

Bowring, Janine Zara (2018) *The SaPI master repressor and the bacteriophage Duts: a tale of parasitism, evolution, and interspecific transfer*. PhD thesis.

<https://theses.gla.ac.uk/8904/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given



University of Glasgow

The SaPI master repressor and the bacteriophage Duts: a tale of parasitism, evolution, and interspecific transfer

A thesis submitted to the University of Glasgow for the degree of Doctor of
Philosophy

By

Janine Zara Bowring
BSc (Hons), MRes

Submitted on the 2nd of October 2017

Institute of Infection, Immunity and Inflammation
College of Medical, Veterinary and Life Sciences
University of Glasgow

Author's declaration

Unless explicitly stated otherwise, all of the experimental work and data analysis presented in this thesis was performed by the named author, either independently or in conjunction with others. This thesis was written in whole by the named author. No part of this research or any part of this document has or will be presented for the fulfilment of any other degree or qualification.

Acknowledgements

Firstly, I would like to thank my supervisor Professor José R Penadés for the opportunity to work in his group, for his good advice throughout my PhD, and for helping me become a better, more dedicated researcher. I would also like to thank Dr Alberto Marina and his group for their help and guidance on the structural side of things; it was lovely to collaborate with them. A big thank you to everyone in the Penadés group for always being willing to soundboard ideas and offer a hand. Thanks to the rest of Level 2, it wouldn't have been nearly as fun without you guys to share the long hours and Friday drinks with! A massive thank-you to my family and friends, your encouragement, love, support and offered ears have helped me through. Finally, a very special thank you to Mărtiņš, you helped this thesis happen with your constant love, support, reassurance, and praise.

Abstract

Staphylococcus aureus pathogenicity islands (SaPIs) form part of the wider family of phage inducible chromosomal islands (PICIs), which are extremely mobile phage satellites. SaPIs are clinically relevant, encoding and disseminating virulence and fitness factors to promote bacterial evolution and pathogenesis. These islands are innately linked to their helper phage's life cycle, requiring phage-mediated induction and packaging for transfer between bacterial hosts. The inherent association between SaPIs and helper phages is one of the best characterised examples of virus satellite interactions in prokaryotic cells. The SaPIs sit passively in the host chromosome, controlled by the SaPI-encoded master repressor protein; the Stl. Interaction between the SaPIs and their helper phages begins when a phage-encoded protein binds to the Stl to de-repress the island. Different SaPIs encode different Stl repressors, which require diverse phage-encoded proteins for derepression. The best characterised phage-encoded inducers are the trimeric dUTPases (Duts), which traditionally prevent the misincorporation of uracil into DNA, but have been proposed as regulatory proteins that induce SaPIbov1. Recently, the structurally dissimilar phage-encoded dimeric Dut of ϕ NM1 has also been shown to mobilize SaPIbov1, but how this is accomplished remains unsolved.

Here, this work expands the family of SaPIbov1 phage-encoded inducers to include the subset of phage-encoded dimeric Duts. This develops upon the theory that specific phage-proteins are required for SaPI induction, broadening this vision to show that SaPIs instead target conserved phage mechanisms. Using SaPIbov1 as an example, this study indicates the Stl_{SaPIbov1} can interact with the structurally distinct dimeric and trimeric Duts, whose only shared value is their role for the phage. Furthermore, this study highlights that, although these two proteins are completely unrelated, they both encode one highly variable region (named motif VI). Indeed, the mechanism for the dimeric Dut interaction with the Stl_{SaPIbov1} is investigated here and a number of parallels to the trimeric Dut mode of action are apparent. These results indicate that the structurally divergent dimeric and trimeric Duts interact with the SaPIbov1 Stl using analogous, but distinct mechanisms that represent a fascinating example of convergent evolution. Furthermore, investigations into the SaPIbov1 Stl indicate

that Stl accomplishes these interactions with the dimeric and trimeric Duts by using separate domains for each interaction. This intriguing Stl evolution is an elegant strategy utilised by the SaPIs to overcome phage-encoded inducer mutation or exchange. This highlights the continual arms race between the phages, which aim to escape SaPI induction, and the SaPIs, which must overcome these changes for island induction to occur. Furthermore, these results highlight the potential for Stl targeting of structurally diverse, functionally related phage proteins to facilitate intra- and inter-specific transfer. Overall, these results highlight the SaPIs as a fascinating subcellular parasite of the phages, which have evolved a novel strategy to target their phage inducer proteins and achieve transfer. Likewise, this study shows the biological significance of the dimeric Duts as regulatory molecules.

Table of contents

Author's declaration	ii
Acknowledgements	iii
Abstract	iv
Table of contents	vi
List of tables	ix
List of figures	x
List of accompanying material	xii
Abbreviations	xiii
Chapter 1 Introduction	1
1.1 <i>Staphylococcus aureus</i>	2
1.2 Bacteriophages	3
1.2.1 Background	3
1.2.2 Lysogenic conversion and transfer of phage-encoded virulence genes	4
1.2.3 Phage structure and taxonomy	8
1.2.4 The life cycle of a bacteriophage	10
1.2.5 Genetic and functional organisation of bacteriophages	12
1.2.6 Control of phage lysogeny	13
1.2.7 Phage integration and excision	14
1.2.8 Phage DNA replication	15
1.2.9 Packaging regulation systems	16
1.2.10 Packaging initiation, DNA recognition and capsid formation	16
1.2.11 DNA translocation machinery and packaging termination	21
1.3 Phage Inducible Chromosomal Islands	23
1.3.1 Overview of the genomic island family	23
1.3.2 Genome organisation and regulation of the SaPIs	26
1.3.3 SaPI-helper phage interactions	34
1.3.4 SaPIs as vehicles for horizontal intra- and intergeneric virulence gene transfer	43
1.3.5 Biology of non-SaPI PICI elements	45
1.4 The dUTPases, metabolic enzymes with a moonlighting signalling role	46
1.4.1 The enzymatic function of the Duts	47
1.4.2 Types of Dut	49
1.4.3 Moonlighting signalling roles of Duts	51
1.4.4 Phage-encoded trimeric Duts as SaPI inducers	54
1.4.5 Phage-encoded dimeric Duts as SaPI inducers	58
1.5 Aims	61
Chapter 2 Materials and methods	64
2.1 Materials	65
2.1.1 Chemicals	65
2.1.2 Enzymes	65
2.2 Bacterial strains and vectors	65
2.2.1 Bacterial strains	65

2.2.2 Storage and growth	65
2.2.3 Plasmid vectors	66
2.3 General molecular biology techniques	75
2.3.1 Chromosomal DNA extraction	75
2.3.2 Plasmid DNA extraction	75
2.3.3 Polymerase chain reaction	75
2.3.4 Klenow Fill-in reactions	76
2.3.5 Agarose gel electrophoresis	77
2.3.6 Restriction digest	77
2.3.7 Gel extraction and PCR purification	77
2.3.8 Ligation reactions	78
2.3.9 Precipitation of genomic DNA	78
2.3.10 Overexpression of pPROEX-HTa vectors	79
2.3.11 Southern blot	79
2.3.12 Western blot	80
2.4 Bacterial transformation.....	80
2.4.1 Preparation of chemically competent <i>E. coli</i> cells	80
2.4.2 Transformation of chemically competent <i>E. coli</i> cells	81
2.4.3 Preparation of electrocompetent <i>S. aureus</i> cells	81
2.4.4 Transformation of electrocompetent <i>S. aureus</i> cells	82
2.5 Biochemical methods	82
2.5.1 Enzyme assay for the quantification of β -lactamase activity in transcriptional fusion plasmids	82
2.5.2 Protein purification	83
2.5.3 Mass spectrometry	83
2.6 Induction of PICIs and phage.....	85
2.6.1 Phage propagation: induction.....	85
2.6.2 Phage propagation: infection.....	85
2.6.3 Bacteriophage titring assay	86
2.6.4 Transduction titring assay	86
2.7 Allelic exchange	87
2.7.1 Point and Deletion mutagenesis using pMAD	87
2.7.2 Introduction of a 3 x FLAG-tag into lysogenic phage DNA	92
2.8 Data analysis	94
2.8.1 Statistical analysis	95
2.8.2 Structural modelling of the wild-type and mutant dimeric Duts	95
2.8.3 Structural modelling of the SaPI _{bov1} Stl.....	95
Chapter 3 The dimeric Duts are a subset of phage-encoded SaPI inducers.....	96
3.1 Background	97
3.2 Results	99
3.2.1 The phage encoded dimeric and trimeric Duts lack homology in sequence and structure	99
3.2.2 Characterising the different families of dimeric Duts.....	101
3.2.3 Induction of SaPIs by phage-encoded dimeric Duts	108
3.2.4 SaPI induction is Dut specific and interferes with phage transfer	113

3.2.5 SaPI induction is triggered by a direct interaction between the dimeric Dut and the Stl repressor	117
3.3 Discussion	119
3.4 Conclusions and future work.....	122
Chapter 4 Characterisation of the dimeric Duts interaction with SaPIbov1	124
4.1 Background	125
4.2 Results	126
4.2.1 The structures of the Φ O11 and Φ DI encoded dimeric Duts	126
4.2.2 The role of the dimeric Duts unconserved extra domain (motif VI) in SaPIbov1 induction	130
4.2.3 The role of the dimeric Duts enzymatic function in SaPIbov1 induction	132
4.2.4 The role of the dimerization interface of the dimeric Duts.....	136
4.3 Discussion	144
4.4 Conclusions and future work.....	147
Chapter 5 Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts	149
5.1 Background	150
5.2 Results	152
5.2.1 Identifying SaPIbov1 Stl repressor homologues in different staphylococcal species ...	152
5.2.2 Pirating conserved phage processes supports inter-specific SaPI transfer.....	158
5.3 Discussion	165
5.4 Conclusions and future work.....	168
Chapter 6 Characterisation of the SaPIbov1 Stl repressor	170
6.1 Background	171
6.2 Results	172
6.2.1 The predicted structure of the SaPIbov1 Stl repressor	172
6.2.2 Identifying the SaPIbov1 Stl residues responsible for DNA recognition for SaPI repression.....	175
6.2.3 Establishing the SaPIbov1 Stl residues involved in repressor dimerization	180
6.2.4 Determining SaPIbov1 Stl residues that interact with the phage-encoded trimeric Duts	184
6.3 Discussion	189
6.4 Conclusions and future work.....	191
Chapter 7 Concluding remarks	193
7.1 The dimeric Duts as a subset of SaPIbov1 inducers	194
7.2 SaPIs exploit conserved phage mechanisms to achieve efficient intra- and interspecific transfer	197
7.3 The SaPIbov1 Stl: a multi-domain protein targeting the conserved phage dUTPase mechanism for derepression.....	199
7.4 Future work.....	203
Appendices	207
Glossary	243
List of references	245
Publications related to this work.....	269

List of tables

Table 1.1 A selection of toxins and virulence factors encoded by phages.....	6
Table 1.2 Virulence and accessory genes encoded by the SaPIs and SaPI-like elements.....	34
Table 2.1 Bacterial strains used in this study	68
Table 2.2 Plasmids used in this study	72
Table 3.1 The transfer of SaPIbov1 and SaPIbov5 by dimeric Duts Φ D1 and Φ NM1.....	115
Table 3.2 Interference of SaPIbov1 and SaPIbov5 with phage Φ D1 and Φ NM1 titres.....	116
Table 5.1 Phage-inducible islands containing a Stl-like repressor homologous to the SaPIbov1 Stl repressor, as identified in this study.....	153
Table 5.2 Known prophages in the Staphylococcal species encoding SaPIbov1 Stl homologues and the presence or absence of a phage-encoded Dut in that phage genome.....	162
Table 6.1 List of point mutations shown in Fig 6.3 and predicted to affect SaPIbov1 Stl DNA recognition, as cloned into the SaPIbov1 Stl of transcriptional fusion vector pJP2085..	178
Table 6.2 List of point mutations shown in Fig 6.4 and predicted to affect SaPIbov1 Stl repressor dimerization, as cloned into the SaPIbov1 Stl of transcriptional fusion vector pJP2085..	183
Table 6.3 List of point mutations shown in Fig 6.5 and predicted to affect Stl repressor interaction with the trimeric Duts, as cloned into the SaPIbov1 Stl of transcriptional fusion vector pJP2085..	187

List of figures

Figure 1.1 Prototypical bacteriophage structures. (A) <i>Myoviridae</i> , T4; (B) <i>Siphoviridae</i> , λ ; (C) <i>Podoviridae</i> , P22.....	9
Figure 1.2 Life cycles of the bacteriophage	11
Figure 1.3 Schematic representation of <i>Siphoviridae</i> Staphylococcal phage genome modules	13
Figure 1.4 Schematic representation of virus particle assembly of dsDNA phages.	18
Figure 1.5 The classical <i>cos</i> and <i>pac</i> models for cleavage and packaging of phage dsDNA concatemers.....	20
Figure 1.6 Comparison of phage inducible chromosomal island genomes	27
Figure 1.7 SaPI transcriptional module and transcriptional control.	29
Figure 1.8 Staphylococcal pathogenicity island replication scenarios.	31
Figure 1.9 SaPI induction by a helper phage..	36
Figure 1.10 Electron micrographs of a mixed helper phage-SaPI lysate.....	38
Figure 1.11 Mechanisms of SaPI interference with phage reproduction.	40
Figure 1.12 The classic enzymatic role of the dUTPases.....	48
Figure 1.13 Dut types and their active sites.	49
Figure 1.14 Monomer structures of trimeric Dut proteins and localisation of the extra motif VI	52
Figure 1.15 Spatial location in Dut80 α^{WT} structure of relevant areas and residues for Dut–Stl interaction.....	55
Figure 1.16 Model of <i>Staphylococcus aureus</i> phage Dut interaction with SaPIs Stl	57
Figure 1.17 The known structures of dimeric Duts..	59
Figure 2.1 A schematic representation of the double PCR necessary for the construction of deletion mutant homology arms.....	89
Figure 2.2 A schematic representation of the double PCR necessary for the construction of point mutation homology arms.....	90
Figure 2.3 A schematic representation of pMAD mediated mutagenesis..	91
Figure 2.4 A schematic representation of the triple PCR necessary for the construction of the 3 x FLAG-tag with homology arms.....	93
Figure 2.5 A schematic representation of pMAD mediated 3 x FLAG-tag insertion..	94
Figure 3.1 The dimeric and trimeric Duts have largely different protein sequences.....	100
Figure 3.2 The dimeric and trimeric Duts have completely different predicted structures.....	101
Figure 3.3 Radial phylogenetic tree of identified <i>Staphylococcus</i> phage dimeric Duts.	103
Figure 3.4 Alignment of <i>Staphylococcus aureus</i> phage encoded dimeric Duts shows allelic variants.....	105
Figure 3.5 Comparison of the predicted structures of each phage-encoded dimeric Dut group representative and the solved structure of the <i>Campylobacter jejuni</i> – encoded dimeric Dut.....	107
Figure 3.6 SaPIbov1 and SaPIbov5 replication is induced by different dimeric Duts variants..	109
Figure 3.7 Phylogram of the representative Dut protein sequences of the 9 different dimeric Dut families.	111
Figure 3.8 Pairwise comparisons of the representative Duts for each of the 9 families..	112
Figure 3.9 SaPIbov1 excision and replication following induction of Φ NM1 and SaPIbov5 replication following induction of Φ DI.....	114

Figure 3.10 Protein expression of Dut variants leading to the release of Stl repression of downstream genes.....	118
Figure 3.11 Affinity chromatography of the Φ NM1 and Φ O11 Duts for His-tagged SaPIbov1 Stl.....	119
Figure 4.1 Structures of the Φ DI and Φ O11 encoded dimeric Duts.	127
Figure 4.2 The active sites of the phage encoded dimeric and trimeric Duts do not share any overlapping residues.	130
Figure 4.3 The divergent motif VI is not responsible for the SaPIbov1 inducing capabilities of the dimeric Duts..	132
Figure 4.4 The Φ DI A73L Dut mutant is incapable of inducing SaPIbov1	135
Figure 4.5 A second region containing sequence divergence impacts on SaPIbov1 induction capability.	137
Figure 4.6 The central residue of the dimerization interface plays a role in SaPIbov1 induction..	140
Figure 4.7 The dimerization interface plays a role in SaPI induction, but is not solely responsible for the induction of SaPIbov1.	143
Figure 5.1 Protein alignment of the SaPIbov1 Stl repressor and different homologues from various Staphylococcal species.....	154
Figure 5.2 Phylogram of the SaPIbov1 Stl repressor and different homologues from various Staphylococcal species.....	156
Figure 5.3 Pairwise comparisons of the SaPIbov1 Stl repressor and 6 Stl homologues.....	157
Figure 5.4 Protein expression of trimeric and dimeric Dut variants leading to the release of Stl repression of downstream genes in SaPIbov1 and SaPIbov1 Stl repressor homologue regulatory modules.....	160
Figure 5.5 Protein alignment of (A) ϕ 11- and ϕ Shaem1-encoded trimeric Duts (B) ϕ NM1- and ϕ StB12-encoded dimeric Duts (C) ϕ NM1-encoded dimeric Dut and ϕ 575- and ϕ 879-encoded MazG family proteins.	164
Figure 6.1 Predicted structures of the SaPIbov1 Stl repressor protein constructed by the I-TASSER and Phyre2 servers	173
Figure 6.2 GeneSilico Metadisorder server protein sequence prediction of any disordered regions in the SaPIbov1 Stl repressor	174
Figure 6.3 Schematic of the pJP2085 transcriptional fusion plasmid containing the regulatory module of the SaPIbov1, including part of the <i>integrase</i> gene, the <i>stl</i> repressor, the <i>str</i> anti-repressor, and a partial <i>xis</i> fused to a <i>blaZ</i> reporter gene.	175
Figure 6.4 Identifying residues responsible for DNA recognition in the SaPIbov1 Stl repressor...	177
Figure 6.5 Identifying residues responsible for repressor dimerization in the SaPIbov1 Stl repressor..	182
Figure 6.6 Identifying residues responsible for Dut interaction in the SaPIbov1 Stl repressor.....	186

List of accompanying material

Appendix 1 List of primers used in this study.....	207
Appendix 2 Templates and confidence values of phage encoded dimeric Dut models generated by I-TASSER for Fig 3.12	221
Appendix 3 The 59 phage encoded dimeric Duts used to construct Fig 3.	222
Appendix 4 CLC alignment used to construct Fig 3.9.....	233
Appendix 5 CLC alignment used to construct Fig 3.10 and 3.11.....	234
Appendix 6 Evolved phages that grow in presence of SaPI	235
Appendix 7 Crystallographic statistics for Φ DI-dUPNPP and Φ O11-dUPNPP structures.....	236
Appendix 8 Predicted structural interference in Φ DI ^{Φ55α2,α6} and Φ 55 ^{ΦDIα2,α6}	238
Appendix 9 Templates and confidence values for Φ DI ^{Φ55α2,α6} and Φ 55 ^{ΦDIα2,α6} generated by I-TASSER for Appendix 8	239
Appendix 10 Conserved motifs identified across different SaPI intergenic regions.....	240
Appendix 11 CLC alignment used to construct Fig 5.2 and 5.3.....	241
Appendix 12 Templates and confidence values for predicted SaPIbov1 StI generated by I-TASSER and Phyre2	242

Abbreviations

ϕ	Phage
Δ	Deletion
%	Percent
$^{\circ}\text{C}$	Degrees Celsius
μg	Microgram(s)
μl	Microliter(s)
μm	Micrometre(s)
μM	Micromolar
amp	Ampicillin
ANOVA	Analysis of variance
<i>att</i>	Attachment site
<i>att</i> ϕ	Attachment site in the circular intermediate of the phage
<i>att</i> _C	Attachment site in the bacterial chromosome
<i>att</i> _L	Left attachment site
<i>att</i> _R	Right attachment site
<i>att</i> _S	Attachment site in the circular intermediate of the SaPI
<i>bla</i> _Z	β -lactamase gene
bp	Base pair(s)
Bulk	Majority DNA (genomic + DNA phage/SaPI)
CaCl_2	Calcium chloride
CCC	Covalently closed circular SaPI DNA
CFU	Colony forming units
DNA	Deoxyribonucleic acid
dsDNA	Double stranded deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
dUPNPP	α,β -imido-dUTP
Dut	dUTPase
dUTP	Deoxyuridine triphosphate
dTTP	Deoxythymidine triphosphate
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
ERP	Excision-replication-packaging cycle
Erm	Erythromycin
<i>ermC</i>	Erythromycin resistance gene C
g	Gram(s)
GI(s)	Genomic island(s)
h	Hour(s)
HGT	Horizontal gene transfer
HPLC	High-performance liquid chromatography
ICTV	International Committee on Taxonomy of Viruses

IPTG	Isopropyl- β -D-thiogalactoside
kb	Kilo base pairs
KSHV	Karposi's Sarcoma-associated Herpesvirus
kV	Kilovolt(s)
LB	Luria-Bertani medium
M	Molar
MC	Mitomycin C
MgCl ₂	Magnesium chloride
MGEs	Mobile genetic elements
MgSO ₄	Magnesium sulphate
min	Minute(s)
ml	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar
MW	Molecular weight
NaCl	Sodium chloride
NCBI	National Centre for Biotechnology Information
ng	Nanogram(s)
nM	Nanomolar
nt	Nucleotide
OD _{xxx}	Optical density at a wavelength XXX nanometres
ORF	Open reading frame
<i>ori</i>	Origin of replication
P	Promoter
PB	Phage base
PBA	Phage base agar
<i>Pcad</i>	Cadmium inducible promoter
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PICIs	Phage inducible chromosomal islands
PLE	PICI-like element
PPAR	Peroxisome proliferator-activated receptor
RBS	Ribosomal binding site
Rpm	Revolutions per minute
SAD	Single-atom detection
SaPIs	<i>Staphylococcus aureus</i> pathogenicity islands
SCC <i>mec</i>	Staphylococcal cassette chromosome <i>mec</i>
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SOS	Stress bacterial response
TerS	Terminase small subunit
TerS ϕ	Phage terminase small subunit

TerS _{SP}	SaPI terminase small subunit
TerL	Terminase large subunit
Tet	Tetracycline
<i>tetM</i>	Tetracycline resistance gene M in gram-positive bacteria
Tris	Tris-(hydroxymethyl) aminomethane
TSST	Toxic shock syndrome toxin
TSA	Tryptic soy agar medium
TSB	Tryptic soy broth
TTC	2,3,5-triphenyltetrazolium chloride
U	Unit
UV	Ultraviolet
V	Volt(s)
wt	Wild type
Xgal	5-Bromo-4-chloro-3-indole- β -D-galactopyranoside

Chapter 1 Introduction

1.1 *Staphylococcus aureus*

Staphylococcus aureus is a non-motile Gram-positive bacterium that is usually coagulase and catalase positive, but oxidase negative. As a member of the *Firmicutes* phylum, *S. aureus* grows in spherical shaped cocci and groups in irregular clusters. *S. aureus* can behave as both a commensal bacterium and a clinically relevant opportunistic pathogen. This bacterium colonises the nares, skin and perineum of humans and has been isolated from several different animals¹⁻⁴. Although *S. aureus* can colonise hosts as a stable commensal, it is a major opportunistic pathogen in man and an important veterinary pathogen. *S. aureus* can cause severe infectious disease and the prevalence of numerous drug-resistant strains has established this species of bacteria as a major healthcare and veterinary burden.

S. aureus can colonise the human nose and persist as commensal bacteria, with 20-30% of the population living as asymptomatic carriers⁵⁻⁷. Asymptomatic carriers of *S. aureus* suffer an increased risk of staphylococcal infection following cutaneous or mucosal injury, and carriage has been implicated in an increased risk of infection following hospital stays or operations⁸⁻¹². Following damage to the cutaneous or mucosal membranes, the microorganism is able to penetrate the underlying tissues leading to skin and soft tissue infections (SSTIs)^{13,14}. As well as causing SSTIs, *S. aureus* can invade the blood, causing bacteraemia¹⁵⁻¹⁷ and eventually leading to severe infections of the organs including endocarditis^{18,19}, pneumonia^{9,20}, septic arthritis^{21,22}, osteomyelitis^{23,24}, and severe sepsis^{25,26}. Furthermore, *S. aureus* can cause a number of toxin-related diseases, including food poisoning²⁷, toxic shock syndrome^{28,29}, necrotic infections³⁰, and scalded skin syndrome³¹. *S. aureus* also causes important diseases in animals, including dermatitis in rabbits³², skeletal infections in poultry⁴, and mastitis in ruminants^{33,34}.

The incidences of *S. aureus* infection in both hospital and community settings have increased over the past few decades^{35,36}. Accompanying this increase in occurrence, this bacterium has also evolved new virulent and drug-resistant variants^{37,38}. Virulence factors allow the bacteria to colonise a range of different tissues and hosts by facilitating mechanisms for host adaptations^{3,4,34,39}. Over the past decades, increasing resistance to antibiotics has led to the emergence of

vancomycin-resistant *S. aureus*^{40,41} and meticillin-resistant *S. aureus* (MRSA)^{42,43}, through the acquisition of the *vanA* cluster and staphylococcal cassette chromosome *mec* (SCC*mec*), respectively. Two new groups of MRSA with enhanced pathogenicity have recently emerged, healthcare-associated MRSA^{17,44} and community-associated MRSA^{45,46}. Healthcare-associated MRSA isolates carry a relatively large SCC*mec* gene and are resistant to all β -lactam and fluoroquinolone antibiotics, although they seldom carry Panton-Valentine leucocidin genes^{47,48}. Panton-Valentine leucocidin genes encode a cytotoxin involved in severe necrotic and haemorrhagic infections. On the other hand, community-associated MRSA isolates carry smaller SCC*mec* elements, but tend to carry Panton-Valentine leucocidin genes and occur in previously healthy younger patients^{49,50}. The infection ability of *S. aureus* is also increased by fitness factors, such as colonisation⁵¹, immune evasion⁵², bacterial multiplication and dissemination, which are expressed at different infection stages and contribute to the rise of *S. aureus* as a superbug⁵³.

One of the mechanisms through which the different pathogenic clones of *S. aureus* have been able to adapt and evolve is through the horizontal transfer of mobile genetic elements (MGEs)^{44,54}. MGEs encode genes to promote independent mobilisation between host genomes⁵⁵. *S. aureus* MGEs also encode important virulence and fitness-related genes, such as toxins, adhesins and invasins^{56,57}. Approximately 20-22% of the *S. aureus* genome is made up of MGEs and non-mobile genomic islands (GIs), including bacteriophages, pathogenicity islands, plasmids, transposons, and SCC*mec*^{54,57-60}. The ability of *S. aureus* to horizontally acquire MGEs encoding virulence and fitness factors enables this bacterium to adapt to new or changing environmental scenarios.

1.2 Bacteriophages

1.2.1 Background

Bacteriophages, or phages, are the most abundant form of MGE. They are viruses of bacteria that infect and replicate within bacterial cells. Phages have been found in all bacteria-containing environments, from the marine and soil ecosystems to human and animal microbiomes⁶¹⁻⁶⁶. They are predicted to reside environmentally at a level of 1-2 orders of magnitude higher than that of their

bacterial hosts. Indeed, they are the most abundant biological entity, with an estimated 10^{31} tailed phages in Earth's biosphere⁶⁷.

Phages shape the bacterial host cell for their own purpose using self-encoded proteins⁶⁸, whilst mediating horizontal gene transfer (HGT) within bacterial populations⁶⁹. The impact of phages on their surrounding ecosystem is far-reaching, influencing the carbon cycle of soil and water environments, the availability of global nutrients, and transfer of a genomic DNA pool to different biomasses⁷⁰⁻⁷². This is on top of the direct effects of the phage-bacteria interaction. For example, phages can confer encoded virulence traits to bacteria, enabling the evolution of human pathogens, or they can act as vectors for bacterial fitness and virulence genes through lysogeny or direct transfer of other MGEs^{44,69,73,74}.

As well as facilitating the direct supply of phage- or MGE-encoded virulence genes to their bacterial hosts, phages appear to have been utilised by the bacteria for the acquisition of beneficial phage-like machinery. For example, R-type pyocins and the type VI secretion system (T6SSs) may have been evolutionarily assimilated by bacteria from phage, as both are structurally similar to the phage-tail structure⁷⁵. These elements increase the virulence of the bacteria encoding them, with the T6SS utilising its tail-like structure to transport effector proteins by injection through the outer membrane of other cells⁷⁶. The bacterial-encoded tail structures of the R-type pyocins are released on cell lysis, binding and killing other bacteria⁷⁷. Collectively, the impact of phages on bacterial evolution and pathogenicity, together with the wider impact upon surrounding ecosystems, indicates the significant role phages play in bacterial warfare and niche colonisation.

1.2.2 Lysogenic conversion and transfer of phage-encoded virulence genes

Temperate phages can form a stable association with bacterial hosts, known as lysogeny. During lysogeny the viral genome of a temperate phage can be integrated into the bacterial chromosome. This integration of temperate phages can affect the genome of the bacterial host, impacting on genetic variability between bacterial species and strains⁷⁸. For example, the *E. coli* Sakai

pathogenic strain O157:H7 contains 18 prophages, and phages make up ~10% of the genome of *Streptococcus pyogenes*⁷⁸. Many phages have also been identified in the *S. aureus* genome⁷⁹. Increasing numbers of phage variants are being identified within bacteria by 'omics' sequencing, helped by improved annotation of phage genes. Phage acquisition is, therefore, an important source of genome variation for virulent strains^{79,80}. Additionally, lysogenic conversion of phages provides the bacterial host with fitness and virulence factors that can be utilised to colonise a niche. These integrated phage-encoded factors are a significant contributor to bacterial pathogenesis^{74,81}.

Phage-encoded virulence factors range from toxins to invasion factors (Table 1.1)^{78,82}. They include superantigens, adhesion factors, DNases, mitogenic factors, invasion factors, and immune evasion factors. Certain toxin-mediated diseases (toxinoses) are caused by toxins encoded by phages, for example botulism, cholera, diphtheria and scarlet fever⁵⁷. *S. aureus* phage virulence genes include enterotoxin A (SEA), which causes food poisoning, staphylokinase toxin (Sak) and Panton-Valentine leucocidin⁸³⁻⁸⁵.

In some cases, expression and release of toxins, as well as their mobilisation, is linked to the lytic cycle of the phage. Linkage of toxin expression with the phage lytic cycle has clinical implications for the treatment of prophage-containing (chromosomally-integrated phage containing) pathogenic bacteria strains. This is because treatment with some antibiotics can trigger the SOS response and initiate the phage lytic cycle. Induction of the phage, in turn, leads to upregulation of toxin expression, and phage-mediated cell lysis would release those toxins from the bacterial cell. One example of toxin expression linked to a phages lytic cycle is the Shiga toxin (Stx) encoded by *E. coli* phages. Treatment of Shiga toxin-producing *E. coli* with bacteriophage-inducing antibiotics, such as fluoroquinolones, increases the risk of haemolytic uremic syndrome⁸⁶. Furthermore, *S. aureus* enterotoxins A, G, and K, as well as Sak have all been linked to expression following phage induction⁷⁴. In this way, lysogenic phages not only transfer toxins to the bacterial genome, but can be involved in the expression of toxins following induction of the phage lytic cycle.

Table 1.1 A selection of toxins and virulence factors encoded by phages.

Protein	Gene	Phage	Bacteria
Diphtheria toxin	<i>tox</i>	B-Phage	<i>Corynebacterium diphtheriae</i>
Neurotoxin	C1	Phage C1	<i>Clostridium botulinum</i>
Shiga toxins	<i>stx1, stx2</i>	H-19B	<i>Escherichia coli</i>
Enterohaemolysin	<i>hly2</i>	φFC3208	<i>E. coli</i>
Cytotoxin	<i>ctx</i>	φCTX	<i>Pseudomonas aeruginosa</i>
Enterotoxins	<i>see, sel</i>	NA	<i>Staphylococcus aureus</i>
P	<i>sep</i>	φN315	<i>S. aureus</i>
A	<i>entA, sea</i>	φ13, φMu50A	<i>S. aureus</i>
Toxin types A and C	<i>speA, speC</i>	T12, CS112	<i>Streptococcus pyogenes</i>
Cholera toxin	<i>ctxAB</i>	CTXφ	<i>Vibrio cholerae</i>
Leukocidin	<i>pvl</i>	fPVL	<i>S. aureus</i>
Superantigens	<i>speA1, speA3, speC, speI, speH, speM, speL, speK, ssa</i>	8232.1	<i>S. pyogenes</i>
Cytolethal distending toxin	<i>cdt</i>	Unnamed	<i>E. coli</i>
Type III effectors	<i>sopE, sseI (gtgB), sspH1</i>	SopEΦ, GIFSY-2, GIFSY-3	<i>Salmonella enterica</i>
Superoxide dismutases	<i>sodC</i>	Sp4, 10	<i>E. coli</i>
	<i>sodC-I, sodC-III</i>	GIFSY-2, Fels-1	<i>S. enterica</i>
	<i>mf2, mf3, mf4</i>	370.1, 370.3, 315.3	<i>S. pyogenes</i>
Mitogenic factors	<i>toxA</i>	Unnamed	<i>Pasteurella multocida</i>
	<i>unnamed</i>	Phisc 1	<i>Streptococcus canis</i>
OMP	<i>bor, eib</i>	λ, λ-like	<i>E. coli</i>
O-antigen acetylase	<i>oac</i>	Sf6	<i>Shigella flexneri</i>

Another effect of lysogenic phages on the bacterial host comes from the actual integration of the phage into the bacterial chromosome. Prophages (integrated phages) can alter the phenotype of the host bacterium if the phages are integrated into a coding region, leading to disruption of host genes^{87,88}. This host phenotype switching through phage integration is known as negative lysogenic conversion. One example of this is seen when staphylococcal phages integrate

into β -haemolysin⁸⁴ and lipase⁸⁹ genes, inactivating them in the process. Insertion of prophages into coding regions is also seen in *Shigella* spp, leading to inactivation of the lysine decarboxylase *cadA* gene and increased virulence⁹⁰. Similarly, temperate phages can act as an anchor point for genomic rearrangements^{78,88}.

