
UBE2D2/UBCH4	P18009	ipaH4.5
UBE2D2/UBCH5B	A0A0H2USG1, Q83RJ4, P18014	ipaH1.4, ipaH3, ipaH7.8
UBE2D3	Q8X509, P18014, D2AJU3	nleG2-3, ipaH7.8, ospG
UBE2D3/UBCH5C	Q8ZNR3, P0CE12, Q83RJ4	sopA, sspH2, ipaH3
UBE2D4	Q8X509	nleG2-3
UBE2E1	Q8Z7T2, D2AJU3	sboD/stoD, ospG
UBE2E1/UBCH6	P18009	ipaH4.5
UBE2E2	Q8X509, D2AJU3	nleG2-3, ospG
UBE2E3	Q8X509	nleG2-3
UBE2K	D2AJU3	ospG
UBE2L3/UBCH7	Q8ZNR3	sopA
UBE2L6	D2AJU3	ospG
UBE2N	D2AJU3	ospG
UBE2N/UBC13	Q8VSD5	ORF169b/ospI
UDP-Glc	O84804, Q8X837	glgA, nleB2
UDP-GlcNAc	B7UI23, P0DUJ7	lifA, sseK3
UFC1	Q8XAL6	nleH1_2/nleH2
UQCR2	D2TK72	nleG1
Usmg5	A0A0F6AZI6	gtgA
VAMP3	Q9Z8A1, P0DPS6	CPn_0442/(CT006), incG
VAMP7	P0CI27	incA
VAMP8	P0CI27, P33546	incA, icsB
VAPA	Q9Z8A0	IncV

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
VAPB	Q9Z8A0	IncV
VASP	Q8X482, Q05608	espF(U)/tccP, ypkA/yopO
VAV	P08538	yopH
VAV1/Cdc24	Q8ZP57	steC
VAV2	O84462	tarP
VCP/p97	P74873	sptP
VapB	O06662	vapC/MvpA
WASP	Q05608	ypkA/yopO
WIP	Q05608	ypkA/yopO
WIP1	Q9FCY7	wtsE
WIP2	Q9FCY7	wtsE
WIPF1	D2TKD7	espF
WRKY33/AT2G38470	Q48B92	avrRps4
WRKY41/AT4G11070	Q48B92	avrRps4
WRKY60/AT2G25000	Q48B92	avrRps4
WRKY70/AT3G56400	Q48B92	avrRps4
XPO2	P0A2N2	vsdE/spvD
YKT6	P33546	icsB
YajL	Q9L9J3	sseK1
Yes1	A0A6M7GZB8	espJ
YipA	A0A0H2W9Z2, Q8D1P8, A0A2U2H2J5, Q8D1P6	tccC2/yipB, tcaA1/yitA, tcbA/yitB, tcaC1/yitC
YipB	Q8D1P5, Q8D1P8, A0A2U2H2J5, Q8D1P6	tccC1/yipA, tcaA1/yitA, tcbA/yitB, tcaC1/yitC
YitA	Q8D1P5, A0A0H2W9Z2, A0A2U2H2J5, Q8D1P6	tccC1/yipA, tccC2/yipB, tcbA/yitB, tcaC1/yitC
YitB	Q8D1P5, A0A0H2W9Z2, Q8D1P8, Q8D1P6	tccC1/yipA, tccC2/yipB, tcaA1/yitA, tcaC1/yitC
YitC	Q8D1P5, A0A0H2W9Z2, Q8D1P8, A0A2U2H2J5	tccC1/yipA, tccC2/yipB, tcaA1/yitA, tcbA/yitB
Yme1I1	A0A0F6AZI6	gtgA
YopB	P23994, P37132, P27474	lcrV, yopD, yopQ/yopK
YopD	P23994, Q06114, P27474	lcrV, yopB, yopQ/yopK
YopN	P69971	yscI
YscF	P23994, P69971	lcrV, yscI
YscI	Q01247	yscF
YscO	P68587	yscP
YscU	P68587	yscP
YscV	P61416	yscX

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
YscY	P61416	yscX
YspB	A0A8B6KWG7	yspD
YspD	Q8KQ84	yspB
ZBP1	D2TT36, A0A0H3JP21	espL2, espL2
ZNF626	Q8X482	espF(U)/tccP
ZNHIT1	Q8XA11	espY1
ZNRF2	Q54150	ipaJ
ZRANB3	A0A3T2V133	ospZ
actin	Q93Q17, Q87GE5	aexT, vopL
alpha-actinin	B7UM99	tir
alpha-catenin	Q05129	espB
alpha-tubulin	D2TKD5, B7UMC8, B7UH72, Q7BU69, Q4G4D4	espG, espG, espG2, virA, eseG
annexin A2/ANXA2	A0A0H3JP21	espL2
annexin-2	D2TT36	espL2
beclin-1	Q8ZMI3	avrA
beta-4-integrin	D5LUP3	aexU
beta-catenin/CTNNB1	Q8ZMI3	avrA
beta-tubulin	Q7BU69	virA
c-Fyn	D2TKE8, B7UM99	tir
c-Rel	Q8X834	nleC
c-Src	B7UM99	tir
cAMP	Q9I1S4	exoY
cGMP	Q9I1S4	exoY
calcium pectate	D4HVP9	hrpW
calmodulin	Q7NWF2, Q8VSJ7, Q8VSL8, A0A0H2US87	copC, ospC1, ospC2, ospC3
calpastatin	Q7BU69	virA
caprin2	A0A0F7WRH6	CPn_1027
cardiolipin	Q3BY51	xopB
caspase-1	Q56019, P18011, P17778	sipB, ipaB, yopM
caspase-11	D2TRX8	nleF
caspase-3	Q7NWF2, Q56061	copC, sifA
caspase-4	D2TRX8, Q8XAL7, A0A0H2US87	nleF, nleF, ospC3
caspase-7	Q7NWF2	copC
caspase-8	Q7NWF2, D2TRX8, Q8XAL7	copC, nleF, nleF
caspase-9	Q7NWF2, D2TRX8, Q8XAL7	copC, nleF, nleF
H2B (histone)	O84742	CT_737
H3 (histone)	O84742	CT_737

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
H4 (histone)	O84742	CT_737
cellulose	A0A562K8J2	gunA2
chicken lysozyme C	B7UJB3	ivy
cholesterol	Q63K42, Q9FD10, P18011, P33546	bopA, sseJ, ipaB, icsB
chromatin	Q8VSP9	ospF
cofilin	D2TI21, D2TM85, Q05608	espM2, espM3, ypkA/yopO
cortactin	Q8X482, Q7DB77	espF(U)/tccP, tir
cullin-1	Q8ZPD7	steA
cyclophilin	D4I1J6	avrRpt2
deoxycholate	Q56026, P18013	sipD, ipaD
ensconsin/MAP7	Q8XBX8, Q8X837	nleB1, nleB2
fMet-RNA	O06662	vapC
filamin	A0A0F6AZL3, P0CE12	sseI, sspH2
gelsolin	Q05608	ypkA/yopO
glomulin	P18014	ipaH7.8
glucan	O84874	glgB
glycogen	O84046, O84089, O84250	glgX, malQ, glgP
hFBXO22	Q8ZN18	gogB
hGBP1	Q8VSC3	ipaH9.8
hGBP2	Q8VSC3	ipaH9.8
hGBP4	Q8VSC3	ipaH9.8
hGBP6	Q8VSC3	ipaH9.8
hIFN-gamma	P23994	lcrV
hNAIP	A0A2T5NVZ6	cprI
hNAIP1	P0A223	mxiH
hNAIP2	P41785, P0A225	prgJ, mxiI
hSGT1	P0CE12	sspH2
human lysozyme C	B7UJB3	ivy
importin-alpha 1	Q07061	avrBs4
importin-alpha 1/CalMP-alpha1	P14727	avrBs3
importin-alpha 1/KPNA2	Q8VSP9	ospF
importin-alpha 2/CalMP-alpha2	P14727	avrBs3
inositolhexakisphosphate/IP6	A0A822PPP2	ERS574920_04342
intimin	B7UM99, Q7DB77	tir
kinesin-1	Q8ZMM8, Q56061	pipB2, sifA
kinesin-3	Q56061	sifA
mDia1/DIAPH1	Q9AJW7, Q05608	ipgB2, ypkA/yopO
mFADD	Q9L9J3	sseK1

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
mGBP10	Q8VSC3	ipaH9.8
mGBP11	Q8VSC3	ipaH9.8
mGBP2	Q8VSC3	ipaH9.8
mGBP3	Q8VSC3	ipaH9.8
mGBP6	Q8VSC3	ipaH9.8
mGBP7	Q8VSC3	ipaH9.8
mGBP9	Q8VSC3	ipaH9.8
mTNFR1lmTRAILR	P0DUJ7	sseK3
mTRADD	Q9L9J3	sseK1
mouse caspase-11	A0A0H2US87	ospC3
mouse-GLMN/Q8BZM1	P18014	ipaH7.8
mouse-NLRP1B	P18014	ipaH7.8
myosin-10	Q05129	espB
myosin-1a,myosin-2	Q05129	espB
myosin-1c	Q05129	espB
myosin-5	Q05129	espB
myosin-6	Q05129	espB
nopB	A0A6F8NWU7	nopA
nopX	A0A6F8NWU7	nopA
nucleid acids	A0A2U3QA97	ernA
p130Cas	P08538	yopH
p155-RhoGEF/ARHGEF1	B7UMA2	espH
p300	Q8X834	nleC
p300 acetyltransferase	D2TK70	nleC
p38, p38-beta	A0A0H3JGR6, P0A2M9, D2TML3	nleD, spvC, nleD-1
p50/NFKB1	Q8X834	nleC
p63-RhoGEF/ARHGEF25	B7UMA2	espH
p65(RelA)	A0A0F6B537, A0A0F6AZI6, A0A0F6AZQ0, D2TK70, P18009, Q8X834	gogA, gtgA, pipA, ipaH4.5, nleC
p85:Lck	P08538	yopH
paxilin	P08538	yopH
phospholipids	O85477	yplA
preferredoxin	O54581	dspE/dspA
prgI	A0A0H3NF83	orgC
profilin	P0CE12	sspH2
snapin	Q9Z8P7	incB
talin-1	Q8ZNG2, P18010	sseL, ipaA

T3SE and their targets (continued).

target	effector (Uniprot/NCBI ID)	effector gene name
ubiquitin	Q7NY09, D2TTX7, D2TRY0, D2TI20, Q8ZNR3, Q8ZNG2, Q8Z7T2, A0A0H2UY03, Q83R64, D2A6P4, A0A0H2V170, A0A0H2USG1, Q83RJ4, P18009, P18014, Q8VSC3, D2AJU3	cteC, espX7,nleL, nleG7, nleG8, sopA, sseL, sboD/stoD, ipaH_1/ipaH0722, ipaH_4/ipah1880, ipaH2022/ipaH2202/ipaH4/H7, ipaH_7, ipaH1.4, ipaH3/IpaH1383, ipaH4.5, ipaH7.8, ipaH9.8, ospG
vinculin	P18010	ipaA
xyloglucan	M4PUR5	nopAA/gunA

Table A.7.: Identified localizations of T3SS-secreted proteins Main localizations are marked in **bold**.

localization	frequency
unspecified	253
unknown	159
host cell cytoplasm	55
host cell cytosol	44
host cell cytoskeleton	13
host cell actin pedestals	7
host cell focal adhesions	4
Salmonella-induced filaments (Sif)	3
host cell microtubules	2
host cell filopodia	1
host cell actin stress fibres	1
host cell actin pedestals	1
host cell wall	1
host cell membrane	83
host cell lipid rafts	3
proximal to host cell membrane	3
host cell GM1-enriched lipid rafts	1
host cell caveolae	1
proximal to the host cell membrane	1
host cell nucleus	43
host cell perinuclear region	4
host cell perinuclear vesicles	1
proximal to host cell nuclear membrane	1
host cell golgi apparatus	11
host cell cis-golgi network	3

Identified localization of T3SS-secreted proteins (continued)

localization	**frequency**
host cell golgi membrane	3
host cell mitochondrion	15
host cell endoplasmic reticulum	6
host cell endosome	5
host cell chloroplast	3
T3SS apparatus	42
T3SS needle tip	8
flagellar apparatus	1
extracellular space	13
bacteria-associated	4
proximal to bacteria	1
Yersinia outer membrane	2
bacterial outer membrane	2
bacterial periplasm	1
lipid droplets	2
Salmonella-containing vacuole (SCV)	13
inclusion membrane	30
inclusion lumen	14
near inclusion membrane	4
inclusion membrane microdomains	3
host cell membrane-trafficking organelles	2
Shigella-containing vacuole (SCV)	1
Edwardsiella-containing vacuole (ECV)	1
inclusion lumen	2
periphal at inclusion membrane	1
intrainclusion	1
reticulate bodies at the inclusion membrane	1
plasma membrane-associated vesicles	1

Bibliography

(1) Diseases, G. 2.; Collaborators, I. *Lancet (London, England)* **2020**, *396*, 1204–1222, ppublish.

(2) Collaborators, G. B. D. 2. A. R. *Lancet (London, England)* **2022**, *400*, 2221–2248, ppublish.

(3) Collaborators, A. R. *Lancet (London, England)* **2022**, *399*, 629–655, ppublish.