In addition to genome alteration, prophages protect the host bacteria against infection by other lytic phages⁹¹⁻⁹³. Conversely, bacteria can utilise the presence of prophages in niche competitors to kill these rival bacteria by producing phage-inducing substances⁹⁴. From a human and animal host standpoint, it is possible that phages can confer protection from bacteria in certain situations. A recent study indicated that phages can adhere to mucosal surfaces, providing metazoan hosts with non-host immunity protection from bacterial colonisation and infection⁹⁵. In this way, phages have an influence on both their bacterial hosts and the surrounding niche.

The direct transfer of phage-encoded genes to bacterial hosts through prophage integration forms only part of the HGT enabled by phages, with phages able to transfer bacterially-encoded genomic DNA and MGEs. Phage-mediated transfer of bacterial MGEs plays a vital role in bacterial biology, divergence, and evolution^{44,69}. Generalized transduction forms the best studied and most widespread method by which phages transfer bacterial DNA. In generalized transduction, bacterial DNA is packaged into the phage procapsid enabling transfer to a new bacterial cell. First identified in *Salmonella enterica* serovar *Typhimurium* and *E. coli*, generalized transduction occurs at a relatively low frequency (every $\sim 10^8$ phage infections), but given how numerous phage particles are this makes it a common event in nature⁹⁶⁻⁹⁸. Important genomic and pathogenic islands are transferred between bacteria by generalized transduction, including the *SCCmec* and the *V. cholera* pathogenicity island^{60,98}.

Phages can also transfer parts of the bacterial chromosome through specialised transduction. Specialised transduction occurs when a prophage excises imprecisely from the integration site of the bacterial host, taking the flanking regions of the bacterial chromosome with it. These flanking regions are then packaged with the phage genome and transferred to a new cell as a phage particle^{99,100}.

Additionally, phages are linked to MGEs that can act as phage satellites, including the *S. aureus* pathogenicity islands (SaPIs), the RS1 element of *V. cholera* and the satellite phage P4¹⁰¹⁻¹⁰³. These satellites utilise different mechanisms to take advantage of phage machinery and facilitate their own transfer using adapted phage particles.

It has recently become apparent that levels of phage-mediated gene transfer were previously underestimated. Studies have identified the transfer of ever more bacterial genomic regions and phageomes (DNA encapsidated in phage particles), including the transfer of antibiotic resistance genes^{104,105}. One study found a link between antibiotic treatment of mice and an increase in antibiotic resistance related genes in the phageomes of the gut microbiome¹⁰⁶. Indeed, metagenomic analyses into the human virome found a huge number of phage genomes, containing not just lytic and lysogenic genes but also CRISPR and antibiotic resistance genes¹⁰⁷. The extent of antibiotic resistance gene transfer by phages is contentious, with the thresholds used for viral metagenomic analyses impacting on the level of phage-associated resistance genes seen¹⁰⁸. However, the interplay of bacteria and phages in the human gut microbiome has recently been summarised, with the review calling for more attention to be given to the most abundant biological entities in the human gut, the phages⁶⁴.

Overall, phages are heavily involved in the transfer of virulence and fitness factors through a variety of mechanisms. This highlights phages as an important bacterial MGE, contributing to the emergence of novel pathogenic strains of bacteria. Furthermore, the ability of phages to transfer bacterial fitness genes and impact on the surrounding ecosystems shows the role of these phages in bacterial niche colonisation.

1.2.3 Phage structure and taxonomy

Phage taxonomy has traditionally followed a classification scheme devised by the International Committee on Taxonomy of Viruses (ICTV), which grouped them by family, genus or host range in alphabetical or numerical order. In this system the viruses are grouped based on shared characteristics, for example tail-type and genome configuration and size¹⁰⁹. However, as the availability and ease of whole genome sequencing has increased, the lack of grouping based on genome and

proteome characterisations has led to some shortcomings of the classification systems becoming apparent. From the number of complete phage genomes in NCBI GenBank, 40% are only classified to the level of family, whilst a further 10% are completely unclassified¹¹⁰. This has led to alternative methods of classification being proposed, based on proteomic analysis or functional-structural domains^{111,112}. The classifications described in the following text follow the traditional ICTV guidelines.

Most phages are composed of a multipart head-tail structure. Conventionally, the tailed-phages share certain structural properties, including an icosahedral capsid and a helical tail attached to a hexagonal baseplate structure. A mature virion particle is formed once these components have been independently assembled and joined together. These phages belong to the Caudovirales order and contain double-stranded, linear DNA (dsDNA)¹¹³. Within the Caudovirales the phages are further classified into one of three families: *Myoviridae*^{114,115}, *Siphoviridae*^{116,117}, and *Podoviridae*¹¹⁸. Phage family classifications differ in their tail structure (Figure 1.1). The *Myoviridae* family have long contractile tails (*E. coli* phage P2 and T4), the *Siphoviridae* or “ λ -like viruses” have long non-contractile tails (most *S. aureus* phages, *E. coli* HK97, and *Bacillus subtilis* SPP1), and the *Podoviridae* or “P22-like viruses” have short, non-contractile tails (P22, Φ 29, and T7).

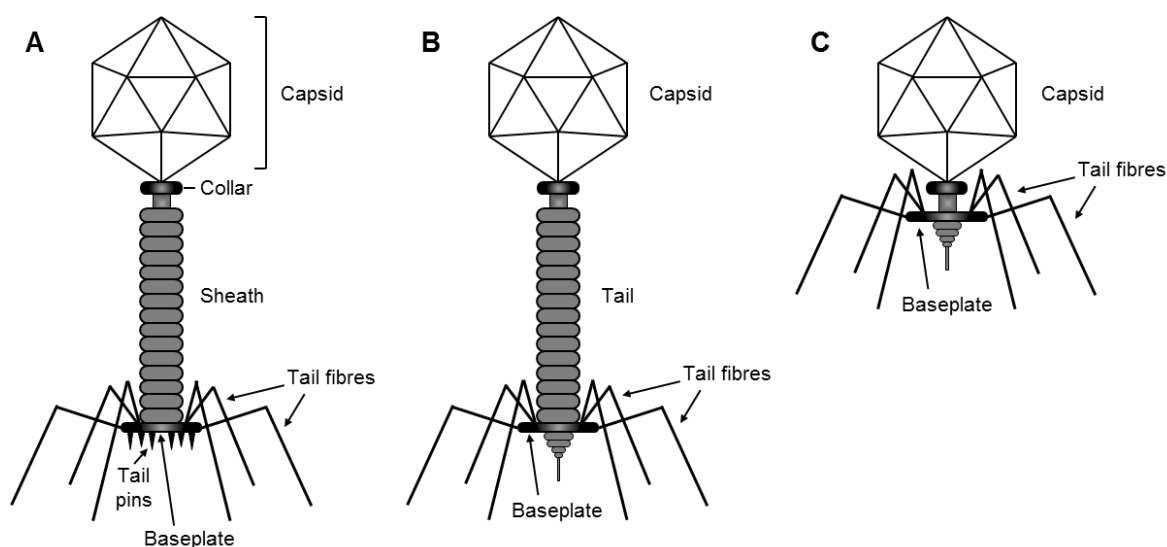


Figure 1.1 Prototypical bacteriophage structures. (A) *Myoviridae*, T4; (B) *Siphoviridae*, λ ; (C) *Podoviridae*, P22. Adapted from Trun et al. 2009¹¹⁹.

1.2.4 The life cycle of a bacteriophage

Phages can be classified into two groups based on the life cycles they undergo, either strictly lytic phages or temperate phages that can also go through the lysogenic cycle. In the lytic cycle, infection of a phage into a recipient bacterial cell or induction of prophage in a lysogenic strain leads to phage replication using host machinery¹²⁰. Assembly of mature virions and lysis of the bacterial cell releases phage progeny to the environment where they can go on to infect a new phage-sensitive recipient. Key to phage infection is the adsorption of the phage on to specific bacterial cell surface receptors, followed by injection of phage DNA into the recipient cell cytoplasm, where transcription and replication occur¹²⁰. Viral proteins are then synthesised and assembled into capsids (heads) and tails, before mature virions are formed. The phage DNA is packaged into the capsids. The phage holin protein forms pores in the cytoplasmic membrane so that phage-encoded endolysin can reach and hydrolyse the bacterial cell peptidoglycan layer, resulting in cell lysis (Figure 1.2)^{121,122}. Within the lytic cycle the different stages are controlled in a temporal pattern, to ensure correctly ordered expression of lytic cycle-associated genes. Expression begins with replication genes, then packaging genes, and finally lysis genes.

Temperate phages can also go through the lytic cycle, in the same manner as the strictly lytic phages. However, they can also enter a non-lytic lifestyle, known as the lysogenic cycle^{78,123,124}. Following DNA injection to the cytoplasm, a temperate phage either initiates the lytic cycle or integrates its DNA into the bacterial chromosome of the host to become a prophage. This prophage expresses a specific phage repressor (called CI in phage λ) that inhibits transcription of most phage genes, including the lytic cycle genes, so the prophage becomes quiescent (Figure 1.2). Following integration, the prophage dsDNA is replicated concomitantly with the bacterial chromosome during cell division, so the prophage is transmitted through generations of bacteria indefinitely (vertical transfer)¹²⁵.

Exit from the lysogenic cycle occurs when the prophage excises from the host chromosome. This excision can be spontaneous, but is normally induced by stress to host cells, such as DNA damage from antibiotics, UV radiation, or thermal stress^{86,126-130}. These stress factors activate the bacterial SOS response, resulting

in the production of SOS-related proteins, including the DNA repair protein RecA. Active RecA (RecA*) initiates CI autocleavage, releasing repression of the prophage lytic genes and initiating the lytic cycle¹³¹. Induction of the lytic cycle leads to phage excision from the host chromosome, replication, and packaging, until mature virions are formed and horizontally transferred to other susceptible bacteria (Figure 1.2).

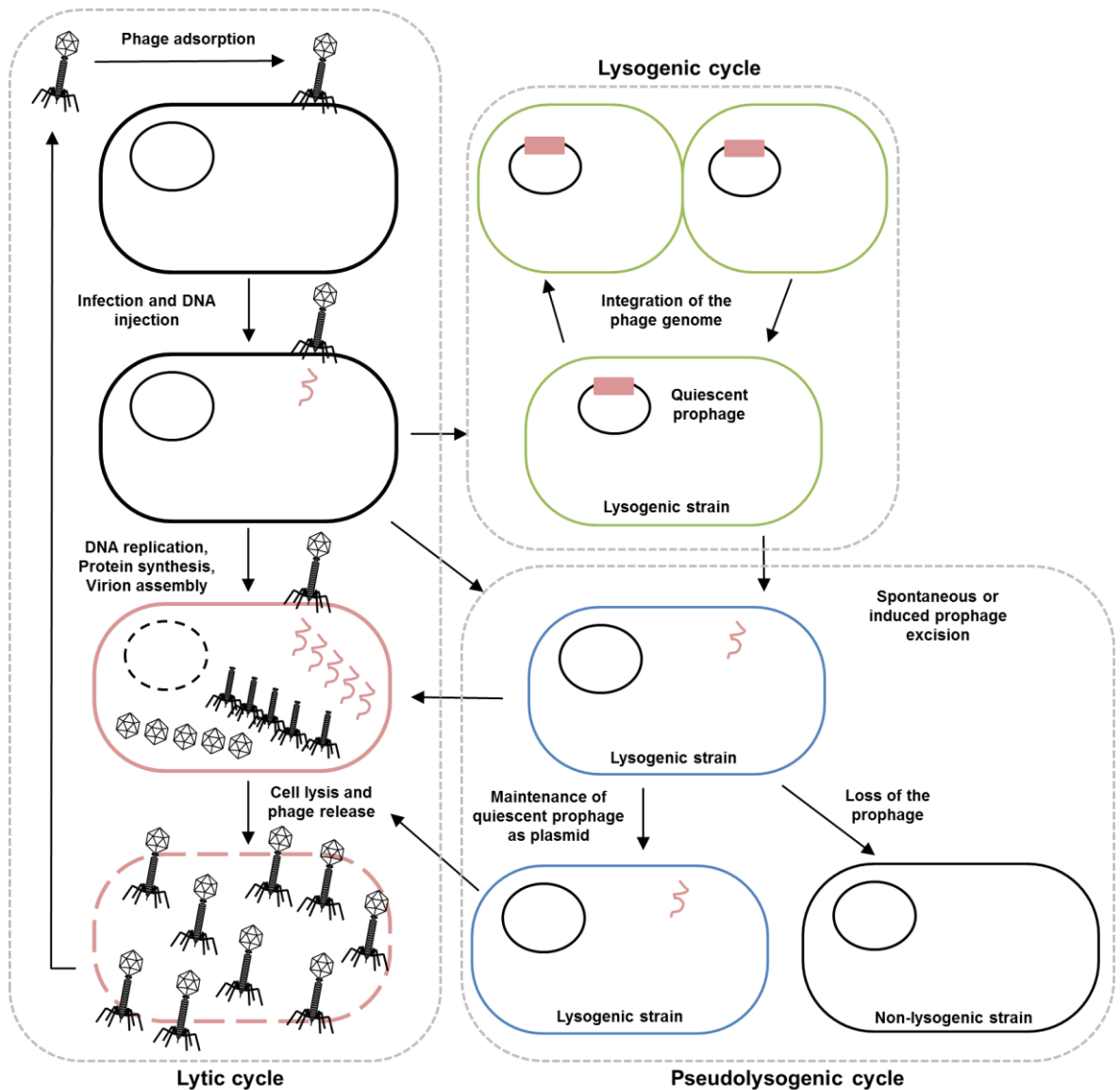


Figure 1.2 Life cycles of the bacteriophage. Virion binds to specific host receptors. Phage adsorption and injection of phage dsDNA into the bacterial cell occurs. The dsDNA can be integrated into the bacterial genome as a quiescent prophage (lysogenic cycle) or the lytic cycle occurs. Quiescent prophages remain integrated in the genome until induction of prophage excision, when the phage enters the lytic cycle. During the lytic cycle, the phage genome replicates as a concatemer while morphogenetic proteins are synthesised. The viral genome is loaded into formed phage capsids and the mature phage particle is created. The bacterial cell is lysed, releasing the new virions, which infect other bacterial cells. Adapted from Fortier *et al.* 2013¹²⁵.

Phages can also undergo a third life cycle, called pseudolysogeny¹³². Pseudolysogeny is a stage of stalled development where the phage resides in the bacterial host cytoplasm as a circular or linear plasmid-like element; neither integrated as a prophage for the lysogenic cycle, nor going through the lytic cycle. Normally triggered by unfavourable growth conditions for the host cell, pseudolysogeny ends either with true lysogeny through phage DNA integration to the host chromosome, or initiation of the lytic cycle following improved growth conditions (Figure 1.2)¹³².

1.2.5 Genetic and functional organisation of bacteriophages

Phage genomes display a modular organisation, with genes grouped according to biological function. As such, modular genetic organisation has been proposed as a parameter of phage classification. This groups phages into different families than the traditional structure-based classifications, for example using this method phage λ , HK97 (Siphovirus), P22 (Podovirus) and some *S. aureus* phages belong to the same family of lambdoid phages¹³³. These lambdoid phages display a highly similar mosaic genome structure, grouping orthologous genes into three functional modules. These modules are the early transcription or lysogeny module, the intermediate transcription or replication module, and the late transcription or packaging and lysis module (Figure 1.3)¹³⁴. This modular organisation facilitates the temporal control of gene expression in an established pattern to mediate the phage life cycle. Hence, following phage infection of a bacterial host cell the lysogeny module is the first to be expressed and determines whether the phage enters the lytic or lysogenic cycle. The intermediate module expresses after initiation of the lytic cycle, leading to expression of genes related to phage replication. Finally, the late module is expressed and encodes the genes for DNA packaging, capsid and tail morphogenesis processes and host-cell lysis, facilitating the end of the lytic cycle in which mature virions are assembled and the host-cell is lysed.

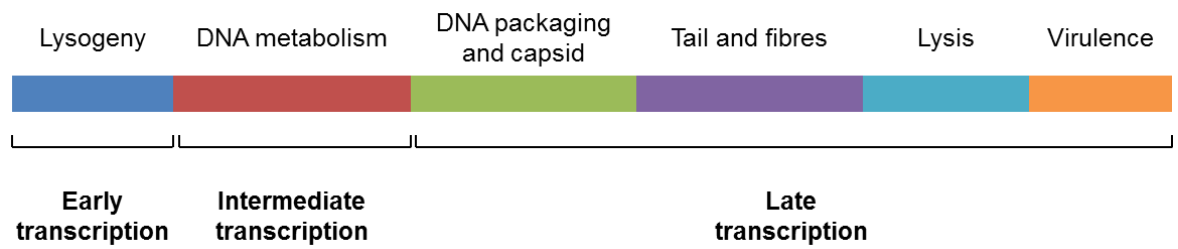


Figure 1.3 Schematic representation of *Siphoviridae* staphylococcal phage genome modules. Coloured boxes represent the different modules; blue for lysogeny, red for DNA metabolism, green for DNA packaging and capsid morphogenesis, purple for tail and fibre genes, teal for cell lysis, and orange for virulence genes.

As well as facilitating the correct gene expression order for the phage lytic cycle, the functional clustering of the phage genome allows modular interchange between different phages^{134,135}. The characteristic mosaic structure of phage genomes is caused by non-homologous recombination, which occurs at high frequency in lytic and temperate phages, and even in defective phages from the bacterial chromosome¹³⁶⁻¹³⁸. On top of this, conserved sequences at the end of each phage module facilitate the exchange of these groups through phage homologous recombination¹³⁹⁻¹⁴². These tactics allow phages to exchange genes performing the same role for the phage, but that are very different in sequence. The ability to cycle through the expression of different, unconserved phage proteins that perform the same function allows a quick and more flexible response to ecological pressures¹⁴³. Such horizontal exchange enables phages to obtain the most advantageous combination of genes from co-infecting phages, to develop an optimal genome adaptation for that particular niche^{78,87}.

1.2.6 Control of phage lysogeny

The early transcriptional module is made up of genes implicated in the establishment and maintenance of the phage lysogenic state¹⁴⁴. The temperate lambdoid phages use an ‘on/off’ genetic switch determining whether the phage enters the lytic or lysogenic cycle following infection. For most λ -like phages, the lysogenic module is made up of two transcriptional regulator genes, the CI repressor and the Cro antirepressor¹⁴⁵. The CI repressor is constitutively expressed and binds the intergenic promoter region between *ci* and *cro*, repressing *cro* and the downstream genes involved in the lytic cycle¹⁴⁶. The CI repressor, therefore, maintains the quiescent prophage of the lysogenic cycle.

On the other hand, the antirepressor Cro promotes the lytic cycle by inhibiting expression of CI synthesis following infection of a phage or induction of a prophage. Without the CI repressor, the lytic genes are expressed and the phage goes through the lytic cycle.

CI and Cro compete for the same intergenic promoter region between the *cl/cro*¹⁴⁶. Since the CI repressor prevents transcription of Cro, this binding site competition determines whether the phage undergoes the lysogenic or lytic cycle. When the bacterial host cell is under stress the bacterial SOS response is triggered. The SOS response activates the bacterial protease RecA, which triggers autolysis of the CI repressor at a specific alanine-glycine bond^{131,147}. Without CI repression, the Cro antirepressor is able to express and this in turn inhibits CI expression. This mediates the transition from a prophage to the lytic state. Furthermore, this mechanism explains host immunity against phage superinfection. When a bacterial host contains a lysogenic phage as a prophage, the host cannot be re-infected by the same phage (or by another carrying the same *cl* gene), since the CI repressor expressed by the prophage inhibits transcription of most of the infecting phage genes⁹¹.

1.2.7 Phage integration and excision

Integration of the phage genome into the host chromosome is vital for the progression of the lysogenic cycle. Following dsDNA injection, the phage dsDNA circularises using complementary 5'-overhangs (*pac* sites) or through cohesive ends (*cos* sites)^{148,149}. The phage genome then integrates into the host chromosome by the Campbell mechanism, with site-specific recombination of the phage attachment site (*att_P*) to a specific bacterial attachment site (*att_C*)¹⁵⁰. This site-specific integration is catalysed by the phage-encoded integrase (Int) within the lysogeny module¹⁵¹⁻¹⁵³. Once the phage genome is integrated, hybrid attachment sites border the prophage (*att_L* and *att_R*). These *att_L* and *att_R* sites act as recognition sites for phage genome excision. Following prophage induction, either spontaneous or SOS-mediated, the Int and neighbouring excisionase (Xis) act in a coordinated manner to catalyse phage DNA excision¹⁵⁴⁻¹⁵⁶. This means the correct ratio of Int:Xis is essential for establishing the lytic or lysogenic cycle.

Once lysogeny has been established, the proteins implicated in the lytic cycle are swiftly broken down. This protein degradation process is performed by host proteases, such as ClpX, ClpP, and others¹⁵⁷. Host-encoded proteins are also implicated in the integration of some phages, with roles played by the integration host factor (IHF) and the Fis protein^{158,159}.

1.2.8 Phage DNA replication

The replication (intermediate) module sits immediately downstream of the *cro* anti-repressor gene. Apart from genes involved in phage replication, the module contains a single origin of replication (*ori*). In λ phages, bidirectional replication is initiated from this *ori*, generating θ -like structures. The replication pattern then switches to rolling circle replication, generating the concatemeric DNA molecules (σ -like structures) that form the substrate for phage packaging^{160,161}. In the λ phages, the replication module consists of the *O* and *P* genes that are combined in a binary compound close to the *ori* site. Phage protein O is a helicase loader that initiates the replication process by binding the λ *ori* site. Protein O requires the host-encoded DNA gyrase to catalyse the negative supercoiling of the phage dsDNA¹⁶². The phage protein P binds the host-encoded replicative helicase, DnaB, which directs P to interact with the supercoiled unwound region of DNA¹⁶³. DnaB mediates widening of the single-strand region, whilst the host-encoded DnaG primase allows bi-directional replication by synthesis of RNA transcript primers¹⁶⁴. All this allows the host-encoded DNA polymerase III to initiate the bidirectional replication process. Essentially, phage proteins O and P commandeer the host polymerase. The replication forks of the single strands form replication intermediates resembling the Greek letter θ , hence θ -like structures.

Once dsDNA synthesis has been completed, the RNA primers are removed by PolA and a host-encoded DNA ligase fills in the gaps. The host-encoded topoisomerase IV resolves the circular DNA dimers^{165,166}. Finally, the replicative fork is interrupted by a poorly understood mechanism and rolling circle replication begins in structures that resemble the Greek letter σ , hence σ -like structures¹⁶⁷. This occurs in the final stages of the replication module, and is considered to be the transition point to the packaging and lysis module, containing the morphogenesis and lysis gene clusters.

1.2.9 Packaging regulation systems

The late-expressing phage gene clusters are linked to the synthesis of different proteins associated with the final assembly of the mature virions and lysis of the host cell. This includes structural proteins (head and tail), proteins facilitating the assembly of virus particles, and proteins responsible for host cell lysis. Regulation of these modules is tightly controlled and associated with phage specific regulatory mechanisms.

In phage λ , late gene expression is positively regulated by protein Q, an anti-terminator transcription factor¹⁶⁸. Q binds a DNA-site, known as *Qut*, that overlaps with the promoter of P_R' . The anti-terminator Q therefore promotes expression of the packaging and lysis module. Transcription from P_R' is constitutive, but without the Q bound to the *Qut* site, transcription stops at the t_R' terminator gene (~200 nucleotides downstream). However, when Q is present transcription extends past the t_R' to transcribe the entire morphogenesis and lysis clusters. In phage λ , this represents a transcript of 26 kb¹⁶⁸.

For the *S. aureus* phages and other phages linked to Gram-positive bacterial hosts, a diverse range of late transcriptional regulators (Ltr) has been described. These also act as positive regulators of the late transcriptional clusters, activating transcription of the morphogenetic and lysis genes. So far, five different families of phage-encoded Ltr proteins have been described: RinA, ArpU, LtrA, LtrB, and LtrC^{169,170}. These Ltr proteins are encoded just downstream of the phage replication module. Binding of the Ltr activators to a regulated promoter region upstream of the *terS* gene activates late gene transcription of the morphogenesis and lysis gene clusters.

1.2.10 Packaging initiation, DNA recognition and capsid formation

The morphogenesis and lysis gene clusters of the lambdoid phages are highly conserved and their mosaic genome structure can be found in phages with a wide-range of host bacteria, from *E. coli* to archaeobacteria^{116,171}. Phage capsid assembly is well conserved throughout these phages and can also be found in eukaryote-infecting viruses, such as herpesviruses¹⁷². Phages were traditionally used as models for understanding capsid assembly and the ATP phage molecular

motor. This is particularly true of phages λ , SPP1, HK97, and P2, which are amongst the most studied lambdoid assembly models¹⁷³⁻¹⁷⁶. Although there are some peculiarities within a few assembly processes or in capsid and tail structures, general packaging principles apply to the majority of phages (Figure 1.4)¹⁷⁷⁻¹⁷⁹.

1.2.10.1 Head assembling

One of the initial steps in the phage packaging process is the formation of an empty shell, or procapsid (Figure 1.4)¹⁸⁰. This procapsid is subsequently filled with the phage genome. A portal protein initiates shell assembly and is a central component for DNA translocation into the procapsid. The portal protein is formed by monomers that group together to produce a dodecameric-shaped cone with a central channel ring¹⁸¹. The DNA is threaded through this central ring. Assembly of the capsid shell is initiated at a single 5-fold vertex through the co-polymerization of major capsid protein monomers and, sometimes, scaffolding proteins¹⁸¹⁻¹⁸⁵.

The procapsid is made up of multiple copies of the major capsid proteins, which form an icosahedron. Most phages encode only a single capsid protein, although the T4 phage encodes two^{179,186}. The capsid proteins of the lambdoid phages all fold in a typical HK97 pattern, despite the fact these proteins sometimes share as little as 10-15% sequence identity. This standard folding includes some conserved structural features, such as an N-terminal N-arm containing α -helical content; an E-loop containing two-stranded anti-parallel β -sheets; a P-domain with a spine helix and long β -sheet; and an A-domain containing a central β -sheet¹⁸⁷. Some phage capsid proteins encode an additional N-terminal scaffold (Δ domain), which limits the conformational probabilities of the protein and supports formation of the procapsid. This Δ domain is often removed by proteolysis after procapsid assembly. Both the T5 and HK97 phages encode scaffolding domains on the capsid protein^{188,189}.

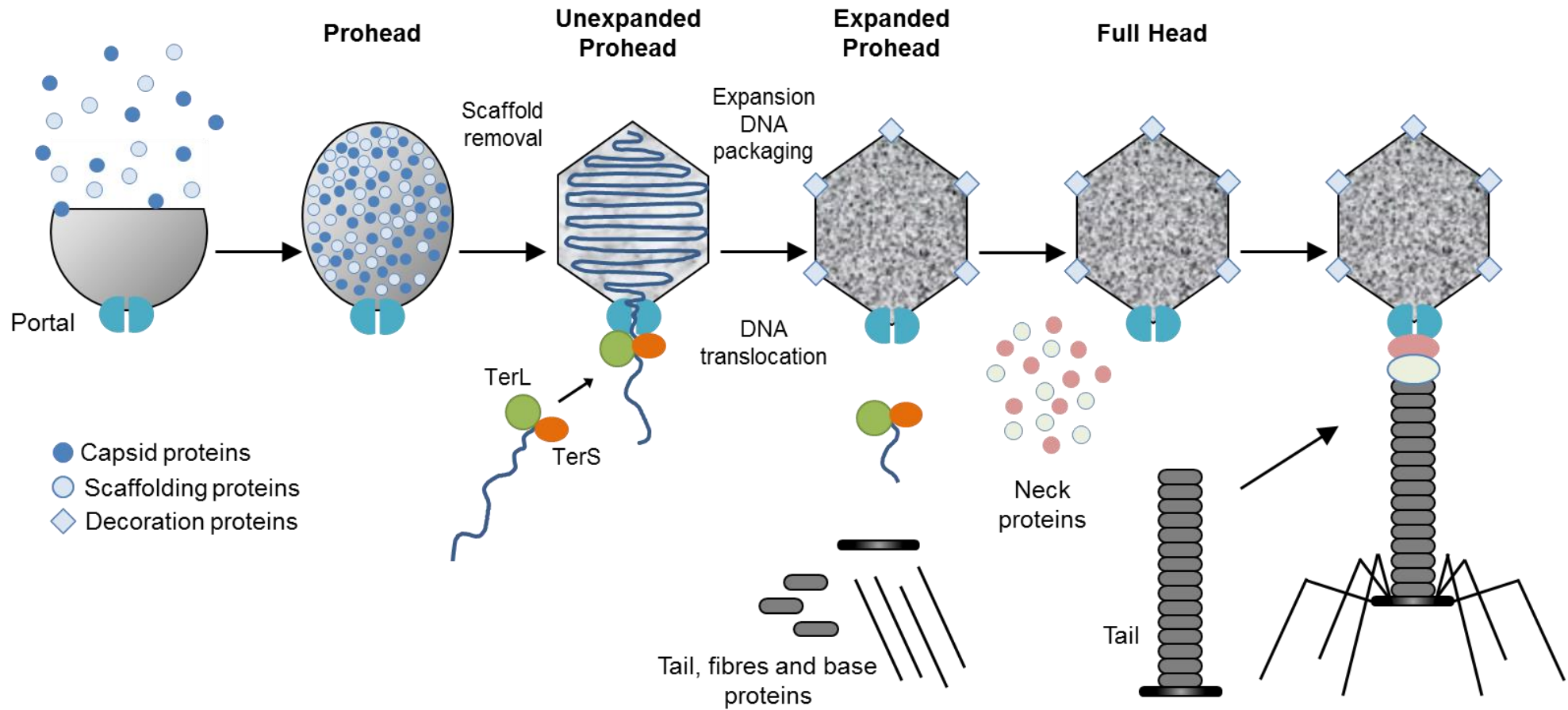


Figure 1.4 Schematic representation of virus particle assembly of dsDNA phages. 1. Assembly of the procapsid for the portal vortex. 2. Packaging and scaffolding protein removal. 3. Dissociation of terminase complex and decorative and neck protein incorporation. 4. Assembly of tail proteins. 5. Unification of capsid, tail and fibres for formation of a mature virus particle. Adapted from Rao *et al.* 2015¹⁹⁰

Other tailed phages, such as P22, encode separate scaffolding proteins, which promote capsid protein interaction and limit the procapsid conformation^{191,192}. The T4 phage encodes six individual scaffolding proteins, although this is unusual and most phages encode only one^{193,194}. Scaffolding proteins form attachment points for capsid proteins, orchestrating capsid protein monomer interaction to determine the final procapsid structure. The scaffolding proteins are usually eliminated from the capsid during maturation, so they control the formation of the capsid structure without contributing to the final phage virion¹⁹¹.

1.2.10.2 Genome recognition and processing

Whilst assembly of the capsid shell occurs, the phage genome undergoes rolling-circle replication. The rolling-circle replication mechanism generates the linear head-to-tail genome concatemers necessary for packaging of the dsDNA. DNA packaging also requires an oligomeric holoenzyme made up of two subunits, the large (TerL) and small (TerS) subunits of the terminase (Figure 1.4)^{190,195}. The TerL subunit has ATPase function that provides energy for translocation of the phage genome into the icosahedral shell¹⁹⁶⁻¹⁹⁹. It also confers a nuclease function that is responsible for cleaving individual copies of the phage genome during packaging^{200,201}. The ATPase domain is found in the N-terminal of the TerL, whilst the C-terminal contains the nuclease domain^{162,202}. The TerS subunit recognises the cleavage sites on the phage genome and modulates the ATPase and nuclease activities of the large subunit²⁰³⁻²⁰⁵. The N-terminal DNA binding domain of TerS binds the cognate recognition sequence to direct the TerL-mediated cleavage of the phage concatemers^{206,207}.