(4) De Kraker, M. E. A.; Stewardson, A. J.; Harbarth, S. *PLoS medicine* **2016**, *13*, e1002184, epublish.

(5) Asokan, G. V.; Ramadhan, T.; Ahmed, E.; Sanad, H. *Oman medical journal* **2019**, *34*, 184–193, ppublish.

(6) Galdiero, S.; Falanga, A.; Cantisani, M.; Tarallo, R.; Della Pepa, M. E.; D'Oriano, V.; Galdiero, M. *Current protein & peptide science* **2012**, *13*, 843–854, ppublish.

(7) Nikaido, H.; Pagès, J.-M. *FEMS microbiology reviews* **2012**, *36*, 340–363, ppublish.

(8) Green, E. R.; Mecsas, J. *Microbiology spectrum* **2016**, *4*, DOI: `10.1128/microbiolspec.VMBF-0012-2015`, ppublish.

(9) Beckwith, J. *Research in microbiology* **2013**, *164*, 497–504, ppublish.

(10) Orfanoudaki, G.; Economou, A. *Molecular & cellular proteomics : MCP* **2014**, *13*, 3674–3687, ppublish.

(11) Monteferrante, C. G.; Miethke, M.; van der Ploeg, R.; Glasner, C.; van Dijl, J. M. *The Journal of biological chemistry* **2012**, *287*, 29789–29800, ppublish.

(12) Simone, D.; Bay, D. C.; Leach, T.; Turner, R. J. *PloS one* **2013**, *8*, e78742, epublish.

(13) Sitsel, O.; Wang, Z.; Janning, P.; Kroczek, L.; Wagner, T.; Raunser, S. *Nature microbiology* **2024**, *9*, 390–404, ppublish.

(14) Lasica, A. M.; Ksiazek, M.; Madej, M.; Potempa, J. *Frontiers in cellular and infection microbiology* **2017**, *7*, 215, epublish.

(15) Sana, T. G.; Voulhoux, R.; Monack, D. M.; Ize, B.; Bleves, S. *Frontiers in cellular and infection microbiology* **2019**, *9*, 473, epublish.

(16) Arechaga, I.; Cascales, E. *Frontiers in microbiology* **2022**, *13*, 917591, epublish.

(17) Famelis, N.; Rivera-Calzada, A.; Degliesposti, G.; Wingender, M.; Mietrach, N.; Skehel, J. M.; Fernandez-Leiro, R.; Böttcher, B.; Schlosser, A.; Llorca, O.; Geibel, S. *Nature* **2019**, *576*, 321–325, ppublish.

(18) Bunduc, C. M.; Fahrenkamp, D.; Wald, J.; Ummels, R.; Bitter, W.; Houben, E. N. G.; Marlovits, T. C. *Nature* **2021**, *593*, 445–448, ppublish.

(19) Russell, A. B.; Hood, R. D.; Bui, N. K.; LeRoux, M.; Vollmer, W.; Mougous, J. D. *Nature* **2011**, *475*, 343–347, epublish.

(20) Morgan, J. L. W.; Acheson, J. F.; Zimmer, J. *Structure (London, England : 1993)* **2017**, *25*, 522–529, ppublish.

(21) Alvarez-Martinez, C. E.; Christie, P. J. *Microbiology and molecular biology reviews : MMBR* **2009**, *73*, 775–808, ppublish.

(22) Lawley, T. D.; Klimke, W. A.; Gubbins, M. J.; Frost, L. S. *FEMS microbiology letters* **2003**, *224*, 1–15, ppublish.

(23) Fronzes, R.; Christie, P. J.; Waksman, G. *Nature reviews. Microbiology* **2009**, *7*, 703–714, ppublish.

(24) Guglielmini, J.; de la Cruz, F.; Rocha, E. P. C. *Molecular biology and evolution* **2013**, *30*, 315–331, ppublish.

(25) MacIntyre, D. L.; Miyata, S. T.; Kitaoka, M.; Pukatzki, S. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 19520–19524, ppublish.

(26) Schwarz, S.; West, T. E.; Boyer, F.; Chiang, W.-C.; Carl, M. A.; Hood, R. D.; Rohmer, L.; Tolker-Nielsen, T.; Skerrett, S. J.; Mougous, J. D. *PLoS pathogens* **2010**, *6*, e1001068, epublish.

(27) Wenren, L. M.; Sullivan, N. L.; Cardarelli, L.; Septer, A. N.; Gibbs, K. A. *mBio* **2013**, *4*, DOI: 10.1128/mBio.00374-13, epublish.

(28) Isberg, R. R.; O'Connor, T. J.; Heidtman, M. *Nature reviews. Microbiology* **2009**, *7*, 13–24, ppublish.

(29) Pallen, M. J.; Beatson, S. A.; Bailey, C. M. *FEMS microbiology reviews* **2005**, *29*, 201–229, ppublish.

(30) Goldberg, T.; Rost, B.; Bromberg, Y. *Scientific reports* **2016**, *6*, 34516, epublish.

(31) Burkinshaw, B. J.; Strynadka, N. C. J. *Biochimica et biophysica acta* **2014**, *1843*, 1649–1663, ppublish.

(32) Stringlis, I. A.; Zamioudis, C.; Berendsen, R. L.; Bakker, P. A. H. M.; Pieterse, C. M. J. *Frontiers in microbiology* **2019**, *10*, 1631, epublish.

(33) Abrusci, P.; McDowell, M. A.; Lea, S. M.; Johnson, S. *Current opinion in structural biology* **2014**, *25*, 111–117, ppublish.

(34) Hu, Y.; Huang, H.; Cheng, X.; Shu, X.; White, A. P.; Stavrinides, J.; Köster, W.; Zhu, G.; Zhao, Z.; Wang, Y. *Environmental microbiology* **2017**, *19*, 3879–3895, ppublish.

(35) Deng, W.; Marshall, N. C.; Rowland, J. L.; McCoy, J. M.; Worrall, L. J.; Santos, A. S.; Strynadka, N. C. J.; Finlay, B. B. *Nature reviews. Microbiology* **2017**, *15*, 323–337, ppublish.

(36) Cascales, E. *Cell* **2017**, *168*, 949–951, ppublish.

(37) Hu, B.; Lara-Tejero, M.; Kong, Q.; Galán, J. E.; Liu, J. *Cell* **2017**, *168*, 1065–1074.e10, ppublish.

(38) Wimmi, S.; Balinovic, A.; Brianceau, C.; Pintor, K.; Vielhauer, J.; Turkowyd, B.; Helbig, C.; Fleck, M.; Langenfeld, K.; Kahnt, J.; Glatter, T.; Endesfelder, U.; Diepold, A. *Nature microbiology* **2024**, *9*, 185–199, ppublish.

(39) Miletic, S.; Fahrenkamp, D.; Goessweiner-Mohr, N.; Wald, J.; Pantel, M.; Vesper, O.; Kotov, V.; Marlovits, T. C. *Nature communications* **2021**, *12*, 1546, epublish.

(40) Bergeron, J. R. C.; Worrall, L. J.; Sgourakis, N. G.; DiMaio, F.; Pfuetzner, R. A.; Felise, H. B.; Vuckovic, M.; Yu, A. C.; Miller, S. I.; Baker, D.; Strynadka, N. C. J. *PLoS pathogens* **2013**, *9*, e1003307, ppublish.

(41) Galán, J. E.; Waksman, G. *Cell* **2018**, *172*, 1306–1318, ppublish.

(42) Schraidt, O.; Marlovits, T. C. *Science (New York, N.Y.)* **2011**, *331*, 1192–1195, ppublish.

(43) Lara-Tejero, M.; Galán, J. E. *EcoSal Plus* **2019**, *8*, DOI: `10.1128/ecosalplus.ESP-0039-2018`, ppublish.

(44) Lefebre, M. D.; Galán, J. E. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 817–822, ppublish.

(45) Rathinavelan, T.; Lara-Tejero, M.; Lefebre, M.; Chatterjee, S.; McShan, A. C.; Guo, D.-C.; Tang, C.; Galan, J. E.; De Guzman, R. N. *Journal of molecular biology* **2014**, *426*, 2958–2969, ppublish.

(46) Diepold, A.; Wagner, S. *FEMS microbiology reviews* **2014**, *38*, 802–822, ppublish.

(47) Kuhlen, L.; Johnson, S.; Zeitler, A.; Bäurle, S.; Deme, J. C.; Caesar, J. J. E.; Debo, R.; Fisher, J.; Wagner, S.; Lea, S. M. *Nature communications* **2020**, *11*, 1296, epublish.

(48) Inoue, Y.; Ogawa, Y.; Kinoshita, M.; Terahara, N.; Shimada, M.; Kodera, N.; Ando, T.; Namba, K.; Kitao, A.; Imada, K.; Minamino, T. *Structure (London, England : 1993)* **2019**, *27*, 965–976.e6, ppublish.

(49) Diepold, A.; Kudryashev, M.; Delalez, N. J.; Berry, R. M.; Armitage, J. P. *PLoS biology* **2015**, *13*, e1002039, epublish.

(50) Bernal, I.; Börnicke, J.; Heidemann, J.; Svergun, D.; Horstmann, J. A.; Erhardt, M.; Tuukkanen, A.; Uetrecht, C.; Kolbe, M. *Journal of molecular biology* **2019**, *431*, 3787–3803, ppublish.

(51) Tachiyama, S.; Chang, Y.; Muthuramalingam, M.; Hu, B.; Barta, M. L.; Picking, W. L.; Liu, J.; Picking, W. D. *The Journal of biological chemistry* **2019**, *294*, 19184–19196, ppublish.

(52) Ernst, N. H.; Reeves, A. Z.; Ramseyer, J. E.; Lesser, C. F. *mBio* **2018**, *9*, DOI: 10.1128/mBio.01050-18, epublish.

(53) Spaeth, K. E.; Chen, Y.-S.; Valdivia, R. H. *PLoS pathogens* **2009**, *5*, e1000579, ppublish.

(54) Majewski, D. D.; Worrall, L. J.; Hong, C.; Atkinson, C. E.; Vuckovic, M.; Watanabe, N.; Yu, Z.; Strynadka, N. C. J. *Nature communications* **2019**, *10*, 626, epublish.

(55) Jensen, J. L.; Yamini, S.; Rietsch, A.; Spiller, B. W. *PLoS pathogens* **2020**, *16*, e1008923, epublish.

(56) Hu, B.; Morado, D. R.; Margolin, W.; Rohde, J. R.; Arizmendi, O.; Picking, W. L.; Picking, W. D.; Liu, J. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 1047–1052, ppublish.

(57) Akeda, Y.; Galán, J. E. *Nature* **2005**, *437*, 911–915, ppublish.

(58) Galán, J. E.; Wolf-Watz, H. *Nature* **2006**, *444*, 567–573, ppublish.

(59) Erhardt, M.; Mertens, M. E.; Fabiani, F. D.; Hughes, K. T. *PLoS genetics* **2014**, *10*, e1004800, epublish.

(60) Minamino, T.; Namba, K. *Nature* **2008**, *451*, 485–488, ppublish.

(61) Minamino, T.; Kinoshita, M.; Namba, K. *Frontiers in microbiology* **2022**, *13*, 864178, epublish.

(62) Wilharm, G.; Dittmann, S.; Schmid, A.; Heesemann, J. *International journal of medical microbiology : IJMM* **2007**, *297*, 27–36, ppublish.

(63) Paul, K.; Erhardt, M.; Hirano, T.; Blair, D. F.; Hughes, K. T. *Nature* **2008**, *451*, 489–492, ppublish.

(64) Diepold, A.; Amstutz, M.; Abel, S.; Sorg, I.; Jenal, U.; Cornelis, G. R. *The EMBO journal* **2010**, *29*, 1928–1940, ppublish.

(65) Marlovits, T. C.; Kubori, T.; Lara-Tejero, M.; Thomas, D.; Unger, V. M.; Galán, J. E. *Nature* **2006**, *441*, 637–640, ppublish.

(66) Journet, L.; Agrain, C.; Broz, P.; Cornelis, G. R. *Science (New York, N.Y.)* **2003**, *302*, 1757–1760, ppublish.

(67) Shaulov, L.; Gershberg, J.; Deng, W.; Finlay, B. B.; Sal-Man, N. *mBio* **2017**, *8*, DOI: 10.1128/mBio.01733-16, epublish.

(68) Ho, O.; Rogne, P.; Edgren, T.; Wolf-Watz, H.; Login, F. H.; Wolf-Watz, M. *The Journal of biological chemistry* **2017**, *292*, 3299–3311, ppublish.

(69) Monjarás Feria, J. V.; Lefebre, M. D.; Stierhof, Y.-D.; Galán, J. E.; Wagner, S. *mBio* **2015**, *6*, e01459–e01415, epublish.

(70) Ngo, T.-D.; Perdu, C.; Jneid, B.; Ragno, M.; Novion Ducassou, J.; Kraut, A.; Couté, Y.; Stopford, C.; Attrée, I.; Rietsch, A.; Faudry, E. *Journal of molecular biology* **2020**, *432*, 166690, ppublish.

(71) Yu, X.-J.; Grabe, G. J.; Liu, M.; Mota, L. J.; Holden, D. W. *mBio* **2018**, *9*, DOI: 10.1128/mBio.01149-18, epublish.

(72) Shen, D.-K.; Blocker, A. J. *PloS one* **2016**, *11*, e0155141, epublish.