Following initial cleavage of the phage concatemer, the terminase complex translocates the phage concatemer from the portal vertex into the procapsid, by ATP hydrolysis. Once the capsid is full and the portal protein undergoes conformational changes, this triggers the TerL to cut the phage dsDNA a second time and the holoterminase complex dissociates from the capsid (Figure 1.4)^{183,208}.

Two different models have been proposed for the cleavage and packaging of phage concatemeric dsDNA¹⁷⁹. Both models propose the TerL/TerS complex as the initiator of phage genome translocation into the procapsid, following

recognition and cleavage of a specific cleavage site. However, the models differ in packaging mechanisms. The models can be described as *pac* or *cos* and this terminology is used to describe phages and cleavage sites, depending on the packaging system (Figure 1.5).

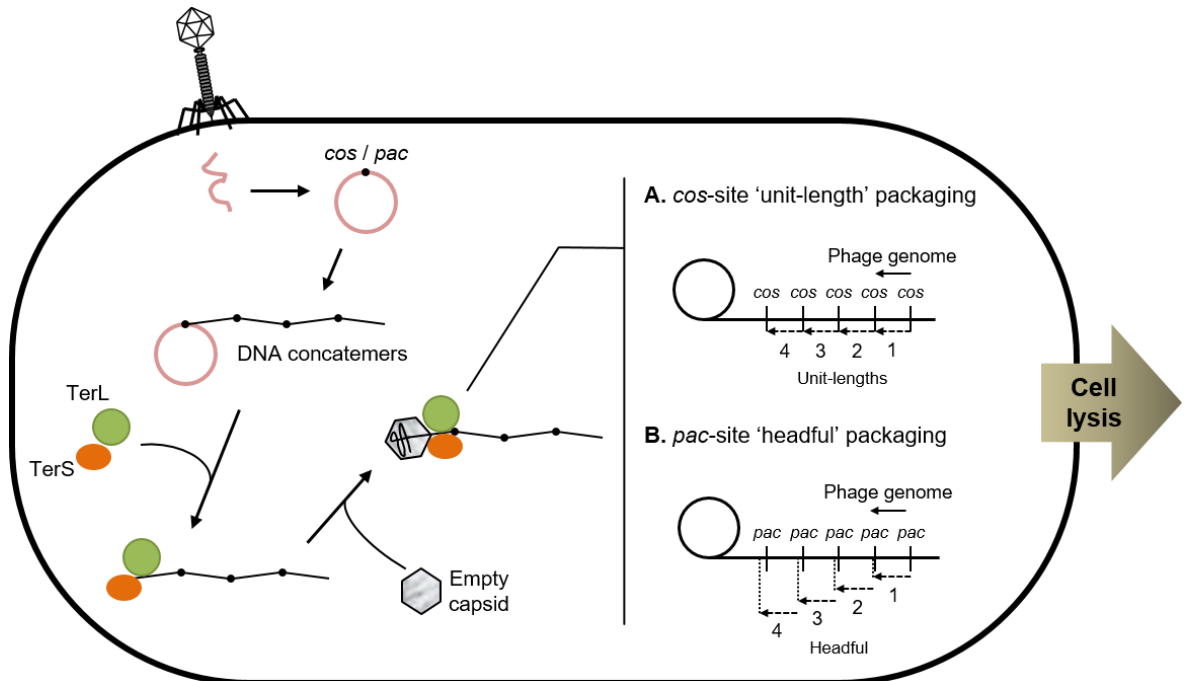


Figure 1.5 The classical *cos* and *pac* models for cleavage and packaging of phage dsDNA concatemers. Both models are identical up to the point of packaging finalisation. (A) Phages that encode *cos* cleavage sites are cut by the TerL at a second specific *cos* site, which produces a unit-length encapsidated molecule the size of the phage genome. (B) For phages with *pac*-sites the termination cleavage by TerL is sequence-independent, occurring once the procapsid is fully loaded. This 'headful' mechanism creates units slightly larger than the phage genome. Adapted from Penadés *et al.* 2015²⁰⁹.

With the *pac* phages, TerL generates the second cleavage of the phage concatemer in a non-sequence-specific manner once the capsid is full (Figure 1.5)¹⁷⁹. This packaging system, known as “headful packaging”, allows more than one phage genome unit length to be packaged (up to 110%), which provides redundant terminal sequences for circularisation. The *pac* phages include phages T4, P22, SPP1, P1, and the *S. aureus* phages $\phi 11$ and $\phi 80\alpha$ ^{173,183,204,210,211}. These *pac* phages package redundant sequences at the end of their genomes that make up 2-10% of the phage genome depending on the phage²¹². These redundant sequences allow the phage genome to circularise via homologous recombination following phage infection of a recipient bacterium. Only a single *pac* cleavage site is required for *pac* phage packaging, for the initial concatemer cleavage that

triggers dsDNA translocation. Because of this limited sequence requirement and the increased packaging capacity, the *pac* phages are regarded as the conductors of foreign DNA through generalised transduction²¹³.

On the other hand, *cos* phages require two *cos* cleavage site sequences in the dsDNA concatemers^{179,214,215}. On top of the initial *cos* site that must be cleaved to initiate phage dsDNA translocation, *cos* packaging necessitates a second identical *cos* sequence for the terminating cleavage event (Figure 1.5). This site-specific cleavage produces cohesive ends on the phage genome, allowing circularisation once the phage has infected a bacterial host cell. These *cos* cleavage sites produce unit-length phage dsDNA for encapsidation. The *cos* phages include phages λ , P2/T7, and the *S. aureus* phages ϕ 12 and ϕ SLT^{199,214,216}. The *cos* system of phage λ encodes two cleavage sites in the phage genome, *cosB* and *cosQ*, which determine the initiation and termination of phage dsDNA packaging. Cleavage at these sites produces a phage genome with cohesive ends of 12 bp, called the *cosN* site, allowing subsequent circularisation on infection.

The restrictive characteristics of *cos* packaging led to *cos* phages being disregarded as potential sources of generalised transduction, due to the low likelihood of two pseudo *cos*-sites being found one-unit length apart in other regions of the bacterial genome. Indeed, given the apparent unlikelihood of functionally spaced pseudo *cos*-sites, the *cos* phages have been the main target of phage therapy research on the basis they would not confer advantageous DNA to the bacterial hosts. However, recent studies have identified the ability of *cos* phages to horizontally transfer MGEs such as the *S. aureus* pathogenicity islands, calling into question the assumption that only *pac* phages are agents of other types of HGT²¹⁷.

1.2.11 DNA translocation machinery and packaging termination

During phage genome replication and incorporation into the capsid shell, the capsid goes through a series of stages of maturation¹⁸⁷. One of these stages is the elimination of the scaffold proteins. There are several mechanisms used by different phages for scaffold protein elimination. With phages P22 and phi29, the scaffold proteins exit the procapsid via several holes at the capsid centre^{192,218}. The scaffold proteins of phages HK97 and T4 are proteolytically

cleaved by self-encoded proteases²¹⁹. Packaging of DNA causes the exit or breakdown of the scaffolding proteins, which triggers capsid maturation²²⁰. This stage of capsid maturation leads to capsid expansion to accommodate the entire phage genome (Figure 1.4)²²¹.

Phages facilitate capsid maturation using a number of different methods. In some cases, the formation of covalent bonds between coat subunits enables capsid expansion²²². Alternatively, some tailed phages achieve shell stabilisation via the binding of self-encoded decoration proteins that attach to the outer capsid surface (Figure 1.4)²²³⁻²²⁶. One phage that utilises decoration proteins is phage λ , which encodes protein gpD for packaging of the phage genome²²⁷. Likewise, phage T4 encodes the Soc protein that is required for virion stability during exposure to environments with extreme pH or temperature²²⁸. Another method to enable capsid maturation is the incorporation of minor capsid proteins, albeit at low copy numbers. Although they are not essential to capsid formation, they are crucial for virion infectivity, with infection efficiency lost in their absence. For example, phage P22 encodes three minor capsid proteins that modulate the ejection of dsDNA during infection²²⁹.

Once packaging of the phage dsDNA is complete, the terminase complex dissociates from the portal protein to be replaced by gatekeeper proteins, which form the neck of the phage particle (Figure 1.4)²³⁰. At this point the portal protein complexes with connector proteins for a connector channel to prevent premature exit of the phage dsDNA¹⁷³. The tail structure must then associate to complete the phage particle. The tail structure should be constructed to allow the release of DNA following the binding of the phage fibres to bacterial host receptors²³¹.

The traditional classification of the phages is based on tail structure, with the short-tailed *Podoviridae* and the long-tailed *Siphoviridae* and *Myoviridae*¹¹³. In the case of phage particle assembly, the short-tailed *Podoviridae* differ to the other two families. The *Podoviridae* phages attach the tail proteins sequentially to the newly formed capsid^{232,233}. The *Siphoviridae* and *Myoviridae* families of phages assemble the phage tail using a separate pathway to capsid formation²³⁴.

Siphoviridae and *Myoviridae* tail assembly begins with the creation of an initiator complex, known as the motherboard or baseplate²³⁵. The baseplate constitutes the phage absorption tube from the distal end. Tail protein(s) go through a series of conformational changes during assembly and polymerisation at the baseplate, forming the cylindrical tail region (Figure 1.4). Both families of phages utilise a tape-measure protein that acts as a scaffolding protein during tail polymerisation and determines the tail length²³⁶. Most phage tails are composed of multiple copies of the major tail protein, although there are exceptions. For example, phage SSP1 encodes two major tail proteins, present at a ratio of 3:1²³⁷. The *Myoviridae* phages cover the tail structure with a contractile sheath protein, which contracts to force the tail through the bacterial membrane when the phage binds to the surface of a bacterial host^{194,238}.

Following tail assembly, completion proteins interact with the tail proteins facilitating the binding of the tail and baseplate to the mature capsid (Figure 1.4)^{239,240}. Once the capsid and tail structures are complexed, the long tail fibres are attached to complete phage particle assembly²⁴¹. These fibres recognise specific bacterial host receptors^{242,243}, but can remain in a retracted position to prevent infection if conditions are unfavourable to phage growth²⁴⁴.

Once the phage virion particles are formed, the phage progeny exit the host to infect new recipient bacterial cells. Phages have several strategies to enable release of virions from the host cell. Most phages encode the holin and endolysin proteins. The holin protein accumulates in the cytoplasmic membrane, creating small nonspecific holes¹²¹ through which the endolysin can access the peptidoglycan in the bacterial wall²⁴⁵. The endolysin enzyme degrades the bacterial wall peptidoglycan, leading to cell lysis and phage exodus¹²⁰.

1.3 Phage Inducible Chromosomal Islands

1.3.1 Overview of the genomic island family

The genomic island (GI) family is made up of discrete sections of DNA between ~10 to 200 kb in length²⁴⁶. These GIs are found in the bacterial genome, but are significantly different from the non-mobile, core regions. Differences in the

nucleotide composition, GC content, and codon usage suggest that GIs were acquired into the bacterial chromosome via HGT^{101,246,247}. Generally, GIs are flanked by 16-20 bp repeats, similar to those seen in phages and also named *att_L* and *att_R* sites. These sites act as recognition sites for specific enzyme-mediated cleavage for excision from the host chromosome²⁴⁸.

GIs often include mobility-linked genes, like integrases, plasmid conjugation factors, and phage-related genes. The inclusion of mobility genes implies GI mobility, or the parasitism of other MGEs. Other genes encoded by the GIs provide a selective advantage to the host bacterium, promoting maintenance of the GI within the host chromosome. Depending on what advantage these GI-encoded genes confer, the GIs can be labelled as pathogenic, symbiotic, metabolic, defence, or resistance islands²⁴⁸. The broad definition of GIs means a wide-range of functional elements are included in this family, often with very diverse lifestyles.

The first GI to be identified was the *E. coli* pathogenicity island (PAI), identified in the genome of *E. coli* uropathogenic strains^{249,250}. The PAI elements were categorised as unstable chromosomal regions encoding different virulence characteristics depending on the strain²⁵¹. The PAI elements fall within the definition of GIs, but form only one part of this diverse and broadly distributed family²⁵²⁻²⁵⁴. Different GIs are involved in wide-ranging bacterial processes, including virulence, symbiosis processes, sucrose and aromatic compound metabolism, mercury resistance, and cytochrome *c* biosynthesis²⁵⁵⁻²⁵⁷. Clinical isolates of *S. aureus* have been found to carry significant numbers of GIs. These *S. aureus* GI elements are varied in organisation and gene content and include the *SCCmec*, pathogenicity islands that encode toxic shock syndrome toxin-1 (TSST-1), and the *vSa* genomic islands (*vSaα*, *vSaB*, *vSaγ*)^{47,246,258,259}. Indeed, all the sequenced clinical isolates of *S. aureus* carry more than one of these MGEs²⁶⁰⁻²⁶².

One of the best characterised types of GI are the *S. aureus* pathogenicity islands (SaPIs)²⁶². The SaPIs are phage satellites that are innately linked to specific temperate phages, known as helper phages. The SaPIs parasitize helper phages for their induction and transfer to a new bacterial host cell. All the different SaPIs share a common life cycle, similar to that of plasmid P4²⁶³ and *Sulfolobus*

plasmid pSSVx²⁶⁴ that both exploit their host to achieve replication. The first SaPI mobilisation studies identified that these GIs are induced by certain helper phages, leading to SaPI excision from the chromosome and circularisation of the SaPI DNA. Phage induction of the SaPIs can occur following one of two events; either a resident prophage is induced when environmental stresses trigger the bacterial SOS-response, or a susceptible bacterial cell carrying a SaPI is infected by a helper phage^{126,127,265,266}. The phage-induced SaPIs excise from the host chromosome, undergo replication, and are packaged into particles made up of phage proteins. Following phage-mediated cell lysis, these SaPI particles are intra- and inter-generically transferred to other cells at a high frequency^{217,267}. This series of events has been termed the excision-replication-packaging (ERP) SaPI cycle²⁶⁸.

SaPI gene organisation displays a similar structure to that of the phages and the SaPIs encode certain phage-homologue proteins, including homologues of the integrase, excisionase, primase, replication initiator, and small terminase subunit²⁶². Like the phages, the SaPI genes exhibit a modular arrangement and group by function, with a regulatory module, a replication module, and a packaging module. With prophage-like parallels, SaPI chromosomes are flanked by *att_L* and *att_R* sites made up of 16-20 bp direct repeats²⁶⁹. So far, 5 different SaPI-specific integration sites (*att_C*) have been identified in *S. aureus*, separate to the phage integration sites^{268,270}. So, although the SaPIs are inherently tied to the phage life cycles, they integrate separately into the bacterial genome.

One or more SaPIs have been identified in all the sequenced clinically relevant strains of *S. aureus*. Indeed, most SaPIs encode the cause of toxic shock, TSST-1, together with two or more superantigens^{28,29}. On top of this, some SaPIs encode additional genes involved in biofilm formation (SaPIbov2)²⁷¹, antibiotic resistance (SaPIfusB)^{272,273}, or host adaptation (SaPIbov4, SaPIbov5, SaPIeq1, and SaPIov2)³⁴. As carriers of numerous bacterial virulence and fitness genes, the SaPIs play a significant role in *S. aureus* niche adaptation and pathogenicity.

On a wider scale, SaPI-like elements have now been identified in a number of other Gram-positive bacteria, such as *Streptococcus pyogenes* (SpyCIM1), *Enterococcus faecalis* V583 (EfCIV683), and *Lactococcus lactis* (LICIBIL312)^{262,274-276}. This collective family of highly effective MGEs has been designated the

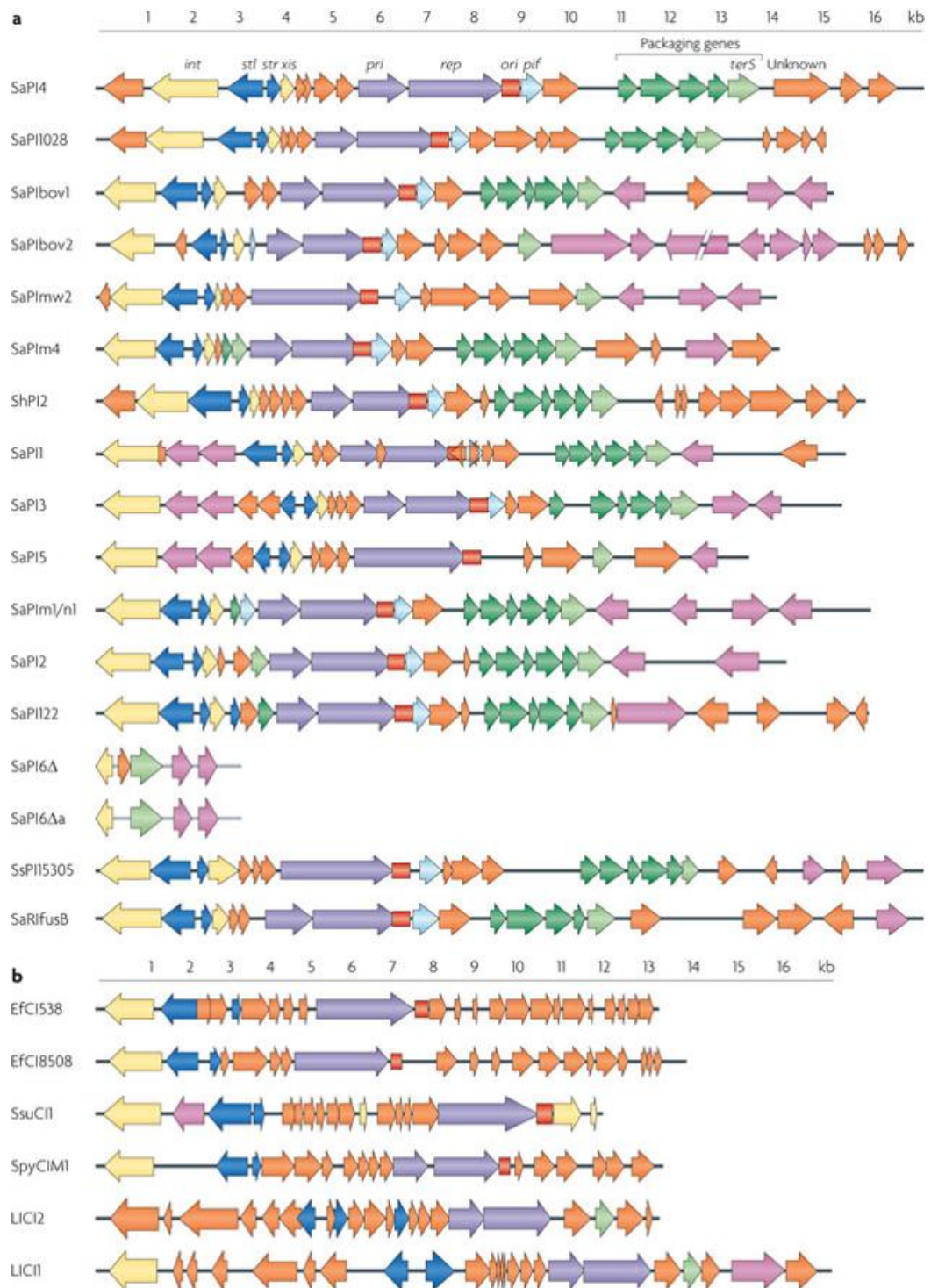
phage inducible chromosomal islands (PICIs)²⁷⁷. PICIs interfere with and utilise specific helper phages for their own high-frequency HGT. They are smaller than the phages that they parasitize, with genomes of approximately one third the size of a phage genome (~15 kb). This novel and ever-expanding family of MGEs plays a major role in bacterial evolution and the development of pathogenicity.

1.3.2 Genome organisation and regulation of the SaPIs

The SaPIs are the best characterised of the PICI family. They exhibit a highly conserved genetic structure that displays parallels to the phages modular gene organisation. The SaPI genome is grouped into modules of genes with related functions, such as transcriptional, integration-excision, replication, morphogenetic, and virulence modules (Figure 1.6)^{262,278}. Some of these functions can be carried out by different proteins, but the functions themselves are conserved between the different SaPIs. The SaPIs are induced by specific phage proteins, but these are often carried on multiple phages. As such, the SaPIs can be induced by several helper phages, so long as the appropriate phage protein is encoded. Following SaPI induction, the ERP cycle is activated. During this cycle some SaPIs encode proteins that enable them to interfere with the helper phage life cycle, leading to increased SaPI packaging and transfer²⁷⁹.

1.3.2.1 SaPI transcriptional control

The regulatory transcriptional module of the SaPIs is made up of the *stl* and *str* genes, which are encoded in a divergent orientation (Figure 1.6). These two genes work together to regulate transcriptional activation or repression of the SaPI genes. Both the Stl and Str proteins encode a helix-turn-helix motif and are poorly conserved between the different SaPIs (Figure 1.7)²⁸⁰. The Stl is the master repressor of the SaPIs, binding the promoter regions between the *stl* and *str* genes to prevent transcription of *str* and the downstream genes²⁷⁸. Stl repression blocks excision and replication of the SaPIs, maintaining the integration of the elements in a quiescent state within the bacterial chromosome.



Nature Reviews | Microbiology

Figure 1.6 Comparison of phage inducible chromosomal island genomes. Genomes are aligned according to the prophage convention, with the integrase gene (*int*) at the left end. Genes are coloured according to their sequence and function: *int* and *xis* (excisionase) are yellow; transcription regulators are dark blue; replication genes (including the primase gene (*pri*) and the replication initiator gene (*rep*) are purple; the replication origin (*ori*) is red; encapsidation genes are green, with the terminase small subunit gene (*terS*) in light green; superantigen and other accessory genes are pink; and *pip* (which functions in phage interference) is light blue. Genes encoding hypothetical proteins are orange. (A) The gene organization of *S. aureus* pathogenicity islands (SaPIs). (B) Putative phage-related chromosomal islands from genera other than *Staphylococcus*; from top, *Enterococcus faecalis*, *E. faecalis*, *Streptococcus suis*, *Streptococcus pyogenes*, *Lactococcus lactis* and *L. lactis*, respectively. SaPIm1/n1 indicates SaPI_n1 (from *S. aureus* str. n315) and SaPI_m1 (from *Staphylococcus aureus* str. mu50), which are essentially

identical. Reproduced from Novick *et al.* 2010, with permission from the Nature Publishing Group²⁶².

The SaPI Stl is a functional homologue of the CI repressor from phage λ ¹⁴⁶. However, unlike the CI, Stl is not degraded on activation of the SOS response. In fact, the SOS-response does not have a direct effect on SaPI activation. Instead, Stl deactivation requires a protein:protein interaction with specific phage-encoded proteins that are produced during the phage cycle (Figure 1.7A)²⁸⁰. These phage-encoded SaPI inducers bind to Stl, removing the repressor from the intergenic promoter to eliminate Stl-mediated island repression. In turn, this activates transcription of the *stl* and *str* promoters to activate the SaPI cycle. Studies in which the Stl repressor was deleted indicate that without the master repressor the SaPI excises and replicates autonomously²⁷⁸. However, these permanently activated SaPIs confer a high cost to the bacterial host, explaining why the characterised SaPIs encode a functional Stl repressor²⁸¹.

Interestingly, the different SaPIs encode different Stl repressors that are highly divergent in sequence (Figure 1.7B). This has resulted in the different SaPI Stl repressors requiring different phage-encoded SaPI inducer proteins for the protein:protein interaction²⁸⁰. As such, depending on the Stl encoded by the SaPI, different phage-encoded SaPI inducers are required for island activation to occur. Because the different phages encode various proteins that perform the same function, only those phages encoding the appropriate SaPI inducer for that specific Stl will be able to activate the encoding island.

The second gene within the transcriptional module is the *str* (Figure 1.7A). The function of Str is currently unknown, although unpublished data suggests it could mediate rightward transcription of the downstream SaPI genes. If this is the case, the Str could act similarly to the phage λ Cro, competing with Stl for binding the intergenic promoters. However, *str* deletion had no impact on the ERP cycle of SaPI_{bov1}²⁷⁸ or SaPI1²⁸². As such, further study is needed to elucidate the role of this gene within the SaPI cycle.

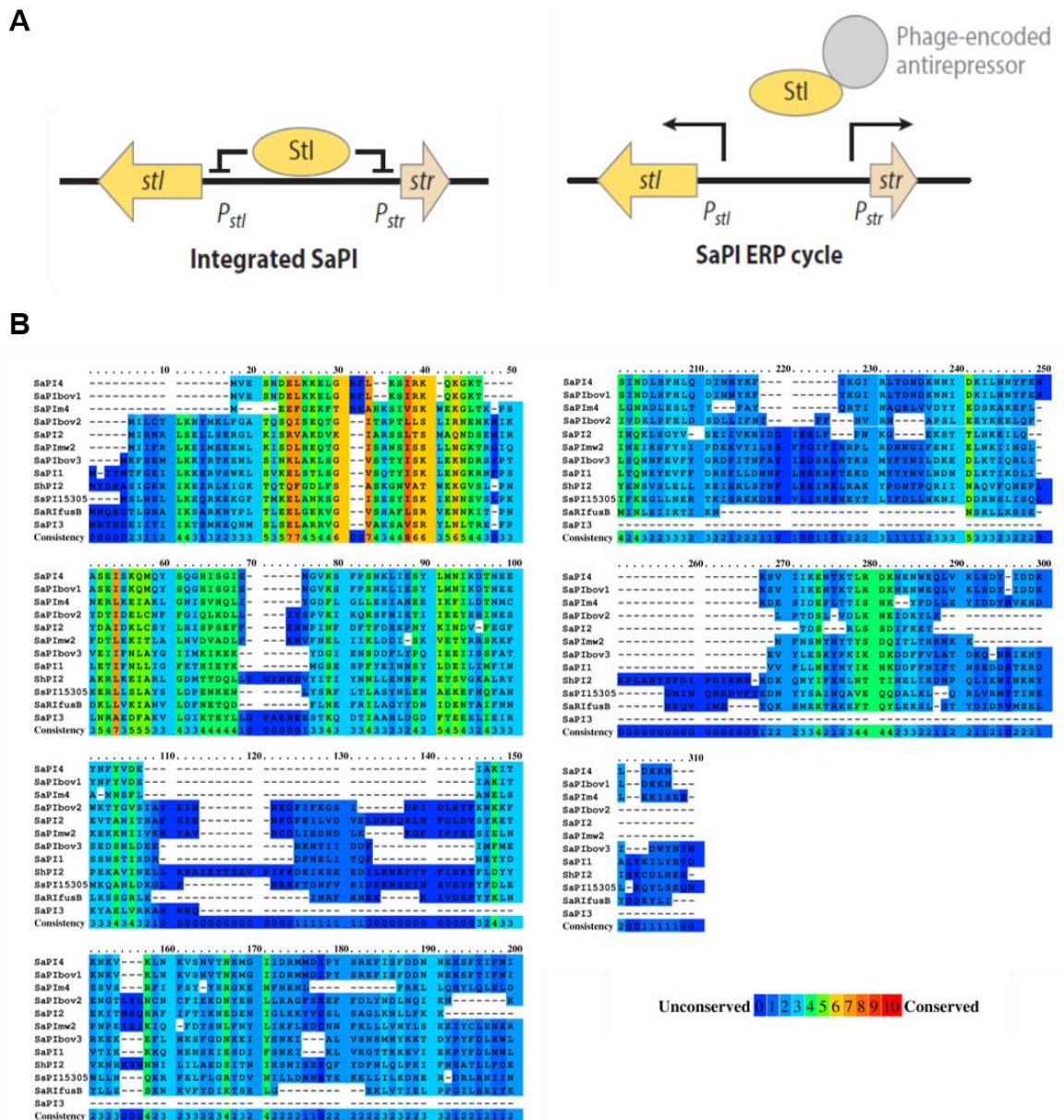


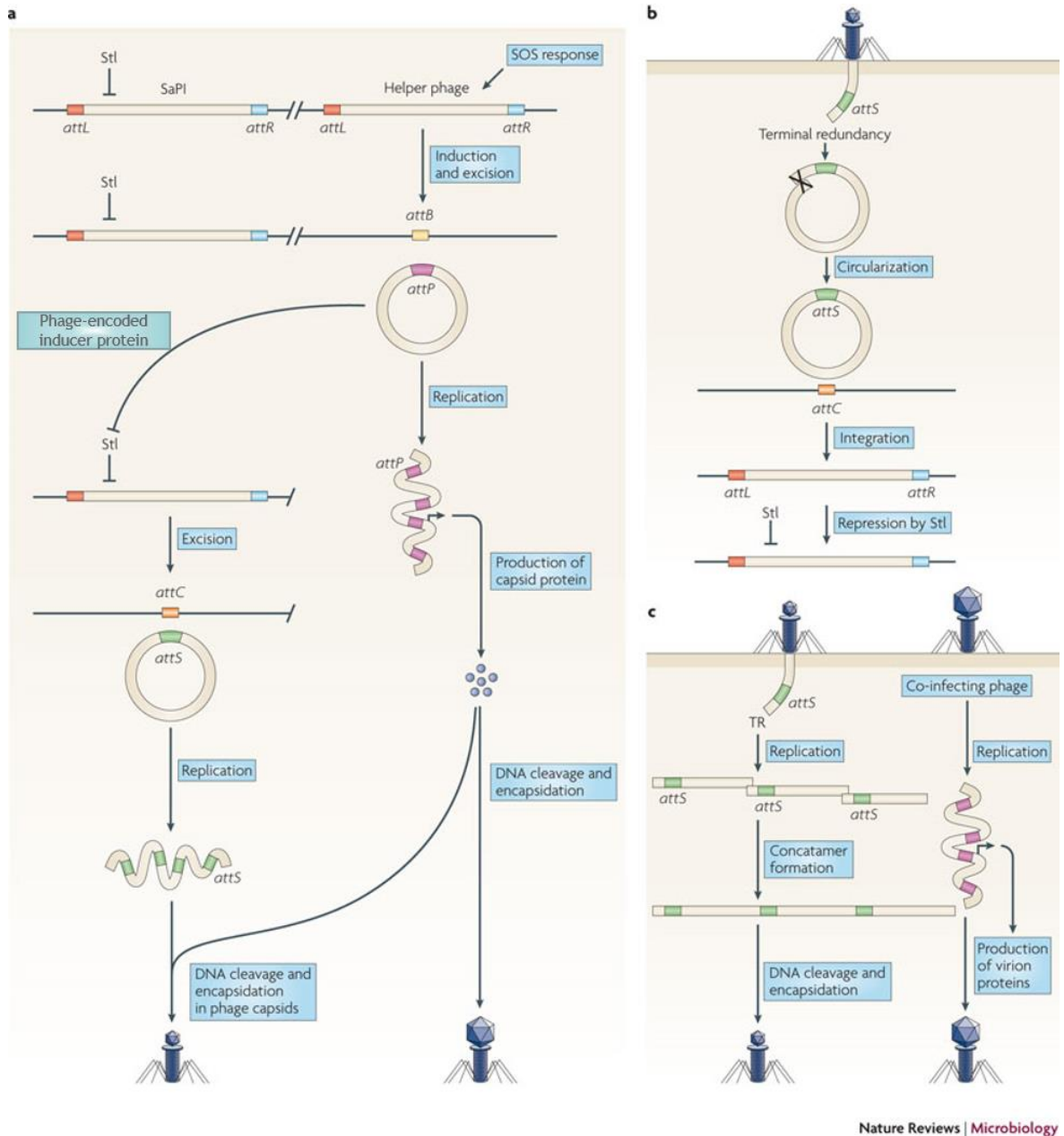
Figure 1.7 SaPI transcriptional module and transcriptional control (A) SaPI induction by a helper phage. Integrated SaPI: stable integration is maintained by the SaPI master repressor Stl, which silences expression of the SaPI genes. SaPI excision-replication-packaging (ERP) cycle induction: interaction with a phage-encoded anti-repressor leads to relief of Stl repression of both the *stl* and *str* promoters (P_{stl} and P_{str}), inducing the SaPI ERP cycle. Adapted from Penadés and Christie 2015²⁸³. (B) A protein sequence alignment of twelve different Stl repressors from various staphylococcal pathogenicity islands (SaPI4, SaPIbov1, SaPIbov2, SaPIImw2, SaPIbov3, SaPIIm4, ShPI2, SaPI1, SaPI3, SaPI2, SsPI5305, SaRifusB). Colours indicate relative sequence conservation at each position, with red being most conserved and blue being least (adapted from alignment generated by PRALINE)^{284,285}.

1.3.2.2 SaPI integration and excision

The integration-excision module is located on either side of the transcriptional modules. This module is made up of an integrase (*int*)^{265,268,270,286} and a gene coding for a protein with excisionase activity (*xis*, Figure 1.6)²⁶⁹. The SaPI-

encoded tyrosine integrase catalyses the integration of circularised SaPI intermediates into the bacterial chromosome. This integration occurs through recombination between specific attachment sites in the bacterial chromosome (*att_C*) and the SaPI (*att_S*), which is similar to the phage *att_P* site. This mechanism creates hybrid flanking regions next to the integrated SaPI, *att_L* and *att_R*. The SaPI sits stably in the bacterial chromosome until helper phage-mediated induction. Once the SaPI has been induced, the *stl* and *str* promoters facilitate expression of the integration-excision module. Coordinated expression of *int* and *xis* allows for the correct cleavage of the *att_L* and *att_R* sites and circularised SaPI intermediates are formed (Figure 1.8)^{268,278,287}.

SaPI integration only ever occurs with one SaPI at each *att_C* site within an individual bacterial chromosome²⁶⁹. However, SaPI integrases have a lower specificity than phage integrases. Indeed, SaPI1 was shown to undergo transduction *in vitro* at an unaffected rate following *att_C* deletion from the bacterial chromosome²⁸⁸. Furthermore, spontaneous excision was exhibited in the absence of a helper phage with both SaPIbov1 and SaPIbov2²⁶⁵. Indeed, spontaneous excision is linked to high rates of transfer by non-inducing phages, although this also correlates with high levels of SaPI loss. It is thought that the potentially lower affinity of Stl-binding of the *stl-str* intergenic region in these islands could allow for spontaneous excision²⁶⁹. These tactics could be useful for exploiting phages that are otherwise non-inducing for the SaPIs or for allowing integration into a new species of genera of bacterial host.



Nature Reviews | Microbiology

Figure 1.8 Staphylococcal pathogenicity island replication scenarios. (A) *SaPI* induction by an SOS-induced helper prophage. *att_P* and *att_S* are the prophage and *SaPI* core attachment sequences, respectively; *att_B* and *att_C* are the prophage and *SaPI* core chromosomal attachment sequences, respectively. After induction and excision of the helper phage, phage-encoded *SaPI* inducer proteins relieve the *Stl*-mediated repression of the *SaPI*, allowing production of *SaPI* proteins. *SaPI* integrase (*Int*) and excisionase (*Xis*) subsequently promote the excision of the *SaPI* through a Campbell mechanism, restoring the *att_C* and *att_S* sites. Subsequent *SaPI* replication produces hundreds of copies in the form of a concatemer, which is cleaved by the terminase complex into individual copies and packed into phage particles consisting entirely of phage proteins but with a smaller head than the phage capsids, owing to the action of *SaPI* proteins. (B) *SaPI* infection. Terminal redundancy allows the incoming *SaPI* DNA to circularize, after which the circular DNA is integrated at the chromosomal *att_C* site by crossover with the *SaPI* *att_S* site. This process requires the *SaPI* integrase. *Stl* silences the expression of *SaPI* genes, keeping the element integrated. (C) Co-infection of a *SaPI* and a helper phage allows the *SaPI* DNA to be replicated prior to integration. The synthesis of phage proteins leads to the production of phage capsids, which are used to package the replicated *SaPI* DNA. As in part a, *SaPI* proteins affect the size of the capsids that are produced. Reproduced with adaptations from Novick *et al.* 2010, with permission from Nature Publishing Group²⁶².

1.3.2.3 SaPI replication

The replication module includes a cognate origin of replication (*ori*) with primase (*pri*) and replicase (*rep*) genes (Figure 1.6)²⁸⁹. The *ori* is situated at the 3'-end of the *rep* gene and has a SaPI-unique organisation, different to the phage *ori*. The SaPI *ori* consists of a series of repeated divergent iterons flanked by AT-rich regions^{289,290}. The expressed product of the *pri* genes does not seem to be essential for replication, although it does improve this process²⁸⁹. The *rep* encoded protein has a helicase function. Rep binds the AT-rich region of the *ori* to initiate replication of the circularised SaPI by melting the regions double strand²⁸⁹.

The structure and function of the *pri-rep-ori* pattern of the module is highly conserved, both throughout the different SaPIs and within the wider PICI family (Figure 1.6). Occasionally, the *pri* and *rep* genes are found fused together, but they are both present within these fusions in the same order. It is assumed that SaPIs replicate by rolling-circle replication, forming a linear genome concatemer that is then processed for encapsidation (Figure 1.8).

1.3.2.4 SaPI packaging and phage exploitation

The SaPIs do not encode any structural proteins for virion formation, instead utilising the structural proteins of the helper phages to form capsids for the packaging of SaPI genomes^{291,292}. However, the SaPIs encode proteins to mediate this process, which are organised into two distinct gene clusters (Figure 1.6). The first cluster encodes the Ppi (phage packaging interference) protein. Ppi blocks the interaction of the phage-encoded TerS subunit with the phage dsDNA, but does not affect the SaPI-encoded TerS. This ensures advantageous packaging of the SaPI in place of the phages, leading to high frequencies of SaPI transfer and, ultimately, interfering with the helper phage life cycle²⁷⁹.

Most SaPIs also contain an 'operon I' comprising of morphogenetic and interference genes that are related to packaging²⁹³. Operon I is controlled by a LexA-dependent promoter that drives the expression of this gene cluster²⁹³. Within operon I, the SaPIs encode a small terminase subunit (TerS_{SP}). This protein complexes with the phage-encoded TerL subunit to recognise and cleave the specific *pac* sequence encoded by the SaPI²⁸⁶.

Operon I also encodes the CpmA and CpmB proteins, which redirect the assembly of the helper phage procapsids to produce smaller capsids that are only viable for SaPI genomes (Figure 1.8)^{292,294}. Because SaPI genomes are smaller than those of the helper phages any phages packaged in these small capsids would be lacking essential genes and non-functional^{286,294,295}. Again, this facilitates greater SaPI transfer, whilst impacting on the phage life cycle. Furthermore, operon I contains the genes *ptiA*, *ptiB*, and *ptiM*, the expression of which blocks the helper phage late gene module to interfere with phage reproduction²⁹⁶.

Remarkably, the LexA-mediated operon I is induced by the SOS-response, so could be of benefit for SaPI-related gene transfer following infection by or induction of a SaPI-non-inducing phage²⁹³. However, during normal SaPI ERP induction this SOS-mediated activation of operon I is unnecessary as the operon seems to be transcribed from the *str* promoter.

1.3.2.5 SaPI accessory genes

The SaPIs encode a number of bacterial virulence and fitness associated genes, including nonessential and accessory genes²⁹. They are usually located at the extreme ends of the SaPI genome, either to the left of the *stl* and *int* region or downstream of the operon I (Figure 1.6). The SaPI accessory genes, or ‘morons’, show similarities to the morons encoded by temperate phages and they have their own promoter regions for expression.

SaPIs encode a number of different virulence factors as accessory genes, including the TSST-1 protein and the superantigen enterotoxins B, C, K, L, and Q (Table 1.2)^{27,29,297-299}. Other virulence factors associated with bacterial pathogenicity are found encoded in SaPIs. For example, biofilm associated protein Bap²⁷¹, iron transporter FhuD³⁰⁰, exfoliative toxin ETA³⁰¹, the staphylococcal complement inhibitor^{52,302}, and the von Willebrand factor-binding protein³⁴ are all found in certain SaPIs. Variations of the von Willebrand factor-binding protein are encoded by the SaPIs, facilitating *S. aureus* coagulation of bovine, ovine, caprine, and equine plasma³⁴. The fusidic acid antibiotic resistance gene *fusB* is also a SaPI-encoded virulence factor, found recently in impetigo-causing clones^{272,273}. A SaPI-like element from *Staphylococcus*

saprophyticus (vSs15305) was even found to encode antibiotic resistance determinants for fosfomycin and streptomycin³⁰³.

Table 1.2 Virulence and accessory genes encoded by the SaPIs and SaPI-like elements.

Protein	Gene	Function	Encoding SaPIs
Toxic shock syndrome toxin 1 (TSST1)	<i>tst</i>	Superantigen	SaPI1, SaPI2, SaPIbov1, SaPIin1, and SaPIIm1
Enterotoxin B	<i>seb</i>	Superantigen	SaPI3
Enterotoxin C	<i>sec</i>	Superantigen	SaPIbov1, SaPIin1, SaPIIm1, SaPImw2, and SePI1
Penicillin-binding protein fragment	<i>ear</i>	Penicillin resistance (<i>E. coli</i>)	SaPI1, SaPI3, and SaPI5
Enterotoxin K	<i>sek</i>	Superantigen	SaPI1 and SaPI3
Enterotoxin L	<i>sel</i>	Superantigen	SaPIbov1, SaPI3, and SePI1
Enterotoxin Q	<i>seq</i>	Superantigen	SaPI1 and SaPI5
Exfoliatin A	<i>eta</i>	Epidermolytic toxin	SaPI2
Biofilm-associated protein (BAP)	<i>bap</i>	Surface expressed protein, adhesion linked	SaPIbov2
Ferrichrome ABC transporter homologue	<i>fhud</i>	Iron transport	SaPIIm4
Multidrug resistance protein	<i>mdr</i>	Multidrug exporter	SaPI122
Aminoglycoside adenyl transferase	<i>aad</i>	Aminoglycoside resistance	SsPI15305
Glutathione thionyl phosphatase	<i>fosb</i>	Fosfomycin resistance	SsPI15305
FusB	<i>fusb</i>	Fusidic acid resistance	SaRIfusB
Ribosome methylase	<i>erma</i>	MLS resistance	SpPI1
Unnamed	<i>unnamed</i>	Phage resistance	LICI1
GNAT family homologue	<i>acta</i>	Acetyl transfer protein	SaPI6Δ

ABC, ATP-binding cassette; MLS, macrolide-lincosamide-streptogramin B

1.3.3 SaPI-helper phage interactions

SaPIs are innately linked to helper phages for their propagation and transfer. To achieve this phage-mediated transfer they exploit the phage life cycle at a number of stages, using several mechanisms to take advantage of the phage for

their own needs. The complementary strategies of the SaPIs interfere with helper phages at different points in phage development, and not every SaPI encodes every interference mechanism²⁷⁹. However, on top of promoting highly efficient SaPI transfer, SaPI-associated manipulation of helper phages also impairs phage reproducibility. Therefore, the bacterial host and neighbours benefit from such SaPI exploitation, due to reduced phage superinfection.

1.3.3.1 SaPI induction and derepression

SaPI utilisation of helper phages begins with SaPI induction itself. As mentioned, SaPIs reside stably in the bacterial chromosome under control of the master repressor, Stl. Whilst Stl remains bound to the divergent *stl* and *str* promoters in the intergenic regions between these two genes, transcription of the SaPI genome is repressed (Figure 1.7). Unlike the phage repressor that undergoes auto-lysis once the bacterial SOS-response has been triggered, the SaPI Stl requires interaction with a phage-encoded protein inducer to initiate island derepression^{280,304-306}. The targeted phage-encoded inducer proteins have an entirely different primary role for the phage, only ‘moonlighting’ as SaPI inducers²⁸⁰. These moonlighting phage proteins bind to the Stl, which dissociates from the intergenic DNA on formation of the phage protein:Stl complex. At this point, the SaPI genes can be expressed for initiation of the ERP cycle.

As previously shown, the Stl repressors encoded by different SaPIs have divergent sequences and so require different phage-encoded proteins for derepression (Figures 1.7 and 1.9)²⁸⁰. Because of this, some phages encode multiple ‘moonlighting’ SaPI inducers to activate a number of SaPIs, whilst others encode none and so induce none. For example, phage $\phi 80\alpha$ induces SaPIbov1, SaPIbov2, SaPIbov5, SaPI1, and SaPI2, indicating that it encodes many SaPI inducers^{217,268,270,307,308}. On the other hand, phage $\phi 80$ only induces SaPIbov1 and SaPI2, whilst phage $\phi 85$ does not induce any of these SaPIs³⁰⁷.

Four phage-encoded ‘moonlighting’ proteins have now been identified as specific SaPI inducers; the phage-encoded trimeric dUTPases (Duts) induce SaPIbov1 and SaPIbov5, the Sri protein induces SaPI1, the $\phi 80\alpha$ ORF15 induces SaPIbov2, and the phage-encoded recombinases induce SaPI2 (Figure 1.9)^{280,304,305,309,310}. The phage Dut proteins’ primary role is regulating dUTP levels by hydrolysing dUTP

to dUMP, but these proteins also bind the identical Stl repressors encoded by SaPIbov1 and SaPIbov5^{304,305,311,312}. Likewise, the Sri protein was first identified as a DnaI-binding protein encoded by phage 77³¹³. Sri binds and blocks the bacterial host helicase, DnaI, to inhibit bacterial DNA replication, but can also bind to the Stl of SaPI1. The phage ϕ 80 α ORF15 has an unknown function for the phage, but binds the Stl of SaPIbov2²⁸⁰. Most recently, the phage recombinases have been shown to play an essential role in phage replication, but are also capable of binding the Stl of SaPI2^{309,310}.

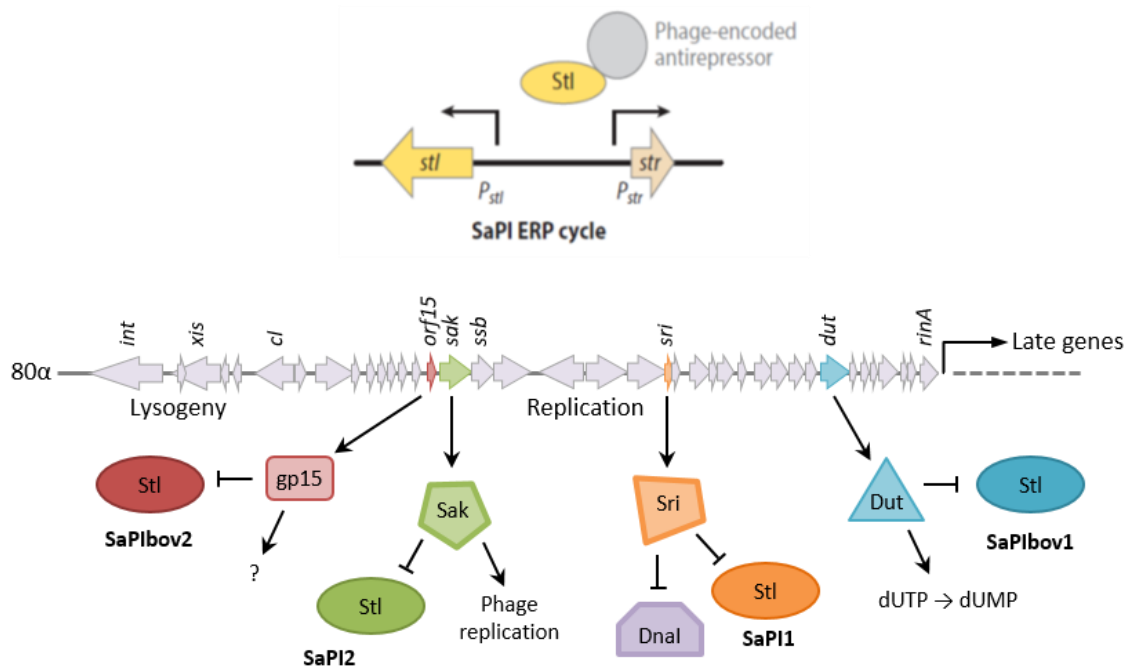


Figure 1.9 SaPI induction by a helper phage. Interaction with a phage-encoded anti-repressor leads to relief of Stl repression of both the *stl* and *str* promoters (P_{stl} and P_{str}), inducing the SaPI excision-replication-packaging (ERP) cycle. Different phage moonlighting proteins function as anti-repressors for different SaPIs. The locations of the known phage-encoded derepression genes are illustrated on the map of the left end of the 80 α genome. Each of these phage gene products targets a different Stl repressor, encoded by a different SaPI, as shown. The known moonlighting functions of these gene products during the phage life cycle are also indicated. Gp15, unknown function; Sak, recombinase; Sri, DnaI binding protein; Dut, dUTPase. Adapted from Penadés and Christie, 2015²⁸³.

The variation in the phage-encoded SaPI inducers' primary phage-related functions indicates the evolutionary pressure for escape of SaPI induction. SaPI induction is inherently damaging for the helper phage, leading to reduced levels of phage transfer²⁷⁹. As such, the phages have continually strived to escape SaPI induction through the exchange or mutation of the phage-encoded inducers, with selection for novel variants with a low affinity to the Stl repressor²⁸¹.

Phages have been described as reducing expression of phage-encoded SaPI inducers in order to avoid detrimental SaPI interference^{280,281}. In this way, the Stl targeting of these phage-encoded moonlighting proteins has driven staphylococcal phage evolution.

On the other hand, this constant inducer protein evolution has challenged the Stl to overcome such escape mechanisms. As ever-increasing varieties of phage-encoded SaPI inducers were created, the Stl repressors co-evolved to target these new proteins. This goes some way to explaining the divergence in sequence seen between the Stl repressors encoded by different islands. Similarly, the targeting of conserved or essential proteins like the phage Duts and recombinases makes phage escape of SaPI induction more difficult, as mutations could inactivate proteins with an important function for the phage. In this way, these MGEs are trapped in a continual co-evolutionary arms race.

1.3.3.2 Assembly and capsid size redirection

One of the most observed strategies of SaPI molecular piracy is the modulation of the phage-encoded capsid assembly pathway for the formation of small SaPI particles^{292,294,295,314}. This mechanism is not limited to the SaPIs, being found more widely throughout the PICI family and in other MGEs^{176,264,286,315}. The formation of small capsids allows the packaging of SaPI genomes in capsids proportionate to their size (~15 kb). Furthermore, it has a secondary effect of reducing phage genome packaging, as the larger helper phage genomes are not functionally packaged in SaPI-sized capsids. Such redirection produces virion particles with $T=4$ icosahedral symmetry, as opposed to the usual $T=7$ phage particles (Figure 1.10)^{101,267,294,295,314,316}. There are several capsid size redirection mechanisms, suggesting this interference function confers an important evolutionary advantage over the helper phage.

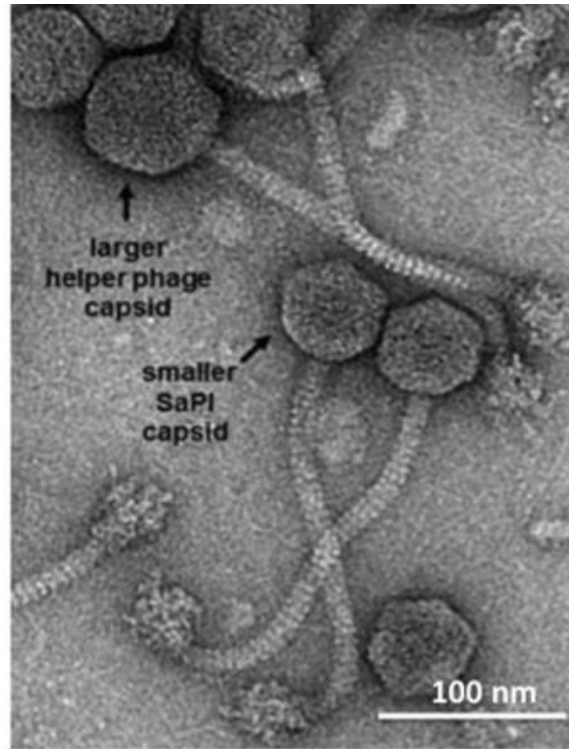


Figure 1.10 Electron micrographs of a mixed helper phage-SaPI lysate. From Novick and Ram 2016, reproduced with permission of Elsevier¹⁰¹.

Some SaPIs control the modulation of capsid size by expressing CpmA and CpmB, which are encoded in the packaging module (Figure 1.6)^{292,294}. These proteins redirect the procapsid size, but are removed after the formation of the procapsid and are not associated with the mature shell (Figure 1.11). CpmB is an internal scaffolding protein, forming dimers with the phage shell proteins²⁹⁴. This protein competes with the phage-encoded scaffolding proteins and shapes the geometry of the capsid to form the reduced icosahedral³¹⁴. Less is known about the role of CpmA. The CpmA protein is found in low copy numbers during the assembly process, indicating it could have a transitory function²⁹⁴. For instance, CpmA could be a regulator of scaffolding core organisation, to prevent the formation of aberrant small capsids. CpmA could also control the access of CpmB to the capsid proteins, stopping phage-encoded scaffolding proteins from binding. Again, the SaPIs encode proteins exhibiting phage interference, but remain interlinked with the phage life cycle for their transfer.

Such capsid size modulators are only able to function in the presence of a compatible helper phage capsid system. Indeed, the efficiency of small capsid formation has been shown to vary between the same SaPI, depending on what

phage is present. For example, SaPI2 can form small capsids for mobilisation in the presence of phage $\phi 80\alpha$, but not with phage $\phi 80$. This is because the $\phi 80$ capsid protein is incompatible with the SaPI2 CpmA and CpmB³⁰⁷. Similarly, size re-modelling in the presence of phage $\phi 80\alpha$ is much more efficient for SaPI1 over SaPIbov1, even though their CpmA and CpmB proteins are almost identical³¹⁷. This suggests that other factors are involved in capsid size modulation.

Not all SaPIs favour packaging into small capsids. SaPI dsDNA can be packaged in both small and phage-sized capsids, and the preference between these two options depends on the SaPI. For example, SaPI1 mainly uses small capsid for dsDNA packaging, whereas SaPIbov1 packages in small and normal sized capsids with near equal efficiency²⁹³. Indeed, surprisingly, both SaPIbov1 and SaPI1 *cpmAB* mutants are transduced at high frequency utilising phage-sized particles^{279,293,294}. Not all SaPIs encode the CpmAB morphogenetic proteins. One such SaPI is the SaPI_{pT1028}, which lacks the *cpmAB* genes. SaPI_{pT1028} is packaged in full-sized phage particles and so allocates three copies of its 15 kb genome in each particle²⁶⁸. Likewise, SaPIbov2 has no CpmB protein and is too large to fit into small capsids (27 kb)²⁶⁸. Therefore, two copies of the SaPIbov2 genome are packaged into each phage-sized capsid. Indeed, variation in packaging modulators seems fairly common, with SaPIbov2 encoding *terS*, *cpmA*, *ptiM*, and *ptiB* genes more similar to the *Staphylococcus epidermidis* pathogenicity islands (SePIs) than any other SaPI. Furthermore, SaPIbov5 (13.5kb) has no homologous proteins matching the established operon I³⁰⁸.

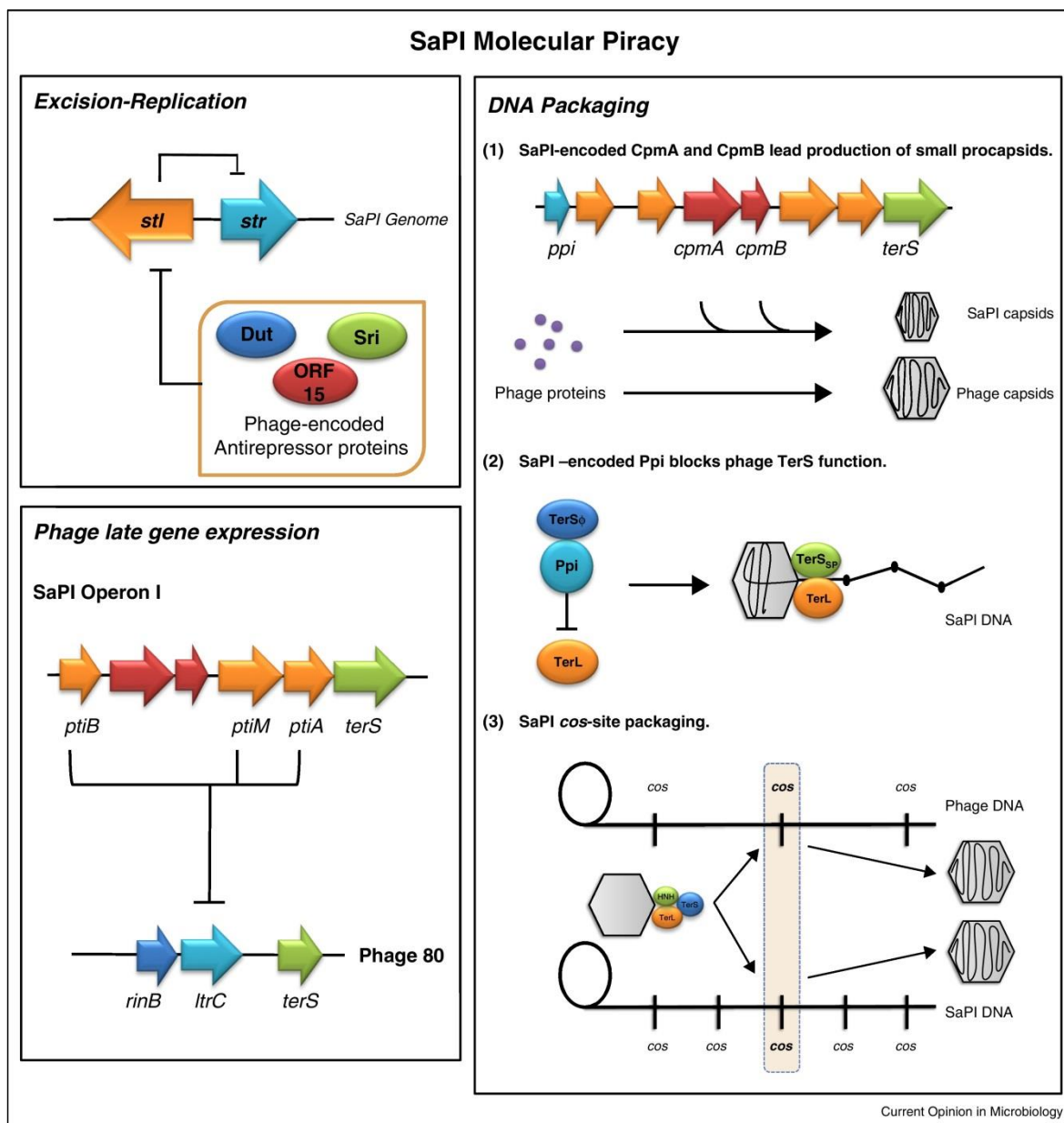


Figure 1.11 Mechanisms of SaPI interference with phage reproduction. Phages induce the SaPI life cycle by expressing proteins that block the activity of the SaPI-encoded StI repressor. SaPI induction blocks phage reproduction using alternative and complementary strategies. These include blockage of the phage LtrC regulator by the PtiABM proteins, generation of SaPI-sized small capsids by expression of the CpmAB proteins, blockage of the phage TerS by the Ppi protein, or package of the SaPI DNA in phage particles by carrying the phage *cos* site in the SaPI genome. Reproduced from Penadés *et al.* 2015, with permission from Elsevier²⁰⁹.

1.3.3.3 DNA packaging and phage interference mechanisms

As described in section 1.2.10.2, phages package their concatemeric dsDNA using one of two different mechanisms, which both rely on a phage-encoded terminase TerS-TerL complex (TerS ϕ and TerL ϕ)^{173,318}. Following TerS ϕ -mediated recognition of a specific *cos* or *pac* sequence in the phage genome, the TerL ϕ cleaves the site and translocates the dsDNA into the procapsid (Figure 1.5). With

pac phages this occurs in a headful manner, whilst *cos* phages are packaged in unit lengths. The SaPIs exploit both these packaging mechanisms via alternative strategies (Figure 1.11).

Classical SaPIs use the headful packaging system, but mediated by their own TerS_{SP}^{279,293,319}. As with the phages, SaPI replication forms concatemeric DNA. The TerS_{SP} recognises the SaPI *pac* site and, having formed a complex with the TerL ϕ , directs cleavage to this *pac* sequence before translocation of the SaPI dsDNA into the capsid. The SaPI *pac* site is located in the intergenic region upstream of the *terS_{SP}* gene and cleavage here leaves redundant ends that enable the SaPI to circularise on infection of a new bacterial cell³¹⁹.

The binding of TerS_{SP} to the specific SaPI *pac* site generates high frequencies of SaPI packaging, and deletion of the TerS_{SP}-encoding gene leads to decreased SaPI transfer equivalent with generalized transduction levels^{286,293}. However, TerS_{SP} alone does not mediate phage interference, with normal levels of phage titre in strains containing a TerS_{SP} SaPI mutant^{286,293}. For the TerS_{SP} to promote translocation of SaPI dsDNA, the TerL ϕ and the TerS_{SP} must be compatible. The *cos* phage ϕ 13 is capable of inducing the *pac* SaPI1, but no SaPI virions are produced because the *cos* phages are not compatible with the *pac* SaPI system (TerS_{SP})²⁷⁰.

The *pac* SaPIs also encode a phage packaging interference (Ppi) protein that confers an advantage for SaPI packaging (Figure 1.11). The Ppi binds to the phage TerS subunit to promote TerS_{SP} binding to TerL ϕ ^{279,286}. The mechanism by which the Ppi binds the TerS ϕ is not yet known. Structural differences between the TerS_{SP} and the TerS ϕ prevent Ppi interference with the TerS_{SP}^{291,295}. Two classes of Ppi have been identified so far; class I from SaPIbov1 interferes with phages ϕ 80 α and ϕ 11, whilst class II from SaPI1 interferes with ϕ 12²⁷⁹. Ppi homologues have also been identified in other Gram-positive bacteria, including *Eubacterium limosum* KIST612, *Aerococcus viridans* ATCC 11563 and *Clostridium bolteae* ATCC BAA613²⁷⁹. This suggests that expression of Ppi proteins is a widespread SaPI and PICI mechanism for interference with phage packaging.

Although headful packaging interference is more common, an alternative strategy is employed by the *cos* SaPIs. The *cos* SaPIs have substituted the TerS-

morphogenetic operon I for a ~90 bp fragment that contains a phage *cos* site homologue³⁰⁸. SaPIbov4 and SaPIbov5 encode *cos* sites and can be induced by certain *cos* phages, including ϕ 12 and ϕ SLT. Although these islands lack the classical operon I, they severely interfere with phage reproduction, implying an alternative interference strategy³²⁰. The *cos* SaPIbov5 has been shown to be capable of two different packaging mechanisms. In one, *pac* phages (ϕ 11 and ϕ 80 α) induce SaPIbov5 and package it into full-size capsids using the phage-encoded TerS-TerL terminase. In the second system SaPIbov5 is packaged by the *cos* phages ϕ 12 and ϕ SLT. SaPIbov5 encodes similar *cos* sites and flanking regions to these *cos* phages, allowing for SaPI DNA encapsidation by these phages³⁰⁸. One novel feature of the *cos* phage-SaPI packaging system is the requirement of a HNH nuclease on top of the TerS-TerL complex. Therefore, the *cos* SaPIs can parasitize the HNH-mediated *cos*-site packaging of the *cos* phages³⁰⁸. Furthermore, it has recently been shown that these *cos* SaPIs are capable of modulating capsid size using the Ccm protein, through an unknown mechanism. The effects of this Ccm protein are reminiscent of the operon I encoded CpmAB proteins³²⁰.

1.3.3.4 Phage late gene transcription interference

Certain SaPI-encoded proteins have recently been identified as interfering with the reproduction of the helper phage. Classified as the phage transcription inhibitor (Pti) proteins, these modulate expression of the late gene transcription operon of some helper phages^{169,321}. This operon is related to morphogenetic, assembly, packaging, and lysis functions that are required by both phage and SaPI. Transcription of the phage late module is modulated by the super family of late transcription regulators (Ltr), including RinA, ArpU, LtrA, LtrB, and LtrC^{169,170}.

The SaPI Pti proteins interfere with phage reproduction by binding to the Ltr proteins. However, the different Pti variants encoded by the SaPIs differentially affect the Ltr variants of certain helper phages. For example, the SaPI2 PtiA blocks phage ϕ 80 reproduction by binding the phage LtrC protein, but does not interfere with RinA. Furthermore, the SaPI2 PtiM protein negatively regulates PtiA to modulate LtrC repression^{170,307,322}. It seems that SaPI-mediated

interference with phage late gene transcription is a finely balanced equation, as the SaPI needs the packaging machinery of the phage for encapsidation.

1.3.4 SaPIs as vehicles for horizontal intra- and intergeneric virulence gene transfer

The parallelisms and inherent association between the SaPIs and their helper phages highlight the co-evolution of these two MGEs, enabling the efficient phage-mediated spread of the SaPIs throughout the genomes of *S. aureus* strains. Indeed, sequencing of pathogenic strains of *S. aureus* has indicated that these strains are poly-lysogenic and can possess more than one island in their genome^{78,246,247,254,260,323}. Furthermore, studies have shown a wide-ranging distribution of SaPIs within different species of the staphylococcal genus. For example, transfer of *pac* SaPIs like SaPIbov2 can occur between *S. aureus* and other staphylococci, including *S. epidermidis* and *Staphylococcus xylosus*²⁶⁸. Evidence has also been found of SaPI transfer to bacterial strains outside of the staphylococci, indicating inter-generic transfer. For instance, both *pac* SaPIs SaPI1 and SaPIbov1 can be transferred to *Listeria monocytogenes*²⁸⁸. Additionally, recent studies have identified *cos* SaPI inter- and intra-generic transfer, with SaPIbov5 transferred to *S. epidermidis*, *S. xylosus*, and *L. monocytogenes* strains²¹⁷.