(73) Büttner, D. *Microbiology and molecular biology reviews : MMBR* **2012**, *76*, 262–310, ppublish.

(74) Dean, P. *FEMS microbiology reviews* **2011**, *35*, 1100–1125, ppublish.

(75) Kaniga, K.; Uralil, J.; Bliska, J. B.; Galán, J. E. *Molecular microbiology* **1996**, *21*, 633–641, ppublish.

(76) Stavrinides, J.; Ma, W.; Guttman, D. S. *PLoS pathogens* **2006**, *2*, e104, ppublish.

(77) Delevoye, C.; Nilges, M.; Dehoux, P.; Paumet, F.; Perrinet, S.; Dautry-Varsat, A.; Subtil, A. *PLoS pathogens* **2008**, *4*, e1000022, epublish.

(78) Gibbs, K. D.; Washington, E. J.; Jaslow, S. L.; Bourgeois, J. S.; Foster, M. W.; Guo, R.; Brennan, R. G.; Ko, D. C. *Cell host & microbe* **2020**, *27*, 129–139.e4, ppublish.

(79) Xing, Q.; Shi, K.; Portaliou, A.; Rossi, P.; Economou, A.; Kalodimos, C. G. *Nature communications* **2018**, *9*, 1773, epublish.

(80) Allison, S. E.; Tuinema, B. R.; Everson, E. S.; Sugiman-Marangos, S.; Zhang, K.; Junop, M. S.; Coombes, B. K. *The Journal of biological chemistry* **2014**, *289*, 23734–23744, ppublish.

(81) LeBlanc, M.-A.; Fink, M. R.; Perkins, T. T.; Sousa, M. C. *Proceedings of the National Academy of Sciences of the United States of America* **2021**, *118*, DOI: 10.1073/pnas.2019566118, ppublish.

(82) Lara-Tejero, M.; Kato, J.; Wagner, S.; Liu, X.; Galán, J. E. *Science (New York, N.Y.)* **2011**, *331*, 1188–1191, ppublish.

(83) Schlumberger, M. C.; Müller, A. J.; Ehrbar, K.; Winnen, B.; Duss, I.; Stecher, B.; Hardt, W.-D. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 12548–12553, ppublish.

(84) Van Engelenburg, S. B.; Palmer, A. E. *Chemistry & biology* **2008**, *15*, 619–628, ppublish.

(85) Kubori, T.; Galán, J. E. *Cell* **2003**, *115*, 333–342, ppublish.

(86) Letzelter, M.; Sorg, I.; Mota, L. J.; Meyer, S.; Stalder, J.; Feldman, M.; Kuhn, M.; Callebaut, I.; Cornelis, G. R. *The EMBO journal* **2006**, *25*, 3223–3233, ppublish.

(87) Lee, S. H.; Galán, J. E. *Molecular microbiology* **2004**, *51*, 483–495, ppublish.

(88) Bagos, P. G.; Nikolaou, E. P.; Liakopoulos, T. D.; Tsirigos, K. D. *Bioinformatics (Oxford, England)* **2010**, *26*, 2811–2817, ppublish.

(89) Spitz, O.; Erenburg, I. N.; Beer, T.; Kanonenberg, K.; Holland, I. B.; Schmitt, L. *Microbiology spectrum* **2019**, *7*, DOI: 10.1128/microbiolspec.PSIB-0003-2018, ppublish.

(90) Delepelaire, P. *Biochimica et biophysica acta* **2004**, *1694*, 149–161, ppublish.

(91) Schesser, K.; Frithz-Lindsten, E.; Wolf-Watz, H. *Journal of bacteriology* **1996**, *178*, 7227–7233, ppublish.

(92) Sory, M. P.; Boland, A.; Lambermont, I.; Cornelis, G. R. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 11998–12002, ppublish.

(93) Anderson, D. M.; Schneewind, O. *Science (New York, N.Y.)* **1997**, *278*, 1140–1143, ppublish.

(94) Anderson, D. M.; Schneewind, O. *Molecular microbiology* **1999**, *31*, 1139–1148, ppublish.

(95) Ramamurthi, K. S.; Schneewind, O. *Journal of bacteriology* **2002**, *184*, 3321–3328, ppublish.

(96) Ramamurthi, K. S.; Schneewind, O. *Molecular microbiology* **2003**, *50*, 1189–1198, ppublish.

(97) Lloyd, S. A.; Norman, M.; Rosqvist, R.; Wolf-Watz, H. *Molecular microbiology* **2001**, *39*, 520–531, ppublish.

(98) Karavolos, M. H.; Roe, A. J.; Wilson, M.; Henderson, J.; Lee, J. J.; Gally, D. L.; Khan, C. M. A. *Journal of bacteriology* **2005**, *187*, 1559–1567, ppublish.

(99) Amer, A. A. A.; Åhlund, M. K.; Bröms, J. E.; Forsberg, Å.; Francis, M. S. *Journal of bacteriology* **2011**, *193*, 6683–6700, ppublish.

(100) Niemann, G. S.; Brown, R. N.; Mushamiri, I. T.; Nguyen, N. T.; Taiwo, R.; Stufkens, A.; Smith, R. D.; Adkins, J. N.; McDermott, J. E.; Heffron, F. *Journal of bacteriology* **2013**, *195*, 2119–2125, ppublish.

(101) Arnold, R.; Brandmaier, S.; Kleine, F.; Tischler, P.; Heinz, E.; Behrens, S.; Niinikoski, A.; Mewes, H.-W.; Horn, M.; Rattei, T. *PLoS pathogens* **2009**, *5*, e1000376, ppublish.

(102) Pais, S. V.; Milho, C.; Almeida, F.; Mota, L. J. *PloS one* **2013**, *8*, e56292, ppublish.

(103) Da Cunha, M.; Milho, C.; Almeida, F.; Pais, S. V.; Borges, V.; Maurício, R.; Borrego, M. J.; Gomes, J. P.; Mota, L. J. *BMC microbiology* **2014**, *14*, 40, epublish.

(104) Zhou, X.; Nydam, S. D.; Christensen, J. E.; Konkel, M. E.; Orfe, L.; Friel, P.; Call, D. R. *Applied and environmental microbiology* **2012**, *78*, 3492–3494, ppublish.

(105) Miao, E. A.; Miller, S. I. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 7539–7544, ppublish.

(106) Lloyd, S. A.; Sjöström, M.; Andersson, S.; Wolf-Watz, H. *Molecular microbiology* **2002**, *43*, 51–59, ppublish.

(107) Buchko, G. W.; Niemann, G.; Baker, E. S.; Belov, M. E.; Smith, R. D.; Heffron, F.; Adkins, J. N.; McDermott, J. E. *Molecular bioSystems* **2010**, *6*, 2448–2458, ppublish.