Typically, the host range of a phage has been classified as the bacteria which that phage can lyse, visualised by the formation of plaques in a lawn of recipient bacteria. This had led to the assumption that phages have a relatively narrow host range³²⁴. However, it has become apparent that, with regard to MGE transfer, phages can undergo ‘silent’ transfer. Silent transfer describes the capability of phages to inject dsDNA into a recipient cell in which they themselves cannot grow. This has been shown for both *pac* and *cos* phages^{217,268,288}. Transfer of SaPIs requires the injection of SaPI dsDNA into an appropriate recipient strain and the circularisation and integration of this DNA into host *att_C* sites. However, SaPI transfer does not require the ability of the helper phage to infect this new host. As such, silent transfer mediates transfer of SaPIs to bacterial hosts that the phages themselves cannot parasitize^{217,268,288}.

SaPIs also have strategies that enable SaPI mobilisation by non-inducing helper phages. When a phage encodes a combination of proteins that does not include the phage-encoded SaPI inducer it has, in essence, ‘escaped’ SaPI parasitism through the SaPI ERP cycle. As such, SaPIs that encode alternative mobilisation strategies gain a significant advantage. Such mobilisation facilitates SaPI transfer through generalised transduction following mispackaging in non-inducing phages, albeit at low frequencies^{268,269}. Such transfer only requires the SaPI genome to be packaged into a phage particle and subsequently injected into a new cell. Following injection into the new host, the SaPIs can integrate at an *att_C* site in the host chromosome using one of two pathways.

In the first system, SaPI mobilisation is *Int* and *Xis*-dependent, but *RecA*-independent. Following non-inducing phage generalised transduction of the SaPI into a new host, the *Int* and *Xis* mediate SaPI excision, circularisation, and integration into an *att_C* site in the recipient strain. The second system involves homologous recombination of the SaPI flanking sequences, mediated by *RecA*. *Int*-dependent SaPI transfer is higher than classical generalised transduction, but SaPI-induction and packaging by an inducing helper phage are necessary for highly efficient SaPI transfer²⁶⁸.

As well as facilitating their own transfer, SaPIs can contribute to the generalised transduction of chromosomal DNA through the SaPI-encoded *TerS_{SP}*^{319,325}. The host chromosome contains a number of pseudo *pac* sequences scattered throughout the host chromosome, to which *TerS_{SP}* can bind. Binding of *TerS_{SP}* to these pseudo *pac* sites initiates mispackaging of downstream chromosomal fragments³²⁵. Often these downstream fragments contain genes encoding virulence and adaptation. Because expression of the SaPI-encoded operon I module encoding the *TerS_{SP}* is mediated by a *LexA* promoter, this region can be activated by the SOS-response²⁹³. Hence, SOS-response activation of a helper prophage also promotes expression of the SaPI-encoded *terS_{SP}*, *pti* and *cpm* genes, independent of SaPI induction. Consequently, the *TerS_{SP}* will bind any *pac* or pseudo *pac* sites in the host chromosome, initiating cleavage and packaging of these regions by the phage *TerL*. This produces SaPI transducing particles containing fragments of bacterial DNA, potentially encoding virulence and adaptation factors. In this way, *TerS*-encoding SaPIs and more widely PICIs also

contribute to HGT in bacteria to modulate niche adaptation and evolution of pathogenicity.

1.3.5 Biology of non-SaPI PICI elements

The discovery of PICIs as the wider family to which the SaPIs belong began with the identification of SaPI-like element in non-*S. aureus* staphylococcal strains. They have since been described in a variety of different Gram-positive genera, including *S. pyogenes* (SpyCIM1), *E. faecalis* V583 (EfCIV583) and *L. lactis* (LICiBIL312)^{262,274,275}. The PICIs display a modular gene structure grouped by function, as seen with the SaPIs (Figure 1.6)²⁶². Certain genes are conserved throughout the PICI family and share the same positioning within the genome. The *int* gene sits at the start of the genome, followed by a *stl-str*-like transcription regulation system with divergent genes. The *pri-rep* genes are conserved between PICIs and *terS* homologues have been identified in a number of cases. Increasing amounts of PICI-like gene clusters are being found by sequence analyses of bacterial chromosomes²⁷⁷. However, *in vitro* and *in vivo* studies are necessary to determine whether these elements are functional and behave in a PICI-like manner.

EfCIV583 was initially, erroneously identified as a prophage (ϕ V7), before this classification was corrected and the name changed to PICI-appropriate nomenclature²⁶². Further investigation identified the phage ϕ p1 as the helper phage, although this study proposed that EfCIV583 was SOS-induced and only required the phage for packaging²⁷⁵. More recently, EfCIV583 has been established as behaving in a typical SaPI-like manner. The latest study found that EfCIV583 is induced by the ϕ p1-encoded Xis protein and can be packaged in phage-sized particles or EfCIV583 small particles. This follows the classic SaPI life cycle of phage-mediated induction followed by packaging independent of capsid size, but with potential phage interference through the creation of small-sized capsids²⁷⁶. Therefore, EfCIV583 behaves as a classic PICI element.

Similarly, the SpyCIM1 is a PICI that was first described in *S. pyogenes* M1 genome strain SF370^{274,326}. As with the SaPIs, this PICI integrates at the *att_C* site in the bacterial chromosome. Interestingly, this integration site of SpyCIM1 separates the bacterial DNA mismatch repair genes *mutS* and *mutL*, so SpyCIM1 PICI

integration/excision acts as an on/off molecular switch for the host²⁷⁴. Excision of *SpyCI_{M1}* during exponential host cell division guarantees DNA replication fidelity, whilst integration promotes the generation of mutations within strains that could enable adaptation to different environmental pressures.

Other *SpyCI_{M1}*-like elements have been identified in the *Streptococcus* genus³²⁷. Although they otherwise have typical PICI characteristics, these islands differ in that they encode a repressor with a LexA- S24-like (cd06529) domain^{274,326}. This domain is inducible by the bacterial SOS-response, unlike the classical PICI repressors that require a specific phage-encoded protein. Furthermore, none of these elements encode an identifiable TerS homologue and their method of dsDNA packaging remains unknown³²⁷. Therefore, the *SpyCI_{M1}*-like elements seem to form a group of unconventional PICI elements.

An increasing number of elements are being termed 'PICI-like'. For example, the PICI-like element (PLE) from Gram-negative *Vibrio cholera* has been proposed to be part of the PICI family because it seems to interfere with phage reproduction³²⁸. However, although this PLE encodes an Integrase, it has no homologues of the regulatory, replication or packaging modules. As such, more studies are needed to confirm whether this PLE is indeed part of the PICI family.

1.4 The dUTPases, metabolic enzymes with a moonlighting signalling role

As mentioned in section 1.3.3.1, different SaPI Stl repressors require different, specific phage proteins for derepression of the island. For the identical SaPI_{bov1} and SaPI_{bov5} Stl repressors this phage-encoded protein is the trimeric dUTP pyrophosphatase (dUTPase; Dut)²⁸⁰. The classic definition of the dUTPases (Duts) is as ubiquitous enzymes that regulate cellular levels of dUTP to prevent the misincorporation of uracil into DNA³²⁹. However, more recently Dut proteins from different models have been implicated in a range of cellular processes, including regulation of the immune system, autoimmunity, and apoptosis³³⁰⁻³³⁴. Furthermore, as phage-encoded SaPI induces the trimeric Duts of phages $\phi 11$ and $\phi 80\alpha$ control the transfer of MGEs^{304-306,311,312}. This implies that certain Duts may have previously unexplored roles outside of dUTP hydrolysis.

1.4.1 The enzymatic function of the Duts

One of the main chemical differences between DNA and RNA is thymine/uracil exchange. The only difference between uracil and thymine is a 5-methyl group that is only found in thymine bases³²⁹. However, this region is not involved in interaction with adenine, so Watson-Crick base pairing with adenine occurs in the same manner for both bases. Most DNA polymerases do not distinguish between uracil and thymine and will incorporate either base depending on the ratio of dUTP or dTTP building block nucleotides available. Furthermore, cytosine deamination converts cytosine to uracil, resulting in mutagenic U:G mismatches that need to be excised³³⁵. The most efficient repair system for U:G mismatches is the major uracil-DNA glycosylase (UNG), which does not distinguish these mismatches from U:A “normal pairs”³³⁶. As such, all uracil-substituted DNA is subjected to uracil-excision repair³³⁷. However, if there is a high dUTP/dTTP ratio, uracil will be reincorporated in the DNA during the base excision repair process. This leads to hyperactive base excision repair, resulting in DNA double strand breaks and chromosome fragmentation until thymine-less cell death occurs³³⁸.

In their classical role, the trimeric Dut enzymes catalyse the cleavage of dUTP to dUMP. This performs two roles, preventing the build-up of dUTP in the cell and producing dUMP, which is the precursor to dTMP synthesis (Figure 1.12)³²⁹. The creation of dUMP by Duts is particularly important for Enterobacteria, Mycobacteria and Plasmodia, as no dCMP deaminases are present to provide the direct input into the thymidine synthesis pathway seen in other organisms^{329,339,340}. Dut knock outs are lethal in *E. coli* and yeast, and it has been assumed that Duts are encoded by almost all free-living organisms and some viruses³⁴¹⁻³⁴³.

The main enzymatic steps of the homotrimeric Duts described so far comprise of rapid substrate binding followed by a relatively slow adoption of the catalytically competent active site conformation. Hydrolysis then occurs and the products are rapidly released³⁴⁴. Trimeric Duts form three active sites in symmetrical fashion and conserved sequence motifs from all three subunits interact with each site making trimer formation indispensable. The active site is

formed to precisely fit a uridine moiety found on the dUTP, conferring specificity³²⁹.

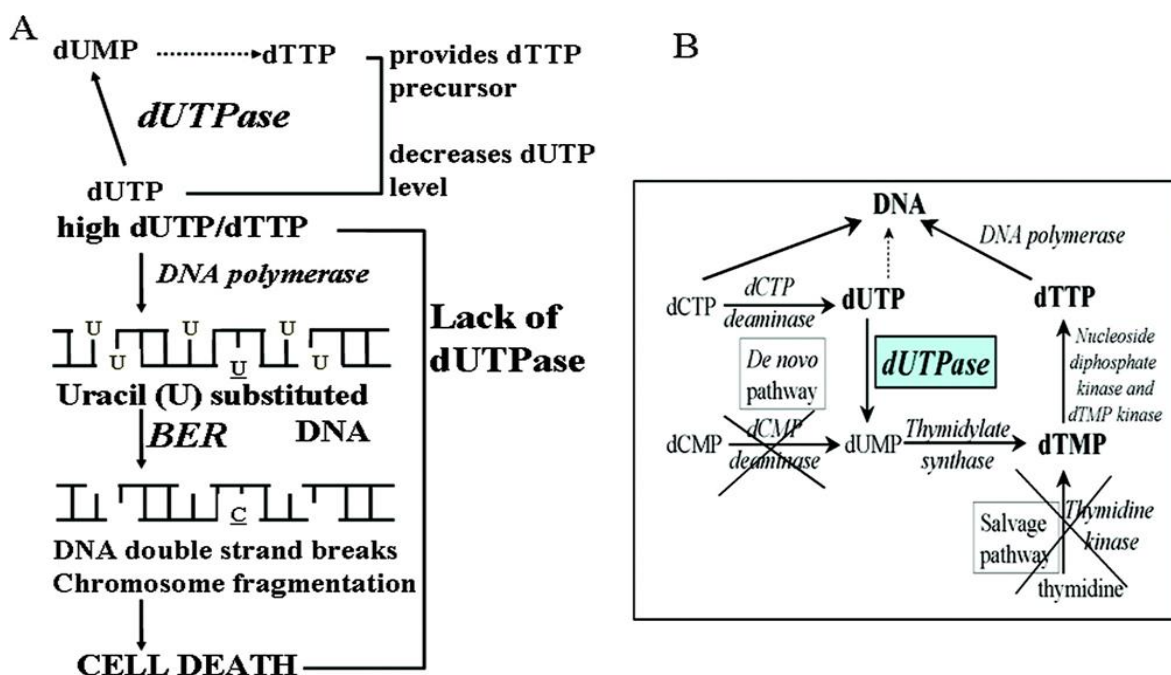


Figure 1.12 The classic enzymatic role of the Duts. (A) Dual role of Dut (dUTPase). In the absence of Dut (i.e., when dUTP/dTTP ratio is high), thymine-replacing uracils will be reincorporated during base excision repair (BER), but cytosine deamination can be correctly repaired (underlined bases U, C). B. *De novo* and salvage pathways for dTTP biosynthesis. Enzymes with a cross are not present in *Enterobacteria*, *Mycobacteria* and *Plasmodia*. Reproduced from Vértessy and Tóth. 2009³²⁹, with permission from American Chemical Society publications.

Recently, the ubiquity of the Dut proteins has been called into question. One study performed bioinformatics investigations into the presence of Duts and found a wide number of bacterial and Archaeal species, including *S. aureus*, lacked any known Dut-encoding gene. These *dut* negative strains were found over a range of phyla, suggesting *dut* loss was not a single, isolated event³⁴⁵.

Similarly, some studies have brought into question the assumption that all DNA uracil incorporation is negative. In some instances, it is absolutely required. For example, uracil found in the chromosome of B cells favoured the generation of antibody diversity³⁴⁶. In the human immunodeficiency virus, DNA generated following primary immune cell infection is highly uracilated to protect the viral DNA from suicidal autointegration³⁴⁷. Furthermore, research on *Drosophila* models found that *Drosophila* can tolerate uracil accumulation in the DNA, indicating a potential role of uracil for metamorphosis in the *Drosophila*³⁴⁸. Dut

proteins are present in all these scenarios, but uracil is necessarily incorporated into the DNA. One theory to explain this apparent repression of the classic Dut dUTP hydrolysis role is that the Duts perform additional functions.

1.4.2 Types of Dut

The Duts have been divided into three families, or types, based on their oligomerisation state. As such, Dut proteins can be monomeric, homodimeric, or homotrimeric enzymes (Figure 1.13). The homotrimeric Duts are found in most eukaryotes, prokaryotes, some DNA viruses, and some retroviruses that infect non-primate cells. This makes the homotrimeric Duts the largest family. A number of homotrimeric Dut structures have been solved, with several X-ray structures available. The homotrimeric Duts are made up of mainly β -pleated sheets, with the *E. coli* and human Duts being the prototypical models^{349,350}. Every protomer from the trimer interacts within the individual active sites of each subunit, linking trimer formation together for enzymatic function³⁴⁹.

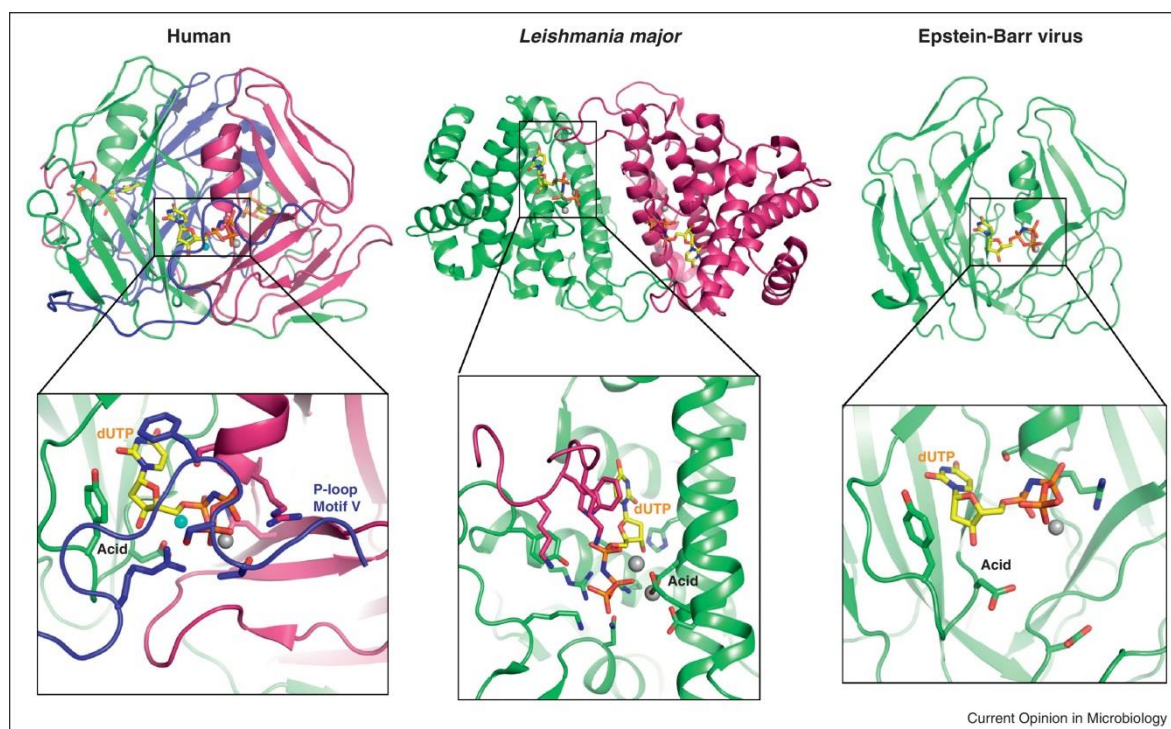


Figure 1.13 Dut types and their active sites. The structures of the different types of Duts are shown in cartoon with the substrate dUTP in sticks. Homotrimeric (protomers in blue, green and deep-red), homodimeric (green and deep-red) and monomeric (green) enzymes are exemplified by the structures of human (PDB code 3EHW), *Leishmania major* (PDB: 2YAY) and Epstein-Barr virus (PDB: 2BT1) Duts, respectively. Close-up views of the respective active sites are shown with the nucleotide and the conserved residues responsible for the catalysis and the substrate binding in sticks. Substrate carbons are coloured in yellow and enzyme carbon atoms in the colour of the corresponding subunit. Non-carbon atoms are coloured red, blue and orange for oxygen, nitrogen

and phosphorus, respectively. The catalytic acidic residues are labelled. Mg ions are shown as grey spheres. The catalytic water in the trimeric Dut is shown as a cyan sphere. In this active centre, the nucleotide is covered by the P-loop motif V. Note that the motif V loop is nonvisible in the monomeric Dut, even if the nucleotide triphosphate was present in the active centre. Reproduced from Penadés *et al.* 2013, with permission from Elsevier³¹².

The trimeric Duts contain five conserved motifs, which make up the active site. These motifs are distributed over the entire sequence. The first four motifs are well-ordered in all known structures (motifs I-IV), however conserved motif V is often disordered and sometimes only observed in a Dut:non-hydrolysable dUTP analog complex^{351,352}. The motif V is a P-loop motif corresponding to the C-terminal region of the protomers. The trimeric Dut motif V shares some features with the P-loops present in many ATPase and GTPase enzyme families, known as P-loop NTPases³⁵³. Motif V is necessary for substrate discrimination and dUTP hydrolysis, as the flexible motif V undergoes a conformational change to cap the active site following ligand binding^{353,354}.

The homodimeric enzymes make up the second family of Duts, or the type 2 Duts. The dimeric Duts are found in the trypanosomatids *Leishmania major*³⁵⁵, *Trypanosoma cruzi*³⁵⁶ and *T. brucei*³⁵⁷, in some bacterial species including *Campylobacter jejuni*³⁵⁸, and in some bacteriophages³⁵⁹. The dimeric Duts are all α -helical and the dimeric Duts encoded by different organisms share a basic module, made up of a rigid domain and a mobile domain (Figure 1.13). The sequences and structures of the different dimeric Duts are completely different to the trimeric Duts. However, in one parallelism to the trimers, the active sites of the dimeric Duts are made up of five motifs. These dimeric Dut motifs are very different to those found in the trimers, though. The dimeric Duts can also hydrolyse both dUTP and dUDP, with different biochemical properties and specificity than seen with the trimeric family³⁵⁵⁻³⁵⁸.

Lastly, the monomeric Duts make up the third family (Figure 1.13). Monomeric Duts are found in mammalian and avian herpesviruses³⁶⁰. Interestingly, these proteins have limited homology to the trimeric Duts, but contain the characteristic five motifs that contribute to the active site³⁶¹. However, in the monomeric enzymes the motifs reshuffle and spread out over a single polypeptide that is double the sequence length of the trimeric Duts. The structural folds of the monomeric Duts mimic those of the trimeric active site

and, indeed, the monomeric active site is nearly identical to the trimeric family. All this indicates that the monomeric Duts have evolved from a gene encoding for a trimeric Dut. A new family of monomeric Dut has been identified that differs to that of the herpesviruses. The monomeric Duts of *Drosophila virilis* and *Caenorhabditis elegans* fuse three copies of a trimeric monomer into a single polypeptide, which is predicted to fold like the trimeric Duts³⁶². It is not yet known how these fused enzymes work.

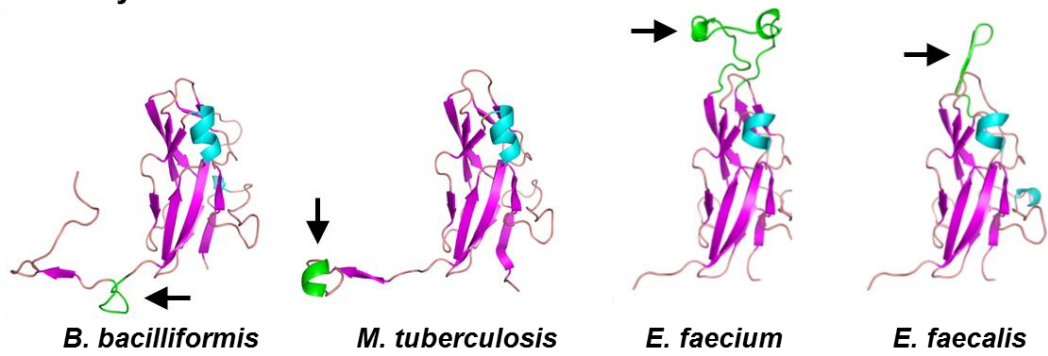
The existence of these different forms of Dut enzyme indicates they are under selective pressures specific to their environment. If their only role was the control of dUTP-levels, then this divergence in structure would probably not be necessary. This is further supported by the existence of extra domains in certain Duts that are involved in non-catalytic functions.

1.4.3 Moonlighting signalling roles of Duts

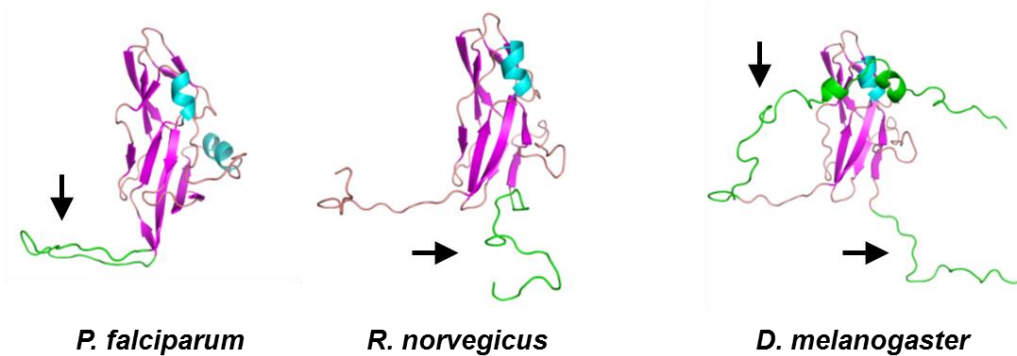
Some Duts encode an extra domain, named motif VI, outwith the five conserved domains required for the enzymatic function. The motif VI extra domains found in some homotrimeric Duts vary between the Dut proteins from different organisms (Figure 1.14). The motif VI is unconserved in both length and position between different Duts. For example, in the *Rattus*, *Plasmodium* and *Drosophila* Duts the motif VI is located in the N-terminal, whilst in *Bartonella*, *Mycobacterium* and alternative *Drosophila* Duts it is found in the C-terminal^{329,334,363}. The extra domains in enterococcal and staphylococcal Duts are positioned centrally (Figure 1.14)³⁰⁴⁻³⁰⁶.

The extra domains have been implicated in the secondary function of Duts as signalling molecules. For example, it has been shown that the rat (*Rattus*) Dut protein interacts with the transcriptional factor PPAR α (peroxisome proliferator-activated receptor alpha) to repress PPAR-mediated transcriptional activation³³⁴. This interaction depends on the motif VI, as well as the C-terminal domain. Likewise, the *D. melanogaster* Dut has a species-specific motif VI encoded at the C-terminal³⁶⁴. The *D. melanogaster* Dut has been implicated in early pupal development, with deletion mutants specifically lethal at this development stage and not in others³⁴⁸.

Prokaryotic Duts



Eukaryotic Duts



Phage Dut

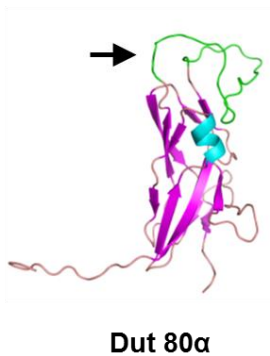


Figure 1.14 Monomer structures of trimeric Dut proteins and localisation of the extra motif VI. The structures of the *Mycobacterium tuberculosis*, *Plasmodium falciparum*, *Drosophila melanogaster* and staphylococcal 80α Dut proteins have been previously experimentally solved. *Bartonella bacilliformis* and enterococcal (*E. faecium* and *E. faecalis*) proteins were modelled against the *Brucella melitensis* and *Streptococcus mutans* Duts, respectively, using Swiss-Model server (<http://swissmodel.expasy.org/workspace/>). Rat Dut was modelled using SAM-T08 server (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html). The conformation of the insertional motif VI is highlighted in green and pointed out by black arrows. Note that the extra domains are located in the N-terminal region (*Rattus*, *Plasmodium* and *Drosophila*), C-terminal region (*Bartonella*, *Mycobacterium* and *Drosophila*) or in the middle (staphylococcal and enterococcal) of the Dut proteins. Although with a different localisation, the extra motif VI from the *Rattus* and the *Mycobacterium* Duts are also involved in signalling. Adapted from Tormo-Más et al. 2013, with permission from Elsevier³⁰⁵.

The Dut protein encoded by *Mycobacterium smegmatis* carries a motif VI domain of five residues in the C-terminal region that is not associated with catalytic activity, but is essential for bacterial viability³⁶⁵. Indeed, when a deletion mutant was complemented with a Dut that was enzymatically functional, but missing this five-residue motif VI the phenotype was unchanged. This suggests a regulatory Dut function conferred by the motif VI that is essential to the bacteria. Similarly, the pathogenic *Mycobacterium tuberculosis* and *Mycobacterium leprae* also encode Duts with a motif VI present, although no signalling function has been found for these Duts as yet^{363,366}.

It is interesting that viruses code for Dut proteins, as generally the host cells they infect will already encode one. It has been proposed that the viral Duts are linked to division in non-dividing cells, but given that there are viruses in nature that replicate efficiently without their own Dut this seems unlikely³⁶⁷. Indeed, studies have indicated that some viral Duts play a role in the modulation of the host immune system. For example, the Epstein-Barr virus (EBV) Dut is involved in one of two pathways EBV uses to control cytokines for promoting infection of monocytes and macrophages³³⁰. The EBV Dut interacts with host Toll-like receptor 2 (TLR2) to recruit the adaptor molecule MyD88, which triggers activation of the NF- κ B pathway. This pathway leads to upregulation of IL-10 and the proinflammatory cytokines TNF- α , IL-1 β , IL-6, and IL-8 in host macrophage cells.

Similarly, the viral MHV-68 Dut is not necessary for virus reproduction, but does block expression of the type I interferon signalling pathway³³¹. Furthermore, the Dut encoded by the Kaposi's Sarcoma-associated herpesvirus targets several cytokine receptors to downregulate the host immune response. These Kaposi's Sarcoma-associated herpesvirus Dut-mediated effects are not dependent on the enzymatic function of the protein. Additionally, the Kaposi's Sarcoma-associated Herpesvirus (KSHV) Dut downregulates NKp44L³³². NKp44L is one of the ligands expressed on virus-infected cells that natural killer (NK) lymphocytes recognise to target cells for destruction. By downregulating expression of this ligand, the KSHV Dut prevents destruction of KSHV infected cells. Again, the KSHV Dut is not dependent on its enzymatic function for this effect. This suggests the herpesvirus-encoded Duts play a role for virus-mediated immune regulation.

1.4.4 Phage-encoded trimeric Duts as SaPI inducers

The trimeric Duts encoded by the *S. aureus* phages play an important role in the HGT of SaPI-encoded virulence factors between bacterial host cells. As SaPI-inducers, these phage Duts activate the SaPI ERP cycle that results in SaPI transfer to a new bacterium²⁸⁰. The *S. aureus* trimeric Duts encode a motif VI domain in the central region, which is required for $\phi 80\alpha$ Dut induction of SaPI_{bov1}³⁰⁵. Given that induction of the SaPI cycle is detrimental to the phage, the maintenance of motif VI across the *S. aureus* phages implies a regulatory role for this region that necessitates its presence within the Dut.

1.4.4.1 The structure of the phage-encoded trimeric Duts

The trimeric *S. aureus* phage Duts encode the five conserved domains present in all the trimeric Duts (motifs I-V) and present a prototypical homotrimeric Dut organisation (Figure 1.15). As such, they have three independent active sites and all three subunits of the trimer contribute to each active site. Motifs I, II, and IV from one subunit bind the triphosphate moiety, motif III from the adjacent subunit is involved in nucleoside binding, and the flexible motif V from the third protomer is a P loop-like element that undergoes a conformational change to cover the bound dUTP triphosphate³⁰⁵.

In addition, the *S. aureus* trimeric Duts encode the extra motif VI (residues 99-124 in Dut ^{$\phi 80\alpha$}). The motif VI of the *S. aureus* trimeric Duts fold as an L-shaped β -hairpin that is laterally projected out from a characteristic channel formed at the trimer core interface^{305,306}. While motifs I-V are conserved, motif VI is highly divergent between the *S. aureus* trimeric Duts and is not required for enzymatic activity³⁰⁵. Given that previously identified signalling roles for the Duts in other organisms were associated with such non-enzymatic extra domains, the *S. aureus* phage Dut motif VI is a promising candidate for a regulatory region. Interestingly, this motif VI also has a role in the *S. aureus* phage Dut-mediated induction of the SaPI cycle.

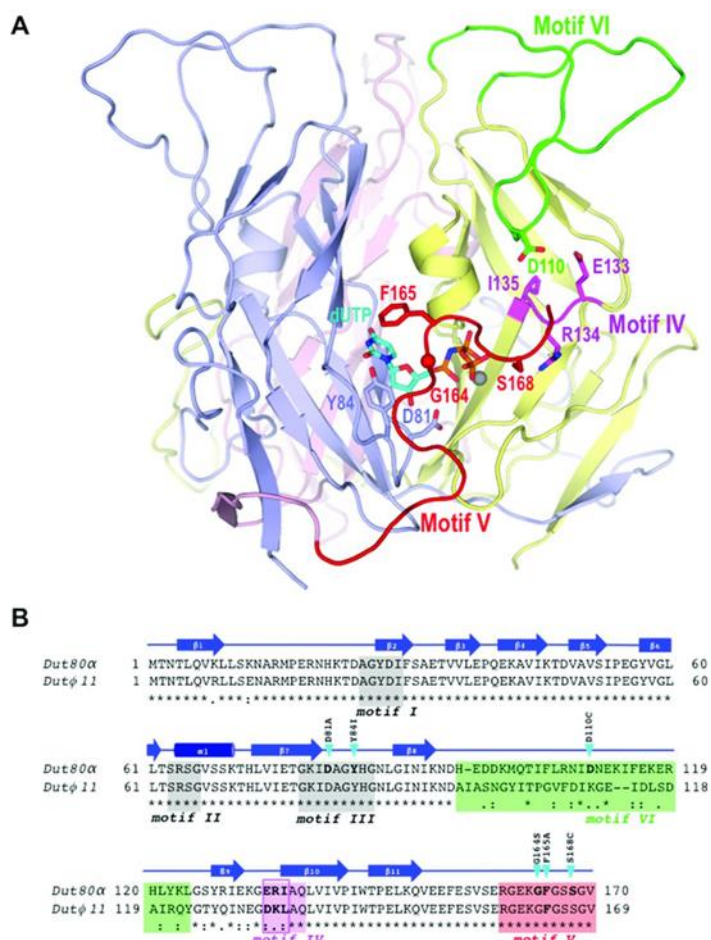


Figure 1.15 Spatial location in Dut80α^{WT} structure of relevant areas and residues for Dut–StII interaction. (A) Structure of homotrimeric (monomers in blue, yellow and pink) Dut80α^{WT} bound to α,β-imido-dUTP (dUPNPP) - Mg (in sticks with carbon atoms in cyan and Mg in grey spheres; PDB 3ZEZ) with one monomer highlighted to show the phage Dut specific motif VI (green), the C-terminal conserved motif V (red) and the exchanged residues from domain IV (magenta). Residues in which mutations are analysed in the Maiques *et al.* 2016 manuscript are shown in sticks and labelled. (B) Structure-based sequence alignment of Dut80α and Dutφ11. The (*) symbol indicates the conserved residues and the (.) or (:) show conserved substitutions. The secondary structural elements are shown and labelled above the amino acid sequence in blue. The five Dut conserved motifs are shaded in grey for motif I, II and III, pink for motif IV and red for motif V, and the *Staphylococcus aureus* phage specific motif VI in green. The point mutations evaluated in the Maiques *et al.* 2016 manuscript and highlighted as sticks in Figure 1A are indicated with cyan arrows. For Dut80α^{ΔVI} and Dutφ11^{ΔVI} motif VI deletional mutants, the highlighted region in green was substituted by Ser-Asn. For Dut80α^{ΔV} and Dutφ11^{ΔV} C-terminal motif V mutants, a stop codon was introduced at the beginning of motif V (S158* and S157* for Dut80α ΔV and Dutφ11ΔV, respectively). Motif IV residues exchanged between Dut80α (Glu133-Arg134-Ile135) and Dutφ11 (Asp132-Lys133-Leu134) are highlighted with a pink square. Reproduced from Maiques *et al.* 2016, under a Creative Commons license³⁰⁴.

1.4.4.2 The role of the trimeric phage Dnt variable motif VI for SaPI-induction

The importance of the *S. aureus* phage Dut motif VI in SaPI-induction has been shown in studies using the $\phi 80\alpha$ Dut^{304,305}. Strategies for analysing the role of motif VI in SaPI-induction included analysis of a similar *S. epidermidis* phage encoded Dut that lacked this extra domain, which was unable to induce the

island. Likewise, mutations deleting the motif VI from the Dut^{φ80α} removed the SaPI-inducing ability from the Dut. However, although Dut^{φ80α} motif VI is important for SaPI-induction, its expression alone is not sufficient to mediate SaPI induction^{304,305}.

There were discrepancies between the results found for Dut^{φ80α} and those found using the Dut from SaPI-inducing phage φ11. Research using the Dut^{φ11} found only slight differences in the SaPI induction between the wt and motif VI mutant^{304,306,368}. However, when the affinities of the two Duts for the SaPI master repressor Stl were checked, Dut^{φ11} had 20 times higher affinity for the Stl than the Dut^{φ80α}³⁰⁴. As such, given that Stl binding requires more than one motif, the Dut^{φ11} could potentially overcome the loss of one interacting motif due to its higher affinity for the Stl. On the other hand, the lower Stl-affinity of the Dut^{φ80α} seems to have led to a more delicately balanced Dut:Stl interaction that cannot overcome the loss of motif VI.

1.4.4.3 The role of the flexible motif V in SaPI induction

Studies investigating the mechanism by which the trimeric *S. aureus* Duts bind the SaPI Stl found that dUTP inhibits the Dut:Stl interaction^{304,368}. It is thought that the Stl repressor and the dUTP compete against one another to bind the phage Duts. Given that the flexible motif V undergoes a conformational change following dUTP binding to cover and stabilise the interaction, the motif V was investigated as a potential second motif involved in Stl interaction.

The Dut^{φ80α} motif V deletion mutant has a decreased affinity for the SaPI Stl, but was still able to bind dUTP³⁰⁵. Other point mutations in the Dut^{φ80α} that prevented the motif V from adopting an ordered conformation had an impact on both Stl and dUTP binding³⁰⁴. With the Dut^{φ11} motif V mutant, the affinity for Stl was identical to the wt, but there was no catalytic activity^{304,368}. The differences between the two phage Duts again implied a difference in the levels of relevance of each motif to Stl binding for Dut^{φ80α} and Dut^{φ11}. Indeed, when a Dut^{φ11} motif V and VI deletional mutant was tested, it had a significantly reduced affinity for the SaPI Stl³⁰⁴. This indicated that both motifs V and VI are important for the binding of Stl.

1.4.4.4 Mechanism for phage-encoded trimeric Dut induction of SaPI_{bov1}

The recent discovery that motif IV is also involved in the Dut:Stl interaction resolved any discrepancies between the Dut^{φ80α} and Dut^{φ11} models³⁰⁴. The motif IV is localised at the junction of motifs V and VI, explaining the synergistic involvement of all these motifs in Stl interaction and SaPI depression. This final piece of the puzzle enabled the proposal of an updated model for the trimeric *S. aureus* phage Dut interaction with the SaPI Stl (Figure 1.16)³⁰⁴.

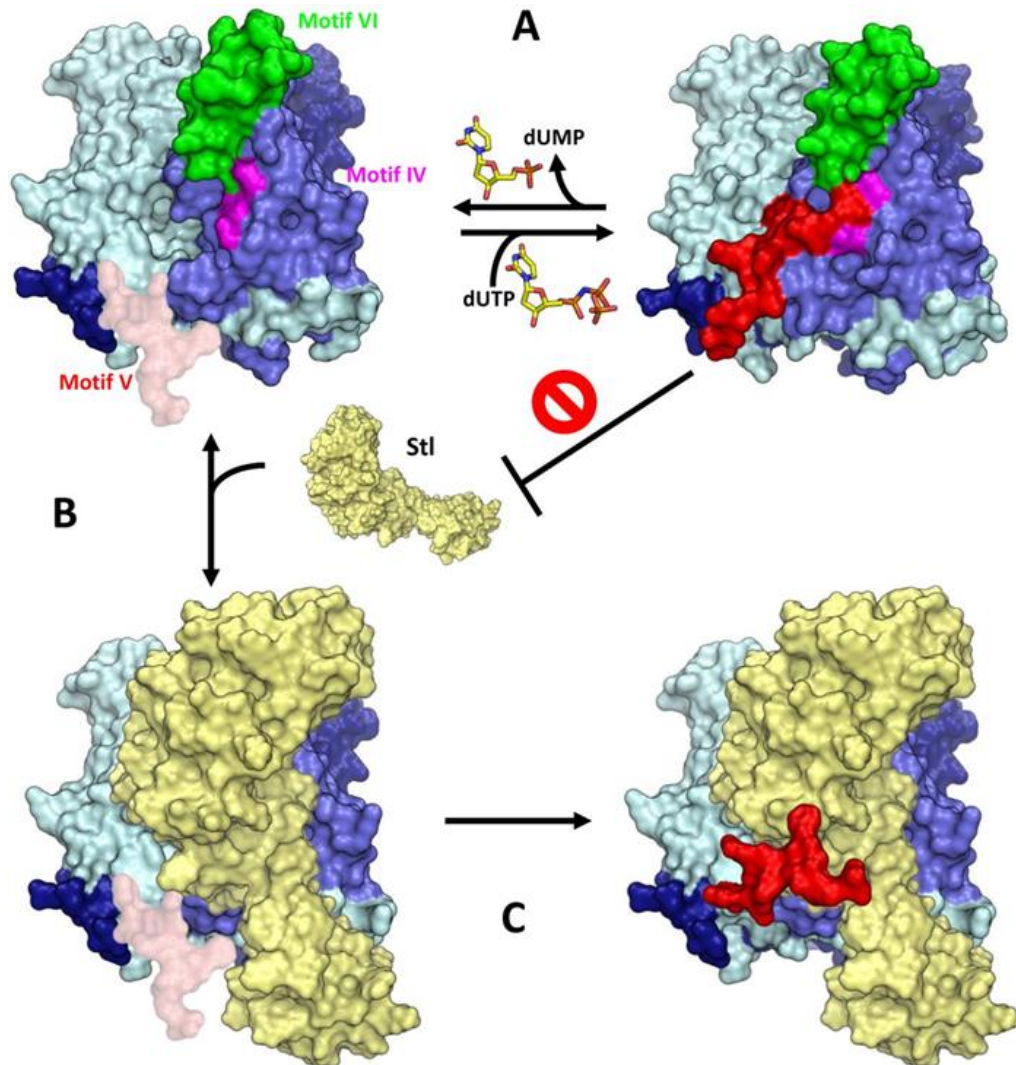


Figure 1.16 Model of *Staphylococcus aureus* phage Dut interaction with SaPIs Stl. (A) Binding and hydrolysis of dUTP by Dut implicates the folding and stabilization of the flexible conserved motif V (pink-red) over the active centre. This disposition of motif V limited the access of Stl to the anchor formed by motif IV (magenta) and motif VI (green) preventing binding. (B) In the Dut *apo* form, the Stl anchor region is accessible due to the motif V flexibility, allowing Stl to interact with residues from motifs IV and VI near the active centre. (C) Once Stl is bound, motif V would interact with Stl, stabilizing the complex. Structures are shown in surface using for Dut (different hues of blue) the experimental structures of Dut^{80α} (PDB 3ZEZ) and for Stl (yellow) an in silico model generated by I-Tasser^{369–371} from the Stl sequence. For clarity, a single Stl is displayed; two additional Stl molecules following the three-fold Dut symmetry should be present to complete the proposed 1:1 Dut–Stl interaction. Reproduced from Maiques et al. 2016, under a Creative Commons Licence³⁰⁴.

In this revised model of Dut:Stl interaction, the Stl recognises *apo* Dut by binding to the motifs IV and VI, close to the active site. Following Stl binding, the flexible motif V undergoes a conformational change to come around and interact with the Stl to stabilise the complex. This blocks the access of dUTP to the active site. Conversely, when dUTP is bound to the active site the motif V also adopts the ordered conformation to stabilise that interaction. In this way bound-dUTP blocks the Stl from binding to the Dut. This explains why the presence of dUTP inhibits Dut:Stl interaction.

This model also explains the differences seen between the Dut^{φ80α} and Dut^{φ11} models. Due to the synergistic involvement of the three motifs, different motifs can be more or less important for the different Duts. Hence, for the Dut^{φ11} motifs IV and VI generate a Stl-anchor of very high affinity, whilst the Dut^{φ80α} anchor site is of low affinity. This is supported by the fact the Dut^{φ11} displayed affinity for Stl 20 times higher than Dut^{φ80α}. As such, the involvement of motif V does not matter for Dut^{φ11} Stl-binding, but is vital for the Dut^{φ80α}. Likewise, the Dut^{φ11} can overcome the deletion of motif VI to continue binding the Stl using motifs IV and V, whilst the Dut^{φ80α} is rendered unable to bind Stl.

All the *S. aureus* phages encode Duts, although some of these are dimeric, rather than trimeric. Given that their inclusion enables SaPI parasitism, to the detriment of the phage, this could indicate an essential function for the phage, separate to the enzymatic role. Indeed, recent studies have found that *dut* deletion mutants are selected against in nature²⁸¹. It has been suggested that the Stl repressors have merged to mimic the structure of a Dut partner, which plays a role in this yet unidentified non-enzymatic function. This would be a fascinating example of molecular parasitism and raises the question, what is this mystery function of the phage-encoded Duts?

1.4.5 Phage-encoded dimeric Duts as SaPI inducers

As mentioned, all the *S. aureus* phages encode a Dut protein²⁸¹. However, some of these *dut* genes encode for a dimeric Dut protein instead of a trimeric one. Only one of the dimeric phage-encoded Duts has been tested for SaPI-inducing capability, the φNM1 Dut³⁵⁹. This Dut induced SaPI_{bov1}, raising the possibility that the dimeric Duts form a second subset of phage-encoded Dut SaPI inducers.

1.4.5.1 The structure and catalytic mechanism of the known dimeric Duts

Less is known about the dimeric Dut family across all organisms, with the larger trimeric family better characterised. The dimeric Duts form part of a larger all α -helical NTP pyrophosphohydrolase superfamily, which also includes the phage T4 dCTPase, phosphoribosyl-ATP pyrophosphatase HisE, the NTP pyrophosphatase MazG, and several uncharacterized protein families³⁷². Indeed, it has been suggested the dimeric Duts evolved from a tetrameric MazG-like ancestor by gene duplication³⁷². Unlike the trimeric Duts where numerous structures have been solved in different organisms, only four dimeric Dut structures have been resolved: *T. cruzi*, *T. brucei*, *C. jejuni*, and *L. major* Duts (Figure 1.17)³⁵⁵⁻³⁵⁸.

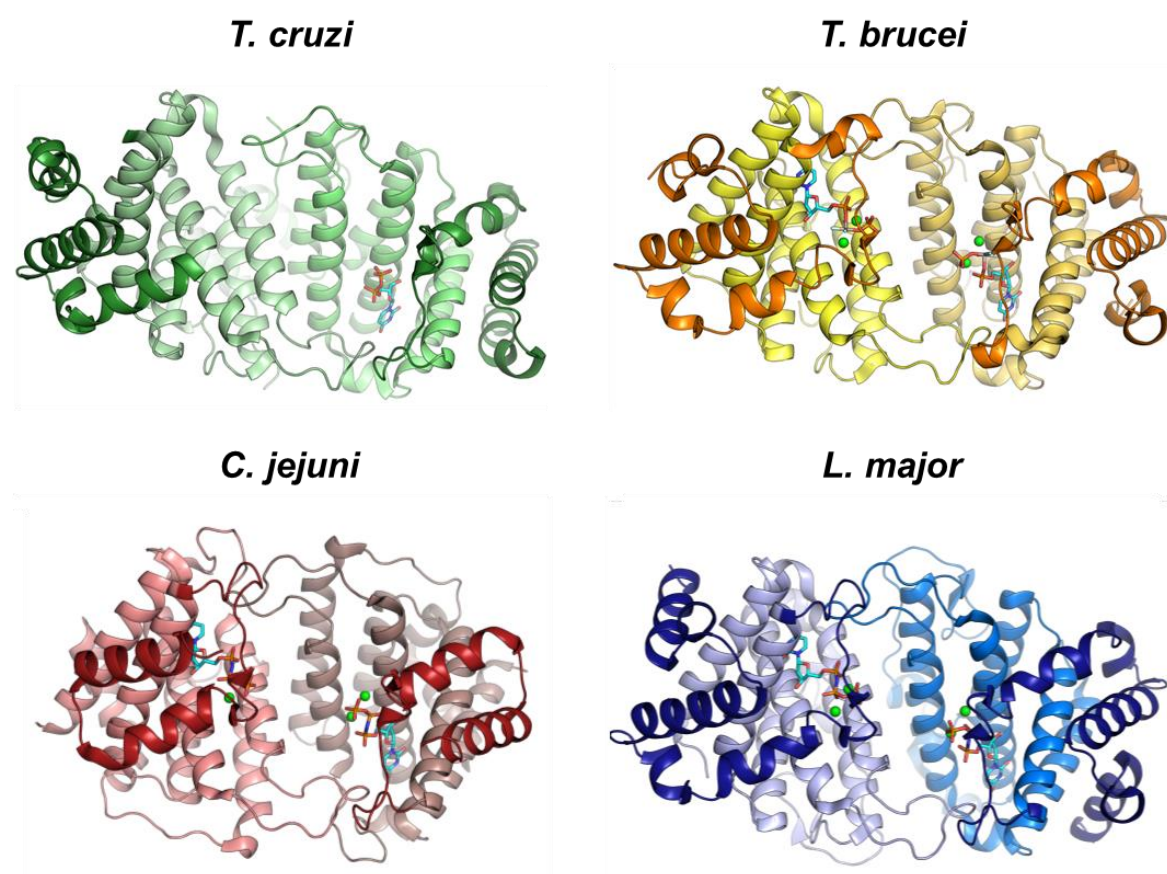


Figure 1.17 The known structures of dimeric Duts. Solved dimeric Dut structures are shown in cartoon representation and each subunit coloured in a tone of green for *Trypanosoma cruzi* (PDB 1OGK), yellow for *Trypanosoma brucei* (PDB 4DL8), red for *Campylobacter jejuni* (PDB 2CIC) and blue for *Leishmania major* (PDB 2YAY). The mobile C-terminal section is coloured in a darker tone. Reproduced from Donderis *et al.* 2017, under a Creative Commons Licence³⁷³.

The dimeric Duts are completely different in structure to the trimeric Duts, being mainly α -helical (Figure 1.17). Furthermore, they can hydrolyse dUDP as well as dUTP, whereas the trimeric Duts are inhibited by dUDP. This is because dUDP does not contain the γ -phosphate group that facilitates the ordered conformation of the trimeric motif V³⁵⁷. On the other hand, this γ -phosphate group is not important for binding and catalysis in the dimeric Duts, so dUDP can bind and be hydrolysed. The dimeric Duts are also subject to product inhibition, which is not so prevalent in the trimeric Dut types. The product inhibition seen with the dimeric Duts is caused by the high number of protein interactions between the dimeric Dut and the dUTP α -phosphate group³⁵⁷. Many of these α -phosphate group interactions remain present in the dUMP product, explaining the product inhibition seen, whilst the α -phosphate interactions are not as numerous with the trimeric Duts.

The four solved structures of the dimeric Duts all display conformational changes on binding of the ligand, open (ligand-free) or closed (ligand-bound). As such, the dimeric Dut proteins can be divided into two distinct domains: the rigid dimerization domain, and the flexible motif that undergoes these extensive conformational changes. The active site is conserved between the dimeric Duts, but the dimeric Duts differ in the number of α -helices that they encode. The *C. jejuni* Dut has a total of 11 α -helices (residues 2-22, 25-29, 38-52, 67-88, 97-106, 111-114, 125-141, 146-161, 165-185, 201-210, 215-225) and two β -strands (195-196 and 199-200)³⁵⁸. The *T. cruzi* Dut has 12 helices (residues 12-29, 33-36, 39-55, 71-93, 102-111, 153-168, 172-186, 190-205, 207-210, 223-231, 245-259, and 264-274), as does the *T. brucei*^{356,357}. However, the *T. brucei* Dut exhibits a larger conformational change following ligand binding than the *T. cruzi* Dut³⁵⁷. Lastly, the *L. major* Dut has 10 α -helices (residues 11-29, 32-35, 38-53, 70-92, 135-150, 154-168, 172-188, 218-222, 224-242, and 252-262) linked by flexible loops³⁵⁵.

The fact that humans encode a trimeric Dut with no extra domain, whilst important human pathogens encode the very different dimeric Duts, has made these dimeric Duts an attractive drug target. This is particularly true of the trypanosomatids³⁷⁴. As such, most research into the dimeric Duts has worked on elucidating the structure and enzymatic function of these Duts with a focus on

drug design. However, recent work has identified that some phages also encode dimeric Duts, including the *S. aureus* phage ϕ NM1^{358,359}.

1.4.5.2 Phage ϕ NM1 as the inducer of SaPIbov1

Initially, researchers identified that phages ϕ NM1 and ϕ NM2, isolated from the *S. aureus* Newman strain, were able to induce SaPIbov1³¹⁷. At that time, it had already been established that SaPIbov1 induction was triggered by the interaction of the SaPIbov1 Stl with the *S. aureus* phage trimeric Duts²⁸⁰. However, when the *dut* genes of the SaPIbov1-inducing phages ϕ 80 α and ϕ NM1/ ϕ NM2 were compared, they had completely different sequences³¹⁷. Indeed, the ϕ NM1 and ϕ NM2 Duts, which are identical, were homologous to the dimeric Duts³⁵⁹.

The structure of the ϕ NM1 Dut has not been solved yet, but it has been established that this Dut is sufficient for SaPIbov1 induction³⁵⁹. As with the trimeric *S. aureus* phage Duts, this induction occurs following a direct protein:protein interaction with the SaPIbov1 Stl³⁵⁹. Given the completely different structures of the dimeric and trimeric Dut families, it is unlikely that the mechanism by which the ϕ NM1 Dut binds Stl has any similarity to the trimeric Dut mechanism. However, as yet the *S. aureus* phage dimeric Dut mechanism of interaction with the SaPIbov1 Stl remains unknown. Furthermore, it is not known if the SaPI-inducing capability of this dimeric Dut is limited to ϕ NM1, or if the other *S. aureus* phage dimeric Duts also exhibit this effect. If this is the case, the phage-encoded dimeric Duts would form a novel subset of SaPI inducers.

1.5 Aims

The life cycle of the SaPIs has been extensively studied, making these pathogenicity islands the best characterised of the PICI family. The induction of the SaPIs by phage-encoded proteins forms a key point of interaction between the SaPIs and the phages. This interaction triggers the SaPI ERP cycle, making it vital for SaPI replication and transfer. Of the known SaPI Stl:phage protein interactions, the SaPIbov1 Stl interaction with the phage-encoded Duts has been most studied. However, although a model has been proposed for the interaction

of Stl_{SaPIbov1} and the trimeric phage Duts, the exact regions of interaction between these proteins is not known. Furthermore, the dimeric Dut encoded by *S. aureus* phage ϕ NM1 has been shown to bind the SaPIbov1 Stl, but little is known about how this interaction occurs. Indeed, none of the other phage-encoded dimeric Duts have been tested for the ability to induce SaPIbov1, opening the possibility that these form a second subset of SaPIbov1 inducers. Therefore, the central purpose of this thesis was to determine whether the phage-encoded dimeric Duts are a subset of SaPIbov1 inducers. If this was the case, another aim would be to discover how two such different proteins can induce the same Stl. Accordingly, the overall aims of this thesis were:

- To identify the distribution of dimeric Duts in *S. aureus* phage genomes and test their ability to induce SaPIbov1 through a direct interaction with the Stl. This would determine if the SaPIbov1 induction of ϕ NM1 was an isolated incident, or if the phage-encoded dimeric Duts form a new subset of SaPIbov1 inducers.
- To identify the structure of the dimeric Duts and utilise this knowledge to investigate the regions involved in SaPI derepression, with an aim to understanding the phage dimeric Dut mechanism of interaction with the Stl_{SaPIbov1}. Such investigations utilised knowledge of the phage-encoded trimeric Dut mechanism of interaction with the Stl_{SaPIbov1} to look for any parallelisms.
- To investigate the implications of the SaPIbov1 Stl repressors apparent ability to interact with unconserved Dut proteins performing the same phage function with regard to intra- and inter-specific transfer in staphylococci. It was hypothesised that the conserved nature of phage Dut proteins would have enabled the spread of SaPIbov1 homologues through different staphylococcal species.
- Finally, to investigate the SaPIbov1 Stl with the aim of establishing the regions involved in DNA binding for island repression, phage-encoded trimeric Duts binding, and phage-encoded dimeric Dut binding. The goal of this investigation was to try to determine whether separate domains or

residues were responsible for these multiple functions, or if the Stl exploited residue overlap in the phage Duts.

Chapter 2 Materials and methods

2.1 Materials

2.1.1 Chemicals

Unless otherwise noted, the reagents used for this thesis were obtained through major suppliers, including Fisher Scientific and Sigma-Aldrich.

2.1.2 Enzymes

Restriction endonucleases and buffers for the DNA manipulation in this study were acquired from New England Biolabs. The DNA polymerase used in PCR reactions for cloning and site-directed mutagenesis was Platinum® Taq DNA polymerase High Fidelity from Invitrogen. For colony PCR checks, DreamTaq DNA polymerase from Thermo Scientific was used. T4 DNA ligase from Thermo Scientific was used for the ligations in this study. For site-directed mutagenesis by the fill-in of 5' overhangs, DNA polymerase I, Large (Klenow) Fragment from Invitrogen was used.

2.2 Bacterial strains and vectors

2.2.1 Bacterial strains

The bacterial strains used in this study are listed in Table 2.1. Cloned constructs were transformed into and tested in *E. coli* strain DH5 α , with obtained plasmids introduced to various *S. aureus* strains by transformation²⁷¹. Protein expression was carried out in *E. coli* BL-21 (DE3) cells.

2.2.2 Storage and growth

Glycerol stocks of bacterial strains were archived at -80°C for long term storage. Single colonies were grown overnight in the appropriate selective media, with the addition of antibiotics as needed. The resultant overnight cultures were stored in 1ml aliquots with 15% glycerol.

S. aureus and other staphylococcal strains were routinely grown on tryptic soy agar (TSA) or in tryptic soy broth (TSB) liquid medium (Oxoid) at 32°C, 37°C or 43°C. Liquid cultures were grown in an orbital shaker at 80 or 120 rpm. Antibiotics (Sigma-Aldrich) were used where appropriate according to

transformed plasmids or chromosomal markers. Antibiotics and chromogenic substrates used included erythromycin ($10\ \mu\text{g ml}^{-1}$ or $2.5\ \mu\text{g ml}^{-1}$), tetracycline ($3\ \mu\text{g ml}^{-1}$), chloramphenicol ($20\ \mu\text{g ml}^{-1}$) and/or X-gal ($20\ \text{mg ml}^{-1}$, Roche).

E. coli strains were routinely grown on Luria-Bertani (LB, Sigma-Aldrich) agar plates or in LB liquid medium at 28°C or 37°C . Liquid cultures were grown in an orbital shaker at 80 or 120 rpm. Antibiotics (Sigma-Aldrich) were used where appropriate to maintain transformed plasmids, these included ampicillin ($100\ \mu\text{g ml}^{-1}$) and kanamycin ($100\ \mu\text{g ml}^{-1}$).

2.2.3 Plasmid vectors

The plasmid vectors used in this project are listed in Table 2.2. Generation of these plasmids was achieved using standard procedures including polymerase chain reaction (PCR), Klenow fill-in reactions, PCR product purification, restriction enzyme digestion, ligation, and gel extraction. These techniques are described in section 2.3. Generally, inserts were amplified by PCR, digested, and ligated into pre-digested plasmid vectors.

Vectors used for the expression of genes in *S. aureus* were assembled by cloning target genes into the pCN51 plasmid, which contains a cadmium-inducible promoter (P_{Cad})³⁷⁵. The *blaZ* transcriptional fusions in *S. aureus* were performed using vector pCN41 or using pRN8298 with the *blaZ* gene cloned downstream of the insert³⁷⁵. Protein purification was performed using the pPROEX-HTa plasmid expressing the proteins of interest in *E. coli* BL-21 DE3.

The constructed plasmids were confirmed by Sanger sequencing performed by Eurofins Genomics, using the primers listed in Appendix 1. Sequence data was verified by analysis of the sequence chr

omatogram using CLC Genomics Workbench 7 and SnapGene, the GSL Biotech LLC sequence-viewing program.

Certain plasmids were obtained by gene synthesis from GenScript HK Limited. These were manufactured as a plasmid preparation and these plasmids are indicated in Appendix 1. Nucleotide sequences of the genes to be synthesised

were supplied to the company and these sequences were optimized and supplied in GenScript standard vectors.

Table 2.1 Bacterial strains used in this study.

The dimeric Duts are a subset of phage-encoded SaPI inducers		
Strains	Description	Reference
DH5α	Host for DNA cloning	
RN4220	Restriction-defective derivate of RN450	376
BL21 (DE3)	<i>E. coli</i> expression strain	Stratagene
JP6774	RN4220 Δspa SaPIbov1 <i>tst::tetM</i>	280
JP1996	RN4220 SaPIbov1 <i>tst::tetM</i>	269
JP11634	RN4220 SaPIbov5:: <i>ermC</i> adjusted	320
JP16140	RN4220 Δspa SaPIbov5 <i>vwm::tetM</i>	This work
JP12491	RN4220 pJP674	280
JP10280	JP6774 pJP1927	This work
JP14649	JP6774 pJP1928	This work
JP11742	JP6774 pJP2040	This work
JP14670	JP6774 pJP2041	This work
JP14834	JP6774 pJP2042	This work
JP10545	JP6774 pJP2043	This work
JP15291	JP6774 pCN51	This work
JP16197	JP16140 pJP1927	This work
JP16198	JP16140 pJP1928	This work
JP16199	JP16140 pJP2040	This work
JP16200	JP16140 pJP2041	This work
JP16201	JP16140 pJP2042	This work
JP16202	JP16140 pJP2043	This work
JP16203	JP16140 pCN51	This work
JP13056	JP12491 pJP1927	This work
JP14809	JP12491 pJP1928	This work
JP13058	JP12491 pJP2040	This work
JP14808	JP12491 pJP2041	This work
JP15292	JP12491 pJP2042	This work
JP13055	JP12491 pJP2043	This work
JP15105	JP12491 pCN51	This work
JP13077	JP1996 ΦNM1 ^{WT}	This work
JP13078	JP1996 ΦNM1 ^{ΔDut}	This work
JP13210	JP11634 ΦNM1 ^{WT}	This work
JP13072	JP11634 ΦNM1 ^{ΔDut}	This work
JP12635	JP1996 ΦDI ^{WT}	This work
JP12436	JP1996 ΦDI ^{ΔDut}	This work
JP12698	JP11634 ΦDI ^{WT}	This work
JP12645	JP11634 ΦDI ^{ΔDut}	This work
JP12866	BL21 (DE3) pJP2055	This work
JP12974	BL21 (DE3) pJP2056	This work
JP14832	BL21 (DE3) pJP1932	This work

JP14833

BL21 (DE3) pJP1933

This work

Characterisation of the dimeric Duts interaction with SaPIbov1

Strains	Description	Reference
DH5α	Host for DNA cloning	
RN4220	Restriction-defective derivative of RN450	376
BL21 (DE3)	<i>E. coli</i> expression strain	Stratagene
JP6774	RN4220 Δspa SaPIbov1 <i>tst::tetM</i>	280
JP11634	RN4220 SaPIbov5:: <i>ermC</i> adjusted	320
JP12491	RN4220 pJP674	280
JP10558	BL21 DE3 pJP2046	373
JP10560	BL21 DE3 pJP2048	310
JP10555	DH5α pJP2049	373
JP11742	JP6774 pJP2040	This work
JP10545	JP6774 pJP2043	This work
JP12281	JP6774 pJP2057	This work
JP12282	JP6774 pJP2058	This work
JP12413	JP6774 pJP2059	This work
JP12278	JP6774 pJP2060	This work
JP12279	JP6774 pJP2061	This work
JP12280	JP6774 pJP2062	This work
JP12414	JP6774 pJP2063	This work
JP12391	JP6774 pJP2064	This work
JP12413	JP6774 pJP2065	This work
JP12388	JP6774 pJP2066	This work
JP12389	JP6774 pJP2067	This work
JP12390	JP6774 pJP2068	This work
JP12275	JP6774 pJP2069	This work
JP12277	JP6774 pJP2070	This work
JP12276	JP6774 pJP2044	This work
JP15291	JP6774 pCN51	This work
JP13058	JP12491 pJP2040	This work
JP13057	JP12491 pJP2044	This work
JP15105	JP12491 pCN51	This work
JP12698	JP11634 ΦDI ^{WT}	This work
JP13222	JP11634 ΦDI ^{A73L}	This work
JP12543	JP6774 pJP2071	This work
JP12052	JP6774 pJP2072	This work
JP14437	JP6774 pJP2073	This work
JP14438	JP6774 pJP2074	This work
JP14439	JP6774 pJP2075	This work
JP14440	JP6774 pJP2076	This work

JP14441	JP6774 pJP2077	This work
JP14442	JP6774 pJP2078	This work
JP14443	JP6774 pJP2079	This work
JP14444	JP6774 pJP2080	This work
JP14445	JP6774 pJP2081	This work
JP14446	JP6774 pJP2082	This work
JP15976	JP6774 pJP2083	This work
JP15917	JP6774 pJP2084	This work

Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts

Strains	Description	Reference
DH5 α	Host for DNA cloning	
RN4220	Restriction-defective derivate of RN450	376
JP12491	RN4220 pJP674	280
JP13056	JP12491 pJP1927	This work
JP14809	JP12491 pJP1928	This work
JP14835	JP12491 pJP1929	This work
JP15109	JP12491 pJP1930	This work
JP15105	JP12491 pCN51	This work
JP15110	RN4220 pJP1924	This work
JP15086	JP15110 pJP1927	This work
JP15085	JP15110 pJP1928	This work
JP15090	JP15110 pJP1929	This work
JP15089	JP15110 pJP1930	This work
JP15106	JP15110 pCN51	This work
JP15091	RN4220 pJP1925	This work
JP15093	JP15091 pJP1927	This work
JP15092	JP15091 pJP1928	This work
JP15095	JP15091 pJP1929	This work
JP15097	JP15091 pJP1930	This work
JP15107	JP15091 pCN51	This work
JP15098	RN4220 pJP1926	This work
JP15100	JP15098 pJP1927	This work
JP15099	JP15098 pJP1928	This work
JP15102	JP15098 pJP1929	This work
JP15104	JP15098 pJP1930	This work
JP15108	JP15098 pCN51	This work

Characterisation of the SaPI _{bov1} StI repressor		
Strains	Description	Reference
DH5 α	Host for DNA cloning	
RN4220	Restriction-defective derivative of RN450	376
JP12666	RN4220 ϕ NM1	This work
RN10616	RN4220 80 α	286
JP8487	RN4220 ϕ 11	This work
JP14101	JP12666 pJP2085	This work
JP14713	JP12666 pJP2086	This work
JP14714	JP12666 pJP2087	This work
JP14715	JP12666 pJP2088	This work
JP14782	JP12666 pJP2089	This work
JP14783	JP12666 pJP2090	This work
JP14716	JP12666 pJP2091	This work
JP14717	JP12666 pJP2092	This work
JP15789	JP12666 pJP2093	This work
JP15792	JP12666 pJP2094	This work
JP15716	JP12666 pJP2095	This work
JP14784	JP12666 pJP2096	This work
JP14718	JP12666 pJP2097	This work
JP14669	JP12666 pJP2098	This work
JP15795	JP12666 pJP2099	This work
JP14668	JP12666 pJP2100	This work
JP14661	JP12666 pJP2101	This work
JP14662	JP12666 pJP2102	This work
JP14725	JP12666 pJP2103	This work
JP15719	JP12666 pJP2104	This work
JP13910	JP12666 pJP2105	This work
JP13907	JP12666 pJP2106	This work
JP14106	JP12666 pJP2107	This work
JP14103	RN10616 pJP2085	This work
JP14719	RN10616 pJP2086	This work
JP14720	RN10616 pJP2087	This work
JP14721	RN10616 pJP2088	This work
JP14734	RN10616 pJP2089	This work
JP14735	RN10616 pJP2090	This work
JP14722	RN10616 pJP2091	This work
JP14723	RN10616 pJP2092	This work
JP15788	RN10616 pJP2093	This work
JP15791	RN10616 pJP2094	This work
JP15715	RN10616 pJP2095	This work
JP14736	RN10616 pJP2096	This work
JP14724	RN10616 pJP2097	This work

JP14667	RN10616 pJP2098	This work
JP15794	RN10616 pJP2099	This work
JP14666	RN10616 pJP2100	This work
JP14659	RN10616 pJP2101	This work
JP14660	RN10616 pJP2102	This work
JP14775	RN10616 pJP2103	This work
JP15718	RN10616 pJP2104	This work
JP13911	RN10616 pJP2105	This work
JP13908	RN10616 pJP2106	This work
JP13905	RN10616 pJP2107	This work
JP14423	JP8487 pJP2085	This work
JP14709	JP8487 pJP2086	This work
JP14727	JP8487 pJP2087	This work
JP14710	JP8487 pJP2088	This work
JP14728	JP8487 pJP2089	This work
JP14729	JP8487 pJP2090	This work
JP14730	JP8487 pJP2091	This work
JP14711	JP8487 pJP2092	This work
JP15787	JP8487 pJP2093	This work
JP15790	JP8487 pJP2094	This work
JP15714	JP8487 pJP2095	This work
JP14731	JP8487 pJP2096	This work
JP14712	JP8487 pJP2097	This work
JP14732	JP8487 pJP2098	This work
JP15793	JP8487 pJP2099	This work
JP14785	JP8487 pJP2100	This work
JP14786	JP8487 pJP2101	This work
JP14787	JP8487 pJP2102	This work
JP14733	JP8487 pJP2103	This work
JP15717	JP8487 pJP2104	This work

Table 2.2 Plasmids used in this study.