(108) Bard, J. A. M.; Goodall, E. A.; Greene, E. R.; Jonsson, E.; Dong, K. C.; Martin, A. *Annual review of biochemistry* **2018**, *87*, 697–724, ppublish.

(109) Wang, Y.; Sun, M.; Bao, H.; White, A. P. *PloS one* **2013**, *8*, e58173, ppublish.

(110) Löwer, M.; Schneider, G. *PloS one* **2009**, *4*, e5917, epublish.

(111) Xue, L.; Tang, B.; Chen, W.; Luo, J. *Bioinformatics (Oxford, England)* **2019**, *35*, 2051–2057, ppublish.

(112) Winnen, B.; Schlumberger, M. C.; Sturm, A.; Schüpbach, K.; Siebenmann, S.; Jenny, P.; Hardt, W.-D. *PloS one* **2008**, *3*, e2178, epublish.

(113) Sorg, I.; Wagner, S.; Amstutz, M.; Müller, S. A.; Broz, P.; Lussi, Y.; Engel, A.; Cornelis, G. R. *The EMBO journal* **2007**, *26*, 3015–3024, ppublish.

(114) Login, F. H.; Wolf-Watz, H. *The Journal of biological chemistry* **2015**, *290*, 26282–26291, ppublish.

(115) Deng, W.; Yu, H. B.; Li, Y.; Finlay, B. B. *Journal of bacteriology* **2015**, *197*, 1263–1275, ppublish.

(116) Mills, E.; Baruch, K.; Aviv, G.; Nitzan, M.; Rosenshine, I. *mBio* **2013**, *4*, DOI: `10.1128/mBio.00303-13`, epublish.

(117) Mills, E.; Baruch, K.; Charpentier, X.; Kobi, S.; Rosenshine, I. *Cell host & microbe* **2008**, *3*, 104–113, ppublish.

(118) Serapio-Palacios, A.; Finlay, B. B. *Current opinion in microbiology* **2020**, *54*, 67–76, ppublish.

(119) Coburn, B.; Grassl, G. A.; Finlay, B. B. *Immunology and cell biology* **2007**, *85*, 112–118, ppublish.

(120) Bao, H.; Wang, S.; Zhao, J.-H.; Liu, S.-L. *Microbiological research* **2020**, *241*, 126591, ppublish.

(121) Agbor, T. A.; McCormick, B. A. *Cellular microbiology* **2011**, *13*, 1858–1869, ppublish.

(122) Yu, X.-J.; Liu, M.; Holden, D. W. *mBio* **2016**, *7*, DOI: 10.1128/mBio.00474-16, epublish.

(123) Liu, J.; Guo, J.-T.; Li, Y.-G.; Johnston, R. N.; Liu, G.-R.; Liu, S.-L. *Journal of basic microbiology* **2013**, *53*, 600–607, ppublish.

(124) Wang, R. F.; Kushner, S. R. *Gene* **1991**, *100*, 195–199, ppublish.

(125) Nyerges, Á.; Csörgő, B.; Nagy, I.; Bálint, B.; Bihari, P.; Lázár, V.; Apjok, G.; Umenhoffer, K.; Bogos, B.; Pósfai, G.; Pál, C. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 2502–2507, ppublish.

(126) Wannier, T. M.; Nyerges, A.; Kuchwara, H. M.; Czikkely, M.; Balogh, D.; Filsinger, G. T.; Borders, N. C.; Gregg, C. J.; Lajoie, M. J.; Rios, X.; Pál, C.; Church, G. M. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117*, 13689–13698, ppublish.

(127) Hoffmann, C.; Galle, M.; Dilling, S.; Käppeli, R.; Müller, A. J.; Songhet, P.; Beyaert, R.; Hardt, W.-D. *PloS one* **2010**, *5*, e12477, epublish.

(128) Sukhan, A.; Kubori, T.; Wilson, J.; Galán, J. E. *Journal of bacteriology* **2001**, *183*, 1159–1167, ppublish.

(129) Hoiseth, S. K.; Stocker, B. A. *Nature* **1981**, *291*, 238–239, ppublish.

(130) Chen, J.; Li, Y.; Zhang, K.; Wang, H. *Genome announcements* **2018**, *6*, DOI: 10.1128/genomeA.00097-18, epublish.

(131) Korbie, D. J.; Mattick, J. S. *Nature protocols* **2008**, *3*, 1452–1456, ppublish.

(132) Kaniga, K.; Trollinger, D.; Galán, J. E. *Journal of bacteriology* **1995**, *177*, 7078–7085, ppublish.

(133) Michiels, T.; Cornelis, G. R. *Journal of bacteriology* **1991**, *173*, 1677–1685, ppublish.

(134) Ménard, R.; Sansonetti, P.; Parsot, C. *The EMBO journal* **1994**, *13*, 5293–5302, ppublish.

(135) Kim, J. S.; Jang, J. I.; Eom, J. S.; Oh, C. H.; Kim, H. G.; Kim, B. H.; Bang, I. S.; Bang, S. H.; Park, Y. K. *Microbiology (Reading, England)* **2013**, *159*, 446–461, ppublish.

(136) Button, J. E.; Galán, J. E. *Molecular microbiology* **2011**, *81*, 1474–1483, ppublish.

(137) Bajaj, V.; Lucas, R. L.; Hwang, C.; Lee, C. A. *Molecular microbiology* **1996**, *22*, 703–714, ppublish.

(138) Lostroh, C. P.; Lee, C. A. *Journal of bacteriology* **2001**, *183*, 4876–4885, ppublish.

(139) Lorenz, R.; Bernhart, S. H.; Höner Zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L. *Algorithms for molecular biology : AMB* **2011**, *6*, 26, epublish.

(140) Takaya, A.; Tomoyasu, T.; Matsui, H.; Yamamoto, T. *Infection and immunity* **2004**, *72*, 1364–1373, ppublish.

(141) Westerhausen, S.; Nowak, M.; Torres-Vargas, C. E.; Bilitewski, U.; Bohn, E.; Grin, I.; Wagner, S. *Molecular microbiology* **2020**, *113*, 1240–1254, ppublish.

(142) England, C. G.; Ehlerding, E. B.; Cai, W. *Bioconjugate chemistry* **2016**, *27*, 1175–1187, ppublish.

(143) Promega Nano-Glo Luciferase Assay System, Instructions for Use of Products N1110, N1120, N1130 and N1150, Promega Corporation: 2022.

(144) Fan, F. J.; Shi, Y. *Bioorganic & medicinal chemistry* **2022**, *72*, 117003, ppublish.

(145) Budach, L.; Feuerpfeil, M.; Ihde, N.; Nathansen, A.; Noack, N.; Patzlaff, H.; Naumann, F.; Harmouch, H. *arXiv preprint arXiv:2207.14529* **2022**.