The dimeric Duts are a subset of phage-encoded SaPI inducers		
Plasmid	Description	Reference
pPROEX-HTa	Expression vector	Invitrogen
pCN51	Expression vector	375
pJP674	pRN8298-Cm-Pint-stl-str-xis fused to blaZ (SaPIbov1)	280
pJP1927	pCN51 3xflag Φ NM1 Dut ^{WT}	This work
pJP1928	pCN51 3xflag Φ O11 Dut ^{WT}	This work
pJP2040	pCN51 3xflag Φ DI Dut ^{WT}	This work
pJP2041	pCN51 3xflag Φ DII Dut ^{WT}	This work
pJP2042	pCN51 3xflag Φ O46 Dut ^{WT}	This work

pJP2043	pCN51 3xflag $\Phi 55$ Dut ^{WT}	This work
pJP2055	pPROEX-HTa His-Stl _{SaPIbov1} , <i>E.coli</i> RBS - Dut Φ NM1	This work
pJP2056	pPROEX-HTa Stl _{SaPIbov1} , <i>E.coli</i> RBS - Dut Φ NM1	This work
pJP1932	pPROEX-HTa His-Stl _{SaPIbov1} , <i>E.coli</i> RBS - Dut Φ O11	This work
pJP1933	pPROEX-HTa Stl _{SaPIbov1} , <i>E.coli</i> RBS - Dut Φ O11	This work

Characterisation of the dimeric Duts interaction with SaPIbov1

Plasmid	Description	Reference
pPROEX-HTa	Expression vector	Invitrogen
pCN41	Expression vector	375
pCN51	Expression vector	375
pJP674	pRN8298-chlor- <i>pint</i> -20-19-18 <i>bla</i> Z (SaPIbov1)	280
pJP2040	pCN51 3xflag Φ DI Dut ^{WT}	This work
pJP2043	pCN51 3xflag $\Phi 55$ Dut ^{WT}	This work
pJP2044	pCN51 3xflag Φ DI Dut ^{A73L}	This work
pJP2046	pET-28a Φ DI Dut ^{WT}	This work
pJP2048	pET-28a Φ O11 Dut ^{WT}	This work
pJP2049	pET-28a $\Phi 55$ Dut ^{WT}	This work
pJP2057	pCN51 3xflag Φ DI Dut ^{$\Phi 55\alpha 4$}	This work
pJP2058	pCN51 3xflag Φ DI Dut ^{$\Phi 55\alpha 4-\alpha 6$}	This work
pJP2059	pCN51 3xflag Φ DI Dut ^{$\Phi 55\alpha 4-\alpha 7$}	This work
pJP2060	pCN51 3xflag Φ DI Dut ^{$\Phi 55\alpha 4-\alpha 5$}	This work
pJP2061	pCN51 3xflag Φ DI Dut ^{$\Phi 55\alpha 5-\alpha 6$}	This work
pJP2062	pCN51 3xflag Φ DI Dut ^{$\Phi 55\alpha 7$}	This work
pJP2063	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI$\alpha 4$}	This work
pJP2064	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI$\alpha 4-\alpha 6$}	This work
pJP2065	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI$\alpha 4-\alpha 7$}	This work
pJP2066	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI$\alpha 4-\alpha 5$}	This work
pJP2067	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI$\alpha 5-\alpha 6$}	This work
pJP2068	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI$\alpha 7$}	This work
pJP2069	pCN51 3xflag Φ DI Dut ^{Q17M}	This work
pJP2070	pCN51 3xflag Φ DI Dut ^{K152*}	This work
pJP2071	pCN51 3xflag Φ DI Dut ^{$\Phi 55$I24-Q82}	This work
pJP2072	pCN51 3xflag $\Phi 55$ Dut ^{ΦDII24-Q82}	This work
pJP2073	pCN51 3xflag Φ DI Dut ^{F41C}	This work
pJP2074	pCN51 3xflag Φ DI Dut ^{F41D}	This work
pJP2075	pCN51 3xflag Φ DI Dut ^{F41A}	This work
pJP2076	pCN51 3xflag Φ DI Dut ^{F41C/A73L}	This work
pJP2077	pCN51 3xflag Φ DI Dut ^{F41D/A73L}	This work
pJP2078	pCN51 3xflag Φ DI Dut ^{F41A/A73L}	This work
pJP2079	pCN51 3xflag Φ DII Dut ^{WT}	This work
pJP2080	pCN51 3xflag Φ DII Dut ^{F41C}	This work

pJP2081	pCN51 3xflag $\Phi 55$ Dut ^{V41C}	This work
pJP2082	pCN51 3xflag $\Phi 55$ Dut ^{V41A}	This work
pJP2083	pCN51 3xflag ΦDI Dut ^{$\Phi 55$ Dimerization Interface $\alpha 2 + \alpha 6$}	This work
pJP2084	pCN51 3xflag $\Phi 55$ Dut ^{ΦDI Dimerization Interface $\alpha 2 + \alpha 6$}	This work

Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts

Plasmid	Description	Reference
pCN51	Expression vector	375
pJP674	pRN8298-chlor- <i>Pint</i> -20-19-18blaZ (SaPIbov1)	280
pJP1924	pRN8298-chlor- <i>Pint</i> -20-19-18blaZ (SaPIC0673)	This work
pJP1925	pRN8298-chlor- <i>Pint</i> -20-19-18blaZ (ShoCI794_SEPI)	This work
pJP1926	pRN8298-chlor- <i>Pint</i> -20-19-18blaZ (ShaCI51-48)	This work
pJP1927	pCN51 3xflag $\Phi NM1$ Dut ^{WT}	This work
pJP1928	pCN51 3xflag $\Phi O11$ Dut ^{WT}	This work
pJP1929	pCN51 3xflag $\Phi IPLA6$ Dut ^{WT}	This work
pJP1930	pCN51 3xflag $\Phi 11$ Dut ^{WT}	This work

Characterisation of the SaPIbov1 StI repressor

Plasmid	Description	Reference
pCN41	Expression vector	375
pJP2085	pCN41- <i>Pint</i> -20-19-18blaZ (SaPIbov1)	This work
pJP2086	pCN41- <i>Pint</i> -20 ^{R22A} -19-18blaZ (SaPIbov1)	This work
pJP2087	pCN41- <i>Pint</i> -20 ^{Q29A} -19-18blaZ (SaPIbov1)	This work
pJP2088	pCN41- <i>Pint</i> -20 ^{F38A} -19-18blaZ (SaPIbov1)	This work
pJP2089	pCN41- <i>Pint</i> -20 ^{S39A} -19-18blaZ (SaPIbov1)	This work
pJP2090	pCN41- <i>Pint</i> -20 ^{E47A} -19-18blaZ (SaPIbov1)	This work
pJP2091	pCN41- <i>Pint</i> -20 ^{N48A} -19-18blaZ (SaPIbov1)	This work
pJP2092	pCN41- <i>Pint</i> -20 ^{N56A} -19-18blaZ (SaPIbov1)	This work
pJP2093	pCN41- <i>Pint</i> -20 ^{G15A} -19-18blaZ (SaPIbov1)	This work
pJP2094	pCN41- <i>Pint</i> -20 ^{N45A} -19-18blaZ (SaPIbov1)	This work
pJP2095	pCN41- <i>Pint</i> -20 ^{N50A} -19-18blaZ (SaPIbov1)	This work
pJP2096	pCN41- <i>Pint</i> -20 ^{Y70A} -19-18blaZ (SaPIbov1)	This work
pJP2097	pCN41- <i>Pint</i> -20 ^{R74A} -19-18blaZ (SaPIbov1)	This work
pJP2098	pCN41- <i>Pint</i> -20 ^{D77A} -19-18blaZ (SaPIbov1)	This work
pJP2099	pCN41- <i>Pint</i> -20 ^{H73C} -19-18blaZ (SaPIbov1)	This work
pJP2100	pCN41- <i>Pint</i> -20 ^{Y113A} -19-18blaZ (SaPIbov1)	This work
pJP2101	pCN41- <i>Pint</i> -20 ^{Y116A} -19-18blaZ (SaPIbov1)	This work
pJP2102	pCN41- <i>Pint</i> -20 ^{D117A} -19-18blaZ (SaPIbov1)	This work
pJP2103	pCN41- <i>Pint</i> -20 ^{D155A} -19-18blaZ (SaPIbov1)	This work
pJP2104	pCN41- <i>Pint</i> -20 ^{Y112A} -19-18blaZ (SaPIbov1)	This work
pJP2105	pCN41- <i>Pint</i> -20 ^{Y106A} -19-18blaZ (SaPIbov1)	This work

pJP2106	pCN41- <i>Pint</i> -20 ^{YY112,113AA} -19-18blaZ (SaPIbov1)	This work
pJP2107	pCN41- <i>Pint</i> -20 ^{112YYSSYD→112AASSAA} -19-18blaZ (SaPIbov1)	This work

2.3 General molecular biology techniques

General DNA manipulations were performed using standard procedures^{377,378}.

2.3.1 Chromosomal DNA extraction

Chromosomal DNA extraction of *E. coli*, *S. aureus* and other staphylococcal strains was performed using the ‘GenElute Bacterial Genomic DNA’ kit from Sigma-Aldrich. Manufacturer’s instructions were followed, except for the additional lysis of staphylococcal bacterial strains, which were incubated with lysostaphin (12.5 µg ml⁻¹, Sigma-Aldrich from *Staphylococcus saprophyticus*) at 37°C for 1 hour (h) prior to extraction.

2.3.2 Plasmid DNA extraction

Plasmid DNA extraction of *E. coli*, *S. aureus* and other staphylococcal strains was performed using the QIAprep®Spin Miniprep Kit (Qiagen). Manufacturer’s instructions were followed, except for the additional lysis of staphylococcal bacterial strains, which were incubated with lysostaphin (12.5 µg ml⁻¹, Sigma-Aldrich from *S. saprophyticus*) at 37°C for 1 hour (h) prior to extraction.

2.3.3 Polymerase chain reaction

PCR reactions were performed using the following PCR thermal cyclers; Eppendorf® Mastercycler Personal, Eppendorf® Mastercycler Nexus Gradient, Techne³Prime Personal Thermal Cycler, and Alpha Cycler 1 PCRmax. The primers used for PCRs were obtained from Thermo Fisher and are listed in Appendix 1. Lyophilized primers were re-suspended in HPLC-grade water to a final concentration of 200 µM. Working primer stock solutions of 10 µM were made with a 1:20 dilution using HPLC-grade water.

Platinum® Taq DNA polymerase High Fidelity (Invitrogen) was used to obtain PCR products for use in cloning. The following PCR mix was used: 10X High Fidelity PCR Buffer (1X final concentration), 50 mM MgSO₄ (2 mM final concentration), 10

mM dNTP Mix (0.2mM final concentration, each dNTP), 10 μ M forward primer (0.2 μ M final concentration), 10 μ M reverse primer (0.2 μ M final concentration), 5-50ng Template DNA, 5 U/ μ L Platinum® Taq DNA polymerase High Fidelity (1 U/reaction final concentration), and autoclaved, distilled water to 50 μ l. Parameters for PCR products <12kb were; initial denaturation at 94°C for 5 minutes, 35X cycles of denaturing at 94°C for 15 seconds, annealing at variable temperatures depending on primer T_m for 30 seconds, and extension at 68°C for 1 minute/kb, then a final extension at 68°C for 10 minutes. PCR products were held at 4°C before being stored at -20°C.

DreamTaq DNA polymerase (Thermo Scientific) was used for detecting generated mutants. Single colonies were selected as DNA template with 1 isogenic colony picked, smeared in an even layer, transferred to a sterile PCR tube with 40 μ l HPLC grade water and vortexed. The following PCR mix was used: 10X DreamTaq Buffer (1X final concentration, with $MgCl_2$ at 2mM), 10 mM dNTP Mix (0.2mM final concentration, each dNTP), 10 μ M forward primer (0.2 μ M final concentration), 10 μ M reverse primer (0.2 μ M final concentration), Template DNA, 5 U/ μ L DreamTaq DNA polymerase (5 U/reaction final concentration), and autoclaved, distilled water to 25 μ l. Parameters for PCR products <2kb were; initial denaturation at 95°C for 5 minutes, 35X cycles of denaturing at 95°C for 30 seconds, annealing at variable temperatures depending on primer T_m for 30 seconds, and extension at 72°C for 1 minute/kb, then a final extension at 72°C for 10 minutes. PCR products were held at 4°C before being stored at -20°C.

All PCR reaction master mixes were scaled up as required, with an additional 10% to correct for pipetting error.

2.3.4 Klenow Fill-in reactions

Klenow Fill-in reactions for cloning were performed using the large fragment of DNA polymerase I as per manufacturer's instructions (Invitrogen). The purified digested DNA (0.5 - 1 μ g) was added to a 1.5 ml microcentrifuge tube on ice, together with 10X REact® 2 Buffer (final concentration 1X), 0.5 mM dNTPs (1 μ l of each), 0.5 U/ μ l Large fragment of DNA Polymerase I (diluted with Klenow Dilution Buffer, final concentration 0.5 U), and made up to 30 μ l with autoclaved, distilled water.

2.3.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate genomic and plasmid DNA. PCR reactions and analytical restriction digests of clones were verified using different percentages of agarose (0.6, 0.7 and 1%, Invitrogen). Gels were prepared with 1X Tris-acetate-EDTA buffer and SybrSafe® (0.1 µg/ml; Thermo Fisher).

The appropriate volume of 5X loading dye (50% v/v glycerol, 50% v/v 1X Tris-acetate-EDTA, bromophenol blue added to the desired colour) was added to samples prior to loading to give a 1X final concentration. Gels were generally run at 120 volts (V) for ~40 minutes, or until the ladder bands were separated and easily visualised. For smaller or larger sample sizes the voltage and time of runs was adjusted to give optimum separation. A 1 kb+ size marker (Invitrogen) was used as a ladder to determine sample band sizes. DNA visualisation was achieved using a uviPRO gold Uvitec transilluminator (Uvitec).

Gel images were edited using PowerPoint (Microsoft) in a standardised manner. Gels were converted to grey scale without any additional manipulation.

2.3.6 Restriction digest

Restriction endonucleases and enzyme buffers for the DNA manipulation in this study were acquired from New England Biolabs and were High Fidelity (HF®) when possible. Restriction digests were performed in a volume of 35 µl, or multiples thereof, according to manufacturer's instructions. For preparative and analytical digests, samples were prepared as follows: 1-3 µg of plasmid DNA or PCR product was digested with a 1X concentration of CutSmart® buffer and 10 units of the appropriate restriction enzyme(s). Analytical digests were incubated at 37°C (or enzyme appropriate temperature) for 3 hours and preparative digests at 37°C overnight. Preparative digests were purified after digestion.

2.3.7 Gel extraction and PCR purification

Digested DNA fragments for cloning were purified before ligation. Where DNA fragments were obtained by preparative plasmid digest, the correct bands were gel extracted using the QIAquick® Gel Extraction Kit (Qiagen) as per the

manufacturer's instructions. PCR products were purified directly from the PCR reaction mix using the QIAquick® PCR Purification Kit (Qiagen), following manufacturer's instructions. DNA concentrations were established using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2.3.8 Ligation reactions

Sticky-end ligation reactions were performed using 50-150 ng of digested plasmid backbone with insert added at a 1:3 molar ratio (plasmid:insert). Ligations additionally contained 10X DNA ligase buffer (1X final concentration) and 1 Weiss U/ μ l T4 DNA ligase (1 U in total, Thermo Scientific), made up to a total volume of 20 μ l with autoclaved, distilled water. Reactions were incubated at 22°C for 1 h, before inactivation of the enzyme at 65°C for 10 min.

Blunt-end ligations were performed using 40-300 ng of plasmids prepared using the Klenow fill-in reaction. Ligations additionally contained 10X DNA ligase buffer (1X final concentration), 50% PEG 4000 solution (total concentration 5%) and 1 Weiss U/ μ l T4 DNA ligase, LC (5 U in total, Thermo Scientific), made up to a total volume of 20 μ l with autoclaved, distilled water. Reactions were incubated at 22°C for 1 h, before inactivation of the enzyme at 65°C for 10 min.

2.3.9 Precipitation of genomic DNA

Subject DNA was precipitated using the following protocol: Sodium acetate 3M, pH=5.2 was added to the DNA at a 1/10 ratio (NaAc:DNA), two volumes of chilled 100% ethanol were then added, the mix was inverted several times and incubated at -20°C for 2 h or -80°C for 30 min. The precipitated DNA was pelleted by centrifugation at 4°C with 14,000 rpm for 20 min. The resultant supernatant was discarded and the remaining pellets washed with 1 ml of chilled 70% ethanol for the removal of excess salts. The mix was pelleted by centrifugation at 4°C with 14,000 rpm for 5 min, the supernatant was discarded and the precipitated DNA was dried in an Eppendorf Concentrator 5301 at 45°C for 15 min. The precipitated DNA was re-suspended in 20 μ l of HPLC-grade water.

2.3.10 Overexpression of pPROEX-HTa vectors

Overexpression of different proteins using the pPROEX-HTa vector was performed as previously described³⁷⁹. Overnight cultures of BL21 DE3 containing the appropriate plasmids were used for test expressions in which 1:100 dilutions were grown to 0.4-0.6 OD₆₀₀, 1 ml of culture sampled and pelleted, and the remaining culture induced with varying concentrations of Isopropyl β -D-thiogalactoside (IPTG, 0.3 mM, 1mM, 10 mM, Melford Biolaboratories Ltd.). The cultures were sampled and pelleted following incubation at various temperatures and times; overnight at 25°C, for 6 h at 28°C, or for 3 hours at 37°C. Pellets were kept at -20°C before being re-suspended in 75 μ l of distilled, autoclaved water and 25 μ l of 2X Laemmli sample buffer (Bio-Rad, 2-mercaptoethanol added). Samples were heated at 95°C for 10 min, put on ice for 5 min and centrifuged for 10 mins. The un-induced and differentially induced samples were run on an SDS-PAGE gel, before being stained in Coomassie Brilliant Blue (50% Methanol, 10% glacial acetic acid, and 2.5 g Coomassie Brilliant Blue R-250, Bio-Rad) and de-stained (40% methanol, 10% Glacial acetic acid) until protein expression bands were clear.

For protein over-expression, the most efficient inducing conditions from the test expression were used and the expression culture was up-scaled to 100 ml - 1 L, depending on the down-stream applications. Following induction of the culture with IPTG and incubation at the appropriate temperature for the appropriate time, the cultures were pelleted in a Falcon™ 50 mL conical centrifuge tube and stored at -80°C until ready for protein purification.

2.3.11 Southern blot

Samples were taken at 0, 60, and 90 min following phage/SaPI induction with mitomycin C (Sigma-Aldrich, from *Streptomyces caespitosus*), or at 3 h following plasmid induction with 1-5 μ M cadmium (CdCl₂, Cadmium chloride hemi(pentahydrate), Sigma-Aldrich). Samples were pelleted and frozen at -20°C until all samples were obtained. The samples were re-suspended in 50 μ l lysis buffer (47.5 μ l TES-Sucrose and 2.5 μ l lysostaphin at 12.5 μ g ml⁻¹, Sigma-Aldrich from *S. staphylolyticus*) and incubated at 37°C for 1 hour. 55 μ l of SDS 2% proteinase K buffer (47.25 μ l H₂O, 5.25 μ l SDS 20%, 2.5 μ l proteinase K at 12.5 μ g

ml⁻¹, Sigma-Aldrich from *Tritirachium album*) was added before incubation at 55°C for 30 minutes. Samples were vortexed for at least 20 minutes with 11 µl of 10X loading dye. Cycles of incubation in dry ice and ethanol, then at 65°C were performed. Samples were run on 0.7% agarose gel at 25V overnight. DNA was transferred to a nylon membrane (0.45 mm hybond-N pore diameter, Amersham Life Science) and exposed using a DIG-labelled probe (Digoxigenin-11-dUTP alkali-labile, Roche) and anti-DIG antibody (Anti-Digoxigenin-AP Fab fragments, Roche) as per the suppliers protocol, before washing and visualisation^{377,380}. The primers used to obtain the labelled probes are shown in Appendix 1.

2.3.12 Western blot

Samples were taken at 0, 60, and 90 min following phage/SaPI induction with mitomycin C (Sigma-Aldrich, from *S. caespitosus*), or at 3 h following plasmid induction with 1-5 µM cadmium (CdCl₂, Cadmium chloride hemi(pentahydrate), Sigma-Aldrich). Preparation of *S. aureus* samples for running/transfer was performed by re-suspending pellets in 200 µl digestion/lysis buffer (50 mM Tris-HCl, 20 mM MgCl₂, 30% w/v raffinose) plus 1 µl of lysostaphin (12.5 µg ml⁻¹), mixed briefly, and incubated at 37°C for 1 h. 2X Laemmli sample buffer (Bio-Rad, 2-mercaptoethanol added) was added to the samples, which were heated at 95°C for 10 min, put on ice for 5 min and fast touch centrifuged. Samples were run on SDS-PAGE gels (15% Acrylamide, Bio-Rad 30% Acrylamide/Bis Solution) before transferring to a polyvinyl difluoride transfer membrane (Thermo Scientific, 0.2 µM) using standard methods^{377,380}. Western blot assays were performed using anti-Flag antibody probes (Monoclonal ANTI-FLAG® M2-Peroxidase (HRP) antibody produced in mouse, Sigma-Aldrich) as per the protocol supplied by the manufacturer.

2.4 Bacterial transformation

2.4.1 Preparation of chemically competent *E. coli* cells

Chemically competent *E. coli* cells (both DH5α and BL21 DE3 strains) were prepared from overnight cultures grown at 37°C with 120 rpm in LB liquid broth. 100 ml of fresh LB was inoculated with 1 ml of the overnight culture and grown at 37°C with 120 rpm to 0.35 OD₆₀₀ as measured by UV-visible spectrophotometer (Thermo Scientific, Evolution 60S). The culture was split between Falcon™ 50 mL

conical centrifuge tubes and put on ice for 15 min to stop bacterial growth. The tubes were centrifuged at 4°C at 4000 rpm for 10 min, the resultant supernatant was discarded and the pelleted bacteria re-suspended in 15 ml chilled, sterile 0.1 mM CaCl₂. The bacteria were then re-pelleted using the same conditions, the supernatant discarded, and pellets re-suspended in 2 ml chilled, sterile 0.1 mM CaCl₂. Aliquots of 200 µl were frozen at -80°C.

2.4.2 Transformation of chemically competent *E. coli* cells

Transformation of plasmids into chemically competent *E. coli* cells was performed using freshly prepared competent cells or frozen aliquots thawed on ice. 10 µl of ligation mix or purified plasmid DNA were added to aliquots of 100 µl. The reaction mix was incubated on ice for 15 min, heat shocked at 42°C for 90 seconds and placed on ice for 2 min. Immediately after, 600 µl of LB liquid medium was added and samples were incubated at 37°C with 120 rpm shaking for 1 h. Transformations were then plated out on plasmid appropriate selective LB plates (100 µg/ml ampicillin or kanamycin, see Table 2.2 for resistance details). Initially, 100 µl of transformation mixture was plated out; after this the remaining volume was centrifuged at 4000 rpm for 2.5 min, 400 µl of the supernatant removed, and the pellet re-suspended and plated out. Plates were incubated at 37°C for 24 h.

2.4.3 Preparation of electrocompetent *S. aureus* cells

Electrocompetent *S. aureus* cells (derivatives of the restriction deficient RN4220 strain) were prepared from overnight cultures grown at 37°C with 120 rpm in TSB. 100 ml of fresh TSB was inoculated with 1 ml of the overnight culture and grown at 37°C with 120 rpm to 0.2 OD₅₄₀. The culture was split between Falcon™ 50 mL conical centrifuge tubes and put on ice for 15 min to stop bacterial growth. The tubes were centrifuged at 4°C at 4000 rpm for 10 min, the resultant supernatant was discarded and the pelleted bacteria re-suspended in 12.5 ml chilled, sterile 0.5 M sucrose. The bacteria were then re-pelleted using the same conditions, the supernatant discarded, and pellets re-suspended in 12.5 ml chilled, sterile 0.5 M sucrose. Again, the bacteria were then re-pelleted using the same conditions and the supernatant discarded. The pellets were then re-suspended in 1 ml chilled, sterile 0.5 M sucrose and incubated on ice for 20 min.

The cells were pelleted using the same centrifuge conditions, the supernatant discarded, and the pellet re-suspended in 1 ml chilled, sterile 0.5 M sucrose. Aliquots of 200 μ l were frozen at -80°C .

2.4.4 Transformation of electrocompetent *S. aureus* cells

Plasmids extracted from *E. coli* were transformed into *S. aureus* electrocompetent cells prepared from the restriction deficient RN4220 strain, as previously described²⁷¹. For the electroporation, 50 or 200 μ l aliquots of electrocompetent cells were added to 0.1 or 0.2 cm cuvettes, respectively (Gene Pulser® Cuvette, Bio-Rad). For 0.1 cm cuvettes 0.2-0.6 μ g of plasmid DNA was added, and for 0.2 cm cuvettes ~5 μ g of plasmid DNA was added. Electroporation was performed using the Gene Pulser Xcell™ electroporation system (Bio-Rad) on a manual setting (2.5 kV, 25 μ F, 100 Ω). Following the electroporation pulse, 600 μ l of TSB was added to the cuvette and transferred, together with the electroporated cells, to a 1.5 ml microcentrifuge tube. samples were incubated at 37°C with 120 rpm shaking for 1 h. Transformations were then plated out on selective TSB plates appropriate for the plasmid (See Table 2.2 for resistance details). Initially, 100 μ l of transformation mixture was plated out; after this the remaining volume was centrifuged at 4000 rpm for 2.5 min, 400 μ l of the supernatant removed, and the pellet re-suspended and plated out. Plates were incubated at 32 or 37°C for 24 h.

2.5 Biochemical methods

2.5.1 Enzyme assay for the quantification of β -lactamase activity in transcriptional fusion plasmids

For the β -lactamase assays, cells were sampled at 0.2-0.3 OD_{540} and then at 4 and 5 hours without induction or post-induction with 5 μM cadmium (CdCl_2 , Cadmium chloride hemi(pentahydrate), Sigma-Aldrich). Cells were kept at -20°C until sampling was complete. β -lactamase assays, using nitrocefin (BD Diagnostic systems) as substrate, were performed as described³⁸¹, using a ELx808 microplate reader (BioTek). An adjustment was made in reading time, with plates read every 20 seconds for 30 mins. The OD_{490} was plotted over time and the change in OD for the exponential phase was recorded as $\Delta\text{absorbance}/\text{hour}$. The β -lactamase units/ml were then calculated and defined as:

$$\text{Unit } \beta\text{-lactamase activity} = [(\text{slope})(V_d)/(\epsilon_m)(l)(s)]$$

Slope is the $\Delta\text{absorbance}/\text{hour}$, V is the volume of the reaction, d is the dilution factor, E_m is the millimolar extinction coefficient for the nitrocefin ($20,500 \text{ M}^{-1} \text{ cm}^{-1}$ at 486 nm), l is the path length (cm), and s is the sample amount.

2.5.2 Protein purification

Protein overexpression pellets were re-suspended in 35 ml of binding buffer (Tris-HCl 20 mM, NaCl 500 mM, Imidazole 10 mM (99+% crystalline, Acros organics), pH 7.5) supplemented with DNaseI, lysozyme and a protease inhibitor tablet (EDTA free). The re-suspended pellets were sonicated on ice for 15 cycles at 15 seconds on, 45 seconds off (Soniprep 150, MSE (UK) Ltd.). The sonicated solution was decanted into 50ml Nalgene tubes and centrifuged at 4°C at 18000 rpm for 20 min. The lysate was transferred to a new Falcon™ 50 mL conical centrifuge tube, on ice.

Protein purification was performed on the AKTApurifier plus (GE Healthcare). The nickel column was equilibrated using 3-5 columns of filtered water and then the A and B lines washed with at least 3 column volumes of binding buffer and gradient buffer (Tris-HCl 20 mM, NaCl 500 mM, Imidazole 500 mM (99+% crystalline, Acros organics), respectively. The prepared sample was run on to the column at a rate of <1 ml/min. The sample was then washed through with binding buffer, before a gradient was set at 40 ml, 100% with 2 ml fraction sizes. Once the gradient was at 100% gradient buffer (high imidazole), an additional 10 ml of buffer was collected. The peak fractions were analysed using SDS PAGE gels and sent for mass spectrometry for protein confirmation as cut-out gel bands in distilled water.

2.5.3 Mass spectrometry

The proteomic analysis referenced in this thesis was performed at the proteomics facility of SCSIE University of Valencia that belongs to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001. The protocol they use is described below.

Samples for analysis by mass spectrometry were prepared by digestion with sequencing grade trypsin (Promega) as described elsewhere³⁸². The digestion mixture was dried in a vacuum centrifuge, re-suspended in 20 μ L of 2% acetonitrile, 0.1% trifluoroacetic acid, and 1 μ L was spotted onto the matrix assisted laser desorption/ionization target plate. After the droplets were air-dried at room temperature, 1 μ L of matrix (5 mg/mL α -Cyano-4-hydroxycinnamic acid (Sigma) in 0.1% trifluoroacetic acid-acetonitrile/H₂O (1:1, v/v)) was added and allowed to air-dry at room temperature.

The resulting mixtures were analysed in a matrix assisted laser desorption/ionization time-of-flight TOF/TOF™ 5800 system (ABSciex) in positive reflectron mode (3000 shots every position). Five of the most intense precursors (according to the threshold criteria: minimum signal-to-noise: 10, minimum cluster area: 500, maximum precursor gap: 200 ppm, maximum fraction gap: 4) were selected for every position for the MSMS analysis. And, MS/MS data was acquired using the default 1kV MS/MS method. The MS and MS/MS information was sent to MASCOT via the Protein Pilot (ABSciex). The samples were spotted in the plate 150609_161117, in run 14, positions B1 (Control) and B2-B6 (Samples 1-5).

Database searches were performed on a homemade database, with specific sequences added. Searches were done with tryptic specificity, allowing one missed cleavage and a tolerance on the mass measurement of 100 ppm in MS mode and 0.8 Da for MS/MS ions. Carbamidomethylation of Cys was used as a fixed modification and oxidation of Met and deamidation of Asn and Gln as variable modifications. Proteins showing a score higher than homology to the significance threshold were identified with confidence $\geq 95\%$. The protein score in the result report is from an MS/MS search derived from the ions scores. The protein score equates to $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.

2.6 Induction of PICIs and phage

2.6.1 Phage propagation: induction

Mitomycin C (MC, Sigma-Aldrich, from *Streptomyces caespitosus*) was used to induce *S. aureus* strains containing lysogenic bacteriophages, to produce bacteriophage lysates. The addition of MC causes DNA damage, which triggers the SOS response pathway and this initiates the phage lytic cycle.