(146) Dong, X.; Lu, X.; Zhang, Z. *Database : the journal of biological databases and curation* **2015**, *2015*, bav064, epublish.

(147) An, Y.; Wang, J.; Li, C.; Revote, J.; Zhang, Y.; Naderer, T.; Hayashida, M.; Akutsu, T.; Webb, G. I.; Lithgow, T.; Song, J. *Scientific reports* **2017**, *7*, 41031, epublish.

(148) Wang, J.; Li, J.; Yang, B.; Xie, R.; Marquez-Lago, T. T.; Leier, A.; Hayashida, M.; Akutsu, T.; Zhang, Y.; Chou, K.-C.; Selkrig, J.; Zhou, T.; Song, J.; Lithgow, T. *Bioinformatics (Oxford, England)* **2019**, *35*, 2017–2028, ppublish.

(149) Li, J.; Wei, L.; Guo, F.; Zou, Q. *Briefings in bioinformatics* **2021**, *22*, 1918–1928, ppublish.

(150) National Library of Medicine (US) National Center for Biotechnology Information, [Identical Protein Groups `https://www.ncbi.nlm.nih.gov/ipg/` (accessed 12/29/2023).

(151) Eichinger, V.; Nussbaumer, T.; Platzer, A.; Jehl, M.-A.; Arnold, R.; Rattei, T. *Nucleic acids research* **2016**, *44*, D669–D674, ppublish.

(152) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. *Bioinformatics (Oxford, England)* **2012**, *28*, 3150–3152, ppublish.

(153) Chan, Y.-C.; Wu, H.-P.; Chuang, D.-Y. *BMC microbiology* **2009**, *9*, 181, epublish.

(154) Christensen, J. E.; Pacheco, S. A.; Konkel, M. E. *Molecular microbiology* **2009**, *73*, 650–662, ppublish.

(155) Boureau, T.; ElMaarouf-Bouteau, H.; Garnier, A.; Brisset, M.-N.; Perino, C.; Pucheu, I.; Barny, M.-A. *Molecular plant-microbe interactions : MPMI* **2006**, *19*, 16–24, ppublish.

(156) Vanden Bergh, P.; Heller, M.; Braga-Lagache, S.; Frey, J. *Proteome science* **2013**, *11*, 42, epublish.

(157) Nissinen, R. M.; Ytterberg, A. J.; Bogdanove, A. J.; VAN Wijk, K. J.; Beer, S. V. *Molecular plant pathology* **2007**, *8*, 55–67, ppublish.

(158) Deng, W.; de Hoog, C. L.; Yu, H. B.; Li, Y.; Croxen, M. A.; Thomas, N. A.; Puente, J. L.; Foster, L. J.; Finlay, B. B. *The Journal of biological chemistry* **2010**, *285*, 6790–6800, ppublish.

(159) Deng, W.; Yu, H. B.; de Hoog, C. L.; Stoynov, N.; Li, Y.; Foster, L. J.; Finlay, B. B. *Molecular & cellular proteomics : MCP* **2012**, *11*, 692–709, ppublish.

(160) Liu, X.; Lu, L.; Liu, X.; Liu, X.; Pan, C.; Feng, E.; Wang, D.; Niu, C.; Zhu, L.; Wang, H. *Journal of microbiology and biotechnology* **2017**, *27*, 171–178, ppublish.

(161) Lonjon, F.; Turner, M.; Henry, C.; Rengel, D.; Lohou, D.; van de Kerkhove, Q.; Cazalé, A.-C.; Peeters, N.; Genin, S.; Vailleau, F. *Molecular & cellular proteomics : MCP* **2016**, *15*, 598–613, ppublish.

(162) Lampaki, D.; Diepold, A.; Glatter, T. *Journal of proteome research* **2020**, *19*, 543–553, ppublish.

(163) Schumacher, J.; Waite, C. J.; Bennett, M. H.; Perez, M. F.; Shethi, K.; Buck, M. *Frontiers in plant science* **2014**, *5*, 242, epublish.

(164) Pinaud, L.; Ferrari, M. L.; Friedman, R.; Jehmlich, N.; von Bergen, M.; Phalipon, A.; Sansonetti, P. J.; Campbell-Valois, F.-X. *PloS one* **2017**, *12*, e0186920, epublish.

(165) Dallaire-Dufresne, S.; Barbeau, X.; Sarty, D.; Tanaka, K. H.; Denoncourt, A. M.; Lagüe, P.; Reith, M. E.; Charette, S. J. *Microbiology (Reading, England)* **2013**, *159*, 1937–1945, ppublish.

(166) Kühn, S.; Bergqvist, J.; Gil, M.; Valenzuela, C.; Barrio, L.; Lebreton, S.; Zurzolo, C.; Enninga, J. *Cell reports* **2020**, *31*, 107638, ppublish.

(167) Gehre, L.; Gorgette, O.; Perrinet, S.; Prevost, M.-C.; Ducatez, M.; Giebel, A. M.; Nelson, D. E.; Ball, S. G.; Subtil, A. *eLife* **2016**, *5*, e12552, epublish.

(168) Cassady-Cain, R. L.; Blackburn, E. A.; Alsarraf, H.; Dedic, E.; Bease, A. G.; Böttcher, B.; Jørgensen, R.; Wear, M.; Stevens, M. P. *The Journal of biological chemistry* **2016**, *291*, 5803–5816, ppublish.

(169) Li, S.; Zhang, L.; Yao, Q.; Li, L.; Dong, N.; Rong, J.; Gao, W.; Ding, X.; Sun, L.; Chen, X.; Chen, S.; Shao, F. *Nature* **2013**, *501*, 242–246, ppublish.

(170) Newson, J. P. M.; Scott, N. E.; Yeuk Wah Chung, I.; Wong Fok Lung, T.; Giogha, C.; Gan, J.; Wang, N.; Strugnell, R. A.; Brown, N. F.; Cygler, M.; Pearson, J. S.; Hartland, E. L. *Molecular & cellular proteomics : MCP* **2019**, *18*, 1138–1156, ppublish.

(171) El Qaidi, S.; Scott, N. E.; Hardwidge, P. R. *Scientific reports* **2021**, *11*, 3834, epublish.

(172) Dressaire, C.; Moreira, R. N.; Barahona, S.; Alves de Matos, A. P.; Arraiano, C. M. *mBio* **2015**, *6*, e02352–e02314, epublish.

(173) Niebuhr, K.; Giuriato, S.; Pedron, T.; Philpott, D. J.; Gaits, F.; Sable, J.; Sheetz, M. P.; Parsot, C.; Sansonetti, P. J.; Payrastre, B. *The EMBO journal* **2002**, *21*, 5069–5078, ppublish.

(174) Parsot, C. *Current opinion in microbiology* **2009**, *12*, 110–116, ppublish.

(175) Steele-Mortimer, O. *Current opinion in microbiology* **2008**, *11*, 38–45, ppublish.

(176) Hui, X.; Chen, Z.; Lin, M.; Zhang, J.; Hu, Y.; Zeng, Y.; Cheng, X.; Ou-Yang, L.; Sun, M.-A.; White, A. P.; Wang, Y. *mSystems* **2020**, *5*, DOI: 10.1128/mSystems.00288-20, epublish.

(177) Hochreiter, S.; Schmidhuber, J. *Neural Computation* **1997**, *9*, 1735–1780.

(178) Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate, 2016.

(179) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł. u.; Polosukhin, I. In *Advances in Neural Information Processing Systems*, ed. by Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Curran Associates, Inc.: 2017; Vol. 30.

(180) Weininger, D. *Journal of Chemical Information and Computer Sciences* **1988**, *28*, 31–36.

(181) Bonidia, R. P.; Sampaio, L. D. H.; Domingues, D. S.; Paschoal, A. R.; Lopes, F. M.; de Carvalho, A. C. P. L. F.; Sanches, D. S. *Briefings in bioinformatics* **2021**, *22*, DOI: 10.1093/bib/bbab011, ppublish.

(182) Marsella, L.; Sirocco, F.; Trovato, A.; Seno, F.; Tosatto, S. C. E. *Bioinformatics (Oxford, England)* **2009**, *25*, i289–i295, ppublish.

(183) Li, J.; Li, Z.; Luo, J.; Yao, Y. *Computational and mathematical methods in medicine* **2020**, *2020*, 3974598, epublish.

(184) Jing, R.; Wen, T.; Liao, C.; Xue, L.; Liu, F.; Yu, L.; Luo, J. *NAR genomics and bioinformatics* **2021**, *3*, lqab086, epublish.

(185) Ahuja, U.; Shokeen, B.; Cheng, N.; Cho, Y.; Blum, C.; Coppola, G.; Miller, J. F. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 2341–2348, ppublish.

(186) Francis, M. S.; Wolf-Watz, H.; Forsberg, A. *Current opinion in microbiology* **2002**, *5*, 166–172, ppublish.

(187) Jackson, M. W.; Silva-Herzog, E.; Plano, G. V. *Molecular microbiology* **2004**, *54*, 1364–1378, ppublish.

(188) Kopaskie, K. S.; Ligtenberg, K. G.; Schneewind, O. *The Journal of biological chemistry* **2013**, *288*, 35478–35488, ppublish.

(189) Auweter, S. D.; Bhavsar, A. P.; de Hoog, C. L.; Li, Y.; Chan, Y. A.; van der Heijden, J.; Lowden, M. J.; Coombes, B. K.; Rogers, L. D.; Stoynov, N.; Foster, L. J.; Finlay, B. B. *The Journal of biological chemistry* **2011**, *286*, 24023–24035, ppublish.

(190) Allombert, J.; Vianney, A.; Charpentier, X. *Methods in molecular biology (Clifton, N.J.)* **2017**, *1615*, 489–499, ppublish.

(191) Chakravarthy, S.; Huot, B.; Kvitko, B. H. *Methods in molecular biology (Clifton, N.J.)* **2017**, *1615*, 473–487, ppublish.

(192) Qureshi, S. A. *BioTechniques* **2007**, *42*, 91–96, ppublish.

(193) Boute, N.; Lowe, P.; Berger, S.; Malissard, M.; Robert, A.; Tesar, M. *Frontiers in pharmacology* **2016**, *7*, 27, epublish.

(194) Nyerges, Á. et al. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, E5726–E5735, ppublish.

(195) Sanders, B. R.; Townsend, S. E.; Ford, M. L.; Graves, J. L.; Thomas, M. D. *Journal of microbiological methods* **2023**, *204*, 106627, ppublish.

(196) Ellermeier, C. D.; Ellermeier, J. R.; Slauch, J. M. *Molecular microbiology* **2005**, *57*, 691–705, ppublish.

(197) Kim, K.; Golubeva, Y. A.; Vanderpool, C. K.; Slauch, J. M. *Molecular microbiology* **2019**, *111*, 570–587, ppublish.

(198) Cheng, S.; Wang, L.; Liu, Q.; Qi, L.; Yu, K.; Wang, Z.; Wu, M.; Liu, Y.; Fu, J.; Hu, M.; Li, M.; Zhou, D.; Liu, X. *Molecular & cellular proteomics : MCP* **2017**, *16*, 2219–2228, ppublish.

(199) Sarrion-Perdigones, A.; Chang, L.; Gonzalez, Y.; Gallego-Flores, T.; Young, D. W.; Venken, K. J. T. *Nature communications* **2019**, *10*, 5710, epublish.

(200) Dehoux, P.; Flores, R.; Dauga, C.; Zhong, G.; Subtil, A. *BMC genomics* **2011**, *12*, 109, epublish.

(201) Goodman, D. B.; Church, G. M.; Kosuri, S. *Science (New York, N.Y.)* **2013**, *342*, 475–479, ppublish.

(202) Tay, D. M. M.; Govindarajan, K. R.; Khan, A. M.; Ong, T. Y. R.; Samad, H. M.; Soh, W. W.; Tong, M.; Zhang, F.; Tan, T. W. *BMC bioinformatics* **2010**, *11 Suppl 7*, S4, epublish.

(203) Yu, L.; Liu, F.; Li, Y.; Luo, J.; Jing, R. *Frontiers in microbiology* **2021**, *12*, 605782, epublish.

(204) Wang, J.; Li, J.; Hou, Y.; Dai, W.; Xie, R.; Marquez-Lago, T. T.; Leier, A.; Zhou, T.; Torres, V.; Hay, I.; Stubenrauch, C.; Zhang, Y.; Song, J.; Lithgow, T. *Nucleic acids research* **2021**, *49*, D651–D659, ppublish.

Dedication

Für Nala und Janina.

Acknowledgements

First, I would like to thank Thomas Marlovits for giving me the opportunity to work on this exciting project, that allowed me to get to know a wide array of scientific disciplines from experimental lab work to the computational development of machine learning pipelines. It helped me refine my strengths and work on my weaknesses.

I would also like to thank Oliver Vesper for his ongoing support in this project, his vast experience and knowledge especially in wet-lab related topics was crucial to advance this project.

A special thank you goes to Barbara Grueter, who not only supported me tremendously by conducting many of the quantification assays but who always had an open ear for me and provided invaluable mental support. With her work during her masters thesis Jiline Ahrendt also contributed to this project. Finally, I would also like to thank past and present members of the Marlovits AG for their unobstructed view on things, helpful ideas and suggestions, as well as encouragements when needed. Thank you!

Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe. Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.