Induction of the lysogenic *S. aureus* strains was performed using overnight, stationary cultures, diluted 1:50 with fresh TSB and grown to 0.2 OD₅₄₀. Once this OD had been reached MC (2 µg ml⁻¹) was added. The induced cultures were incubated in an orbital shaker at 32°C with 80 rpm. Cell lysis usually occurred 4 h post-induction with a standard 5 ml of lysate, although for some bacteriophages an extended incubation overnight at room temperature without shaking was necessary for full lysis of the sample. When sampling for a Southern blot, 1 ml samples were taken throughout the induction incubation, before cell lysis (0, 60, 90 and 120 min). Following lysis, un-lysed bacterial cells were removed by filtration with sterile 0.2 µm filters (Minisart® single use syringe filter unit, hydrophilic and non-pyrogenic, Sartonium Stedim Biotech) and the resultant lysate was stored at 4°C.

2.6.2 Phage propagation: infection

A second technique used to obtain bacteriophage lysates was bacteriophage infection of susceptible bacteria. Following infection, bacteriophages enter the bacteria and progress through the lytic cycle, replicating and increasing the phage population until bacterial cell lysis occurs.

Infection of susceptible *S. aureus* strains was performed using overnight, stationary cultures, diluted 1:50 with fresh TSB and grown to 0.15 OD₆₅₀. Once this OD had been reached the culture was centrifuged at 4°C with 4000 rpm for 10 min and re-suspended in TSB-phage buffer (1:1, phage buffer: 1 mM NaCl, 0.05 M Tris pH 7.8, 1 mM MgSO₄, 4 mM CaCl₂). The re-suspended culture was infected with phage lysate at a ratio of 1:3 or 1:1 (phage:bacteria). Cell lysis usually occurred 4 h post-induction with a standard 5 ml of lysate. Following lysis, un-lysed bacterial cells were removed by filtration with sterile 0.2 µm

filters (Minisart® single use syringe filter unit, hydrophilic and non-pyrogenic, Sartorius Stedim Biotech) and the resultant lysate was stored at 4 °C.

2.6.3 Bacteriophage titering assay

Bacteriophage titering assays were performed in *S. aureus* using strain RN4220 and derivatives thereof as the recipients. RN4220 is restriction-deficient and highly sensitive to bacteriophage infections due to a mutation in the *sau1 hsdR* gene³⁷⁶.

The number of phage particles in a lysate was quantified by a titering assay using a stationary, overnight culture of the appropriate recipient strain, diluted 1:50 with fresh TSB and grown to 0.3-0.4 OD₅₄₀. The phage lysate was setup as serial dilutions using PB, to 10⁻⁸. 50 µl aliquots of the recipient culture were infected with 100 µl of the phage serial dilutions and incubated at room temperature for 10 min. The different dilutions of culture-phage mix were plated out on phage base agar plates (PBA; 25 g of Nutrient Broth No. 2, Oxoid; 7g agar) supplemented with CaCl₂ (final concentration 10mM). The PBA plates were left at room temperature to dry before being incubated at 37 °C for 24 h. The number of plaques formed was counted and the plaque forming units (PFU) were calculated as follows:

$$\frac{\text{PFU}}{\text{ml}} = \frac{\text{number of plaques}}{(d \times v)}$$

d = dilution of plate counted

v = volume of lysate dilution plated

2.6.4 Transduction titering assay

All the PICIs used in this project contained *tetM*, *tetA* or *ermC* antibiotic cassettes and were used as markers to select for PICI particles on selective TSA plates (10 µg/ml tetracycline or erythromycin). Transduction titering assays were also performed using the *S. aureus* strain RN4220 as the acceptor.

The number of transductants in a lysate was quantified by a titering assay using a stationary, overnight culture of the acceptor strain, diluted 1:50 with fresh

TSB and grown to 1.4 OD₅₄₀. Once appropriately grown, CaCl₂ was added to the culture (final concentration 4.4 mM). The lysate was setup as serial dilutions using PB, to 10⁻⁶. 1 ml aliquots of the acceptor culture were infected with 100 µl of the serial dilutions and incubated at 37°C for 20 min. The incubation period allows the PICIs to infect the acceptor bacteria. The different dilutions of culture-lysate mix were plated out on TSA plates containing the PICI-appropriate antibiotics and sodium citrate (17 mM) to inactivate any free phages, preventing phage infection of the acceptor^{323,383}. The TSA plates were left at room temperature to set before being incubated at 37°C for 24 h. The number of colonies formed was counted and the colony forming units (CFU) were calculated as follows:

$$\frac{\text{CFU}}{\text{ml}} = \frac{\text{number of colonies}}{(\text{d} \times \text{v})}$$

d = dilution of plate counted

v = volume of lysate dilution plated

2.7 Allelic exchange

Selective mutations of studied prophage genomes were carried out by allelic exchange, using a variety of strategies.

2.7.1 Point and Deletion mutagenesis using pMAD

Genes of interest in the *S. aureus* chromosome were targeted for allelic exchange by homologous recombination using the pMAD vector, as previously described³⁸⁴. The pMAD vector enables homologous recombination in a two-step process; the initial integration of the whole vector, and then deletion of the vector backbone.

The pMAD vector contains the thermosensitive origin of replication for Gram-positive bacteria from pE194, an erythromycin resistance cassette, and the constitutively expressed promoter region of the *S. aureus clpB* gene fused to a promoterless *bgaB* gene from *Bacillus stearothermophilus*. The *bgaB* gene encodes a thermostable β-galactosidase; expression of this enzyme driven by the

constitutive promoter allows the identification of clones carrying the plasmid on X-gal agar plates by blue-white screening³⁸⁴.

For deletion of genes from the bacterial chromosome, the flanking regions of genes to be deleted were cloned into the pMAD vector to serve as homology arms for specific targeting vectors. These flanking regions were amplified and fused together by a double PCR reaction, as shown in Figure 2.1. For the initial PCRs (PCR 1), one set of primers was used to amplify the 5' homology arm (*primers 1* and *2*), whilst another was used for amplification of the 3' homology arm (*primers 3* and *4*). Following purification of these initial PCR products, the DNA concentrations were measured by Nanodrop 1000 spectrophotometer (Thermo Scientific). The 5' end of *primer 3* contained a 20 bp overlap, homologous to the 3' end of the 5' homology arm; this allowed the fusion of the two initial PCR products by a subsequent PCR (PCR 2). For the fusion PCR reaction, equal concentrations of the two initial PCR products (10 ng) were added for fusion, and the final fused template was amplified using the external primers, *primer 1* and *4*. The final product was purified, digested with the appropriate enzymes, and cloned into digested pMAD vector as previously described.

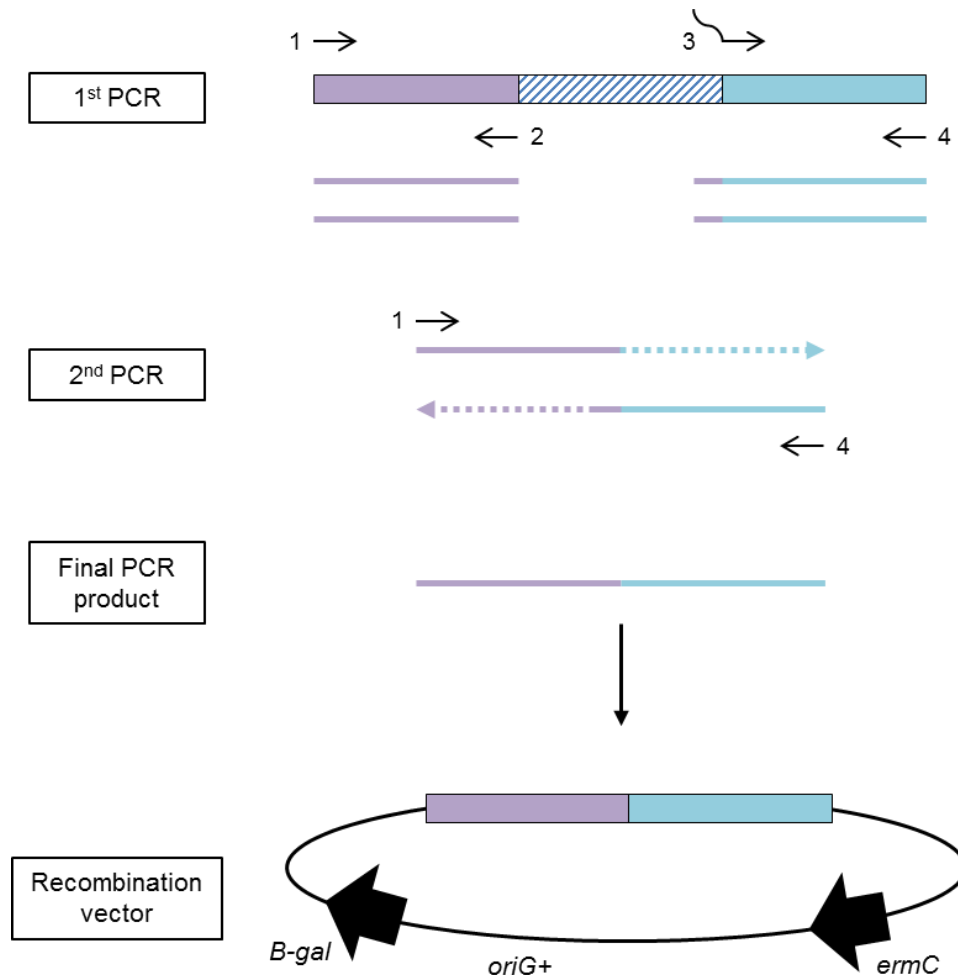


Figure 2.1 A schematic representation of the double PCR necessary for the construction of deletion mutant homology arms. The purple and blue boxes represent the flanking 5' and 3' homology arms. The hatched box illustrates the chromosomal region for deletion. Primer hybridisation sites are shown by arrows, with the *primer 3* tail indicated. The arrangement of the resultant pMAD recombination vector is represented, with the origin of replication labelled and large arrows showing the location of the erythromycin cassette and β-galactosidase gene.

For the creation of point mutations within the bacterial chromosome, a similar double PCR method was used to generate a vector with homology arms flanking the gene for mutation. These were constructed using the same PCR system, however in this case the two internal primers (*primer 2* and *3*) were designed to be complementary and to incorporate the nucleotide mutation centrally, with ~18-20 bp of homologous sequence either side (Figure 2.2). Fusion of the initial PCR products resulted in a final template with the point mutation and flanking regions, which was amplified and processed in the same way as the final deletion PCR product.

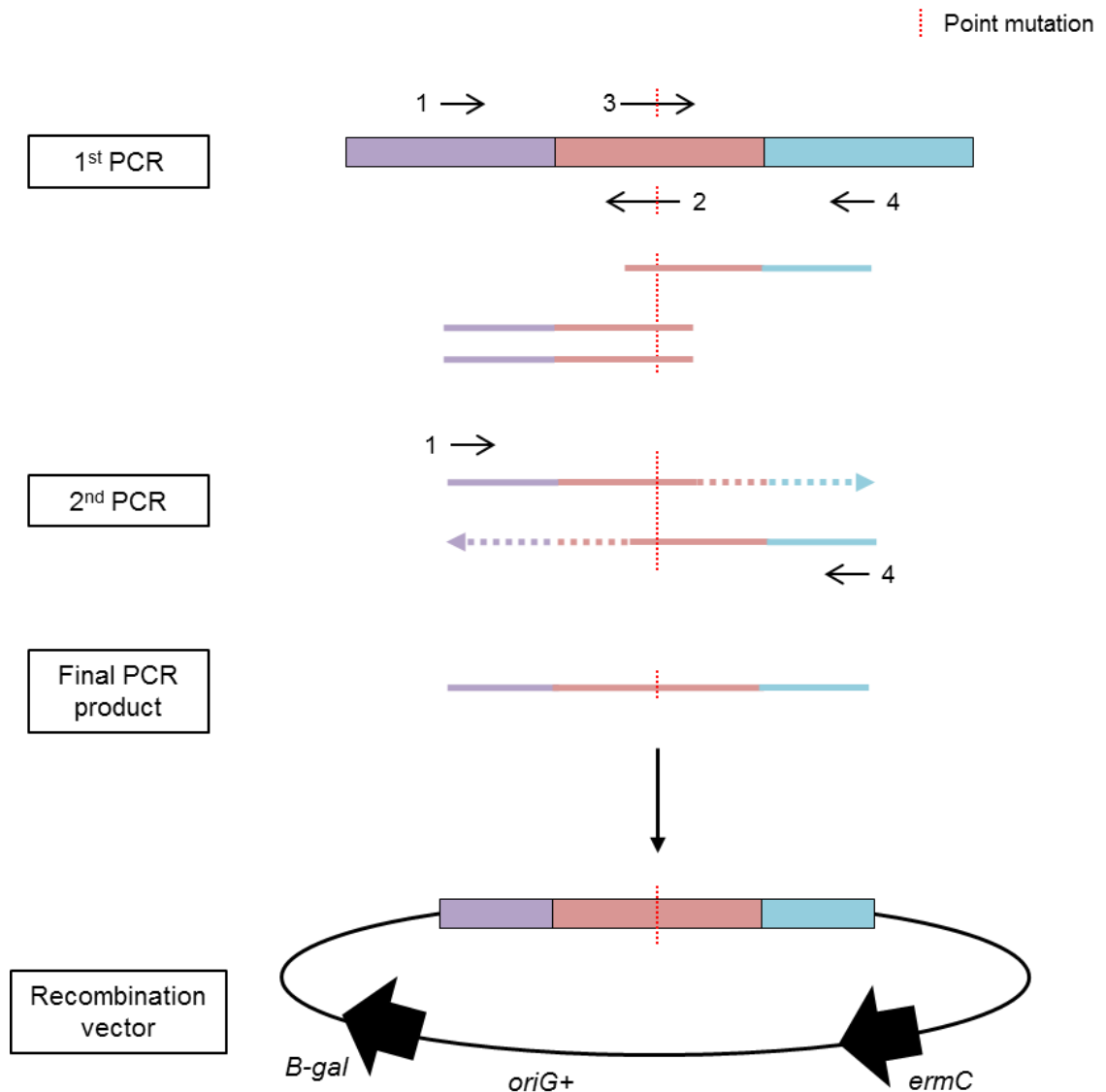


Figure 2.2 A schematic representation of the double PCR necessary for the construction of point mutation homology arms. The red box illustrates the chromosomal gene that is the target for the point mutation (target indicated by the red dashed line). The purple and green boxes represent the flanking genes. Primer hybridisation sites are shown by arrows, with the *primers* 2 and 3 complementary to one another and containing the point mutation. The arrangement of the resultant pMAD recombination vector is represented, with the origin of replication labelled and large arrows showing the location of the erythromycin cassette and β -galactosidase gene.

The cloned pMAD plasmids were confirmed in DH5 α before being transformed to the appropriate *S. aureus* strain by electroporation, as previously described. Colonies were incubated overnight at 32°C on selective TSA plates with erythromycin (2.5 μ g/ml; Erm) and X-gal (200 μ g/ml). The permissive growth temperature (32°C) allowed expression of the pMAD. Using the constitutively expressed β -gal as indicator, single blue candidate colonies were inoculated in 5 ml of TSB with Erm (2.5 μ g/ml) and grown overnight at 32°C with 120 rpm in an orbital shaker. Serial dilutions of the overnight culture (100 μ l, 10^{-2} , 10^{-3} , 10^{-4})

were plated on selective TSA plates with erythromycin (2.5 µg/ml; Erm) and X-gal (200 µg/ml), these were incubated overnight at 42°C. The antibiotic selection along with the 42°C temperature promotes the co-integration of the pMAD, triggering the initial homologous recombination of the plasmid flanking regions with the *S. aureus* chromosomal equivalents (Figure 2.3).

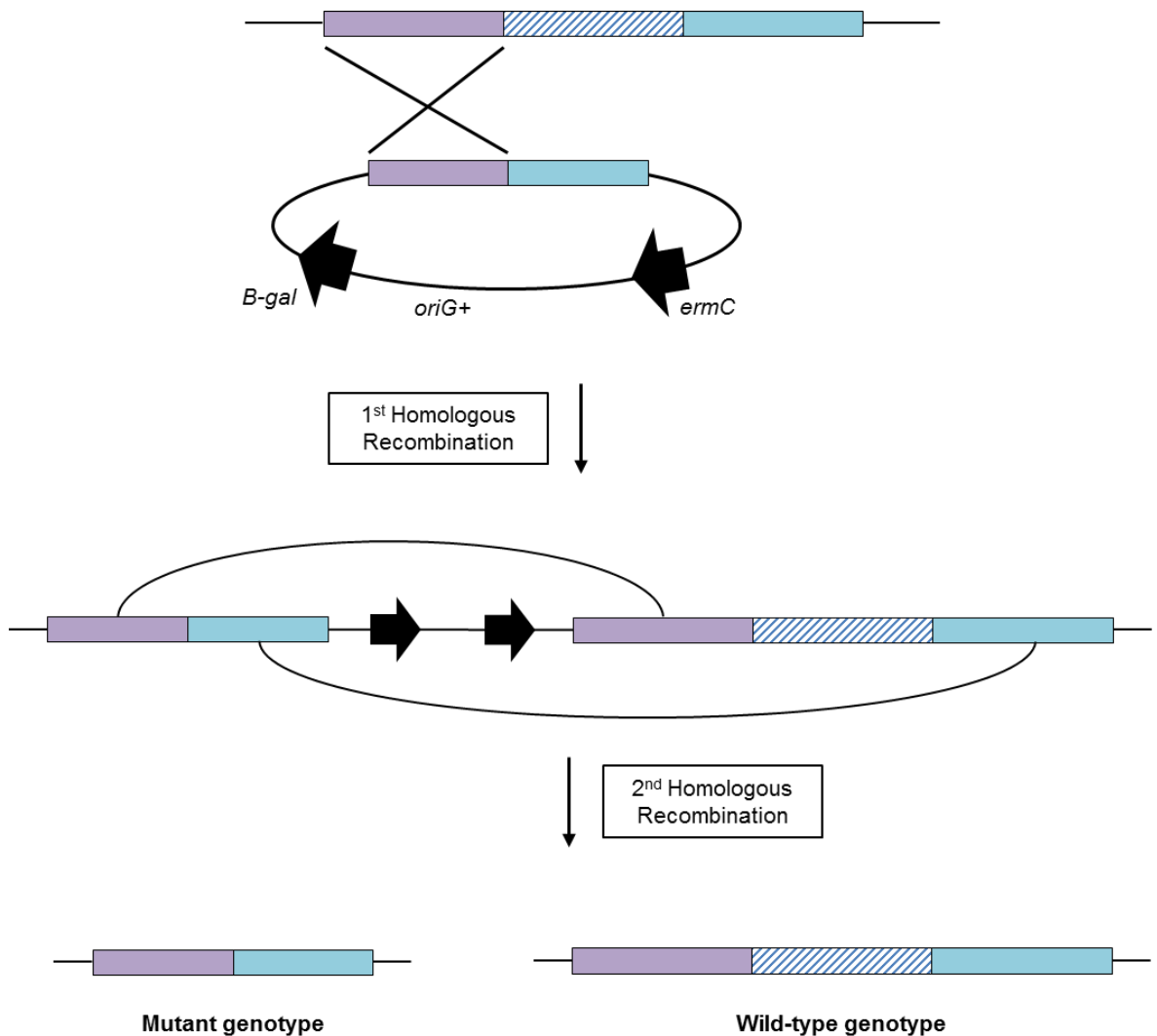


Figure 2.3 A schematic representation of pMAD mediated mutagenesis. Plasmid integration to the chromosome occurs with the first homologous recombination event. At this point, colonies are an intermediate blue (in comparison to free plasmid and no plasmid colonies) and grow in the presence of Erm. The black lines indicate the matching flanking regions from the plasmid insert and the genomic wt DNA. The second homologous recombination leads to plasmid deletion. Following this, bacteria are sensitive to Erm and grow as white colonies. Two genotypes can be produced: deletion/point mutants, or wild-type.

Colonies that were Erm-resistant and an intermediate blue colour, in comparison to the dark blue free-plasmid colonies and the white no-plasmid colonies, were selected. These were inoculated in 5 ml of TSB (no antibiotic) and incubated

overnight at 32°C with 120 rpm, to promote plasmid resolution and loss. Serial dilutions of the overnight cultures (100 µl, 10⁻⁵ and 10⁻⁶) were plated on TSA plates with X-gal (200 µg/ml) and incubated overnight at 42°C. This second recombination leads to one of two genotypes, the original wild-type is maintained or the gene deletion or mutation is introduced.

Resulting white colonies were patched on to both TSA plates and TSA plates with Erm (2.5 µg/ml) and X-gal (200 µg/ml), these were incubated overnight at 37°C. The aim was to identify white colonies that were sensitive to Erm, implying the plasmid had been lost. Potential colonies were tested by colony PCR (see Section 2.3.3) to confirm whether the mutant genotype was present. The PCR product of promising candidates was then purified and sequenced for final confirmation (Eurofins MWG Operon). For point mutations, silent mutations were used in the mutation primers to create enzyme restriction sites for unique cutters so colonies could be tested by colony PCR (see Section 2.3.3) followed by restriction digest (see Section 2.3.6) to confirm the genotype.

2.7.2 Introduction of a 3 x FLAG-tag into lysogenic phage DNA

To allow the assessment of gene expression levels by western blot (see Section 2.3.12) during phage replication, a 3 x FLAG-tag (23 amino acids) was inserted before genes of interest using pMAD (Section 2.7.1). The targeting vector was generated by cloning a 3 x FLAG-tag with the gene of interest as the 3' homology arm and the upstream DNA region as the 5' homology arm. This was achieved by a 3-step PCR series (Figure 2.4), fusing the homology arms either side of the 3 x FLAG-tag. For the initial PCRs (PCR 1), one set of primers was used to amplify the 5' homology arm (*primers 1 and 2*), a second set was used for amplification of the 3' homology arm (*primers 5 and 6*), and a third set was used to amplify the FLAG-tag (*primers 3 and 4*). The homology arms were then fused either side of the 3 x FLAG-tag (PCR 2 and 3), using the 20 bp overlaps complementary to the FLAG-tag on *primers 2 and 5* as described in Section 2.7.1 and shown in Figure 2.4.

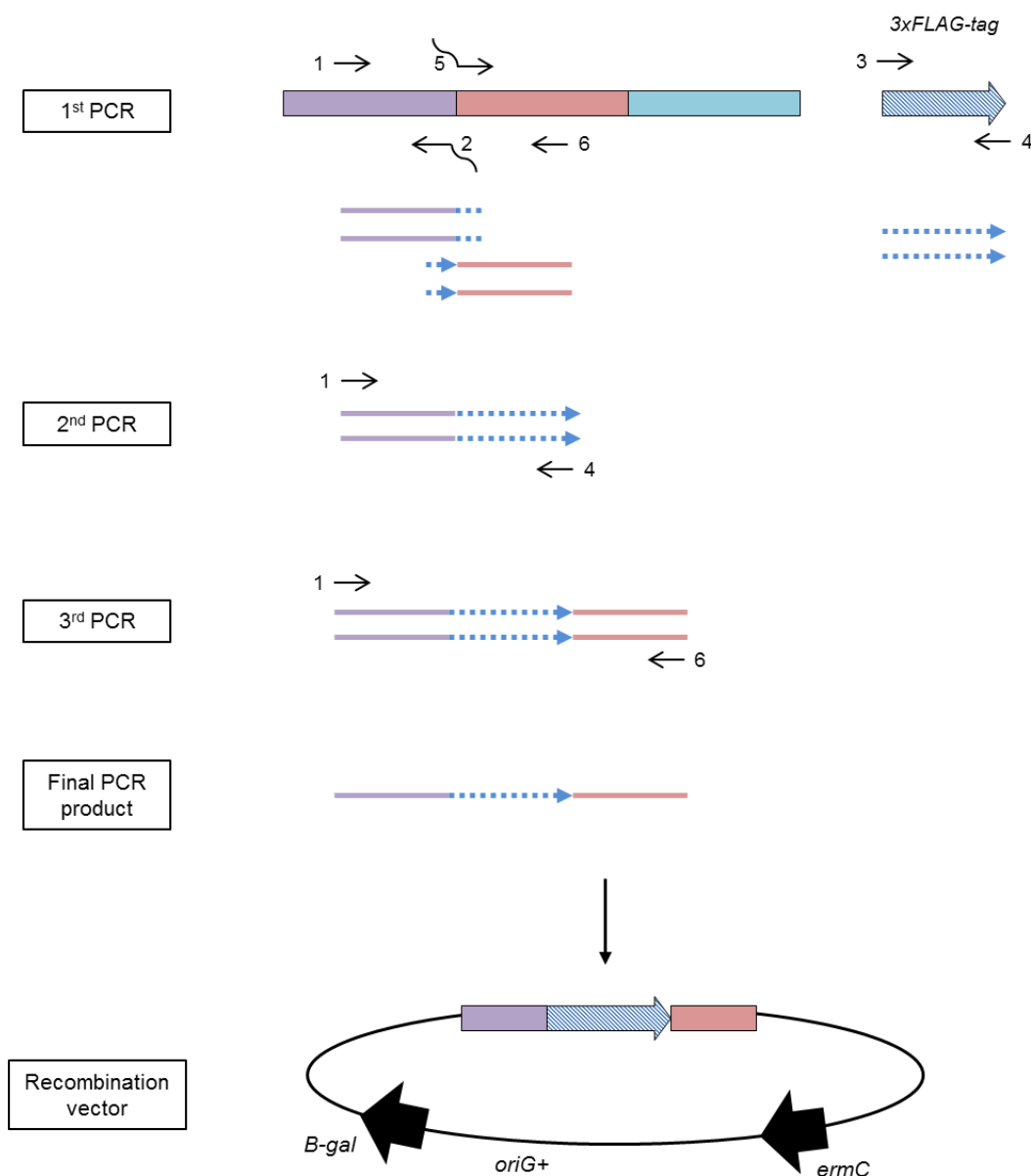


Figure 2.4 A schematic representation of the triple PCR necessary for the construction of the 3 x FLAG-tag with homology arms. The red box illustrates the chromosomal gene that is the target for the upstream, insertion of the 3 x FLAG-tag. The FLAG-tag is shown as a blue, cross-hatched arrow-box. The purple and green boxes represent the flanking genes. Primer hybridisation sites are shown by arrows, with the FLAG-tag complementary *primer 2* and *5* tails indicated. The arrangement of the resultant pMAD recombination vector is represented, with the origin of replication labelled and large arrows showing the location of the erythromycin cassette and β -galactosidase gene.

The cloned pMAD plasmids were confirmed in DH5 α before being transformed to the appropriate *S. aureus* strain by electroporation (Section 2.4.4). The homologous recombination was performed as described in Section 2.7.1, with clones containing the inserted 3 x FLAG-tag selected (Figure 2.5).

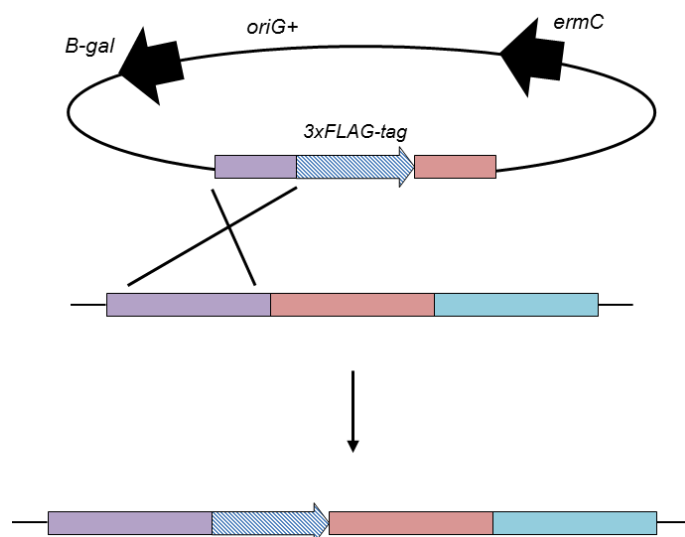


Figure 2.5 A schematic representation of pMAD mediated 3 x FLAG-tag insertion. Plasmid integration occurs during the first homologous recombination. At this point, colonies are an intermediate blue and grow in the presence of Erm. The second homologous recombination leads to plasmid deletion and bacteria are sensitive to Erm. Two genotypes can be produced: wild-type or with 3 x FLAG-tag.

2.8 Data analysis

Raw data was organised in Microsoft® Excel® 2010. Phage and PICI titres were calculated as PFU/ml or CFU/ml, respectively. Titres were analysed using GraphPad Prism® 6 (La Jolla). Phage and PICI titre assays were performed in biological and technical triplicate. Results are shown as the mean. Assays using *blaZ* expression vectors were calculated as units/ml. The change in slope/h was calculated in GraphPad Prism® 6 (La Jolla), before transfer to Microsoft® Excel® 2010 for the calculation of units/ml. The units/ml calculations were analysed and graphed using GraphPad Prism® 6 (La Jolla). Expression assays were performed in biological and technical triplicate. Results are shown as the mean with standard error of the mean (SEM).

The NCBI BLAST server programme www.ncbi.nlm.nih.gov³⁸⁵ was used for homology comparisons of sequences against the GenBank database. CLC Genomics Workbench 7 was used for assembling sequences to a reference, for multiple nucleotide and protein sequence alignment, for the construction of phylogenetic trees based on protein sequence, and for other general sequence analysis. The Praline server programme²⁸⁵ was also used for multiple sequence alignment.

2.8.1 Statistical analysis

To compare the expression assays the data was assessed for normal distribution in GraphPad Prism® using the D'Agostino-Pearson normality test on sample sizes of seven or more. The p -values were not significant ($p = \geq 0.05$), indicating the data was normally distributed and so parametric statistical analysis was carried out. The mean, standard deviation, and SEM were calculated and statistical significance determined by 2-way analysis of variance (ANOVA) with Sidak's or Holm-Sidak's multiple comparisons test (as specified in figure legend) to determine significance compared to controls. In all experiments; $p \leq 0.05$ was considered significant.

2.8.2 Structural modelling of the wild-type and mutant dimeric Duts

Prediction models of the three-dimensional structures of the different phage-encoded dimeric Duts were generated using the I-TASSER server³⁶⁹⁻³⁷¹. Models with the highest C-scores were chosen and selected for further structural analysis using PyMOL software^{386,387}. Crystallographic structures were solved and the three-dimensional structure models analysed at the Instituto de Biomedicina de Valencia (IBV-CSIC) by Dr. Alberto Marina and his laboratory group.

2.8.3 Structural modelling of the SaPIbov1 Stl

Prediction models of the three-dimensional structure of the SaPIbov1 Stl were generated using the I-TASSER server³⁶⁹⁻³⁷¹ and the Phyre2 (Intensive mode) server³⁸⁸. The I-TASSER model with the highest C-score was chosen and selected for further structural analysis using PyMOL software^{386,387}.

Chapter 3 The dimeric Duts are a subset of phage-encoded SaPI inducers

Results from this section are included in:

Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer (2017) **Bowring J***, Neamah MM*, Donderis J*, Mir-Sanchis I, Alite C, J Ciges-Tomas R, Maiques E, Mehmedov I, Marina A, Penadés JR. *eLIFE*. DOI: 10.7554/eLife.26487

Convergent evolution involving dimeric and trimeric dUTPases in pathogenicity island mobilization (2017) Donderis J*, **Bowring J***, Maiques E*, Ciges-Tomas R, Alite C, Mehmedov I, Angeles Tormo-Mas M, Penadés JR, Marina A. *PloS Pathogens*. DOI: 10.1371/journal.ppat.1006581

*Denotes joint first authorship

Disclaimer on the work performed:

The mass spectrometry referred to here was performed by Rafael Ciges at the proteomics facility of SCSIE University of Valencia.

3.1 Background

Bacteriophages and PICIs are MGEs that make up a key component of the wider horizontal transfer map. Bacteriophages have been implicated in the maintenance and genomic diversity of prokaryotic populations^{389,390}. They also play an important role in bacterial pathogenicity and evolution^{78,391}. Similarly, PICIs contribute to the spread of bacterial virulence genes and are innately tied to the bacteriophage life cycle, requiring the infection of a helper phage or the induction of a prophage to trigger their induction, as well as for their packaging^{270,283,297,305,325}.

Of the different PICIs, the *S. aureus* Pathogenicity Islands (SaPIs) are the best characterised. As previously described, they reside passively in the bacterial host chromosome under control of the master repressor, Stl²⁷⁸. SaPI-induction is triggered by a phage protein:SaPI protein interaction between the SaPI master repressor (Stl) and an inducing phage protein²⁸⁰. Different SaPIs encode different Stl repressors, with poor sequence conservation between these repressor variants²⁸⁰. As such, inducing phage proteins are specific to certain SaPIs and differ depending on the Stl repressor.

At the moment, 4 different types of phage proteins have been characterised as SaPI inducers, the DNAI-binding protein Sri²⁸⁰, 80 α -encoded ORF15²⁸⁰, different phage-encoded recombinases (Sak, Sak4, RedB and Erf)^{309,310}, and the trimeric Duts (dUTPases)²⁸⁰. The different inducers are each specific to a different Stl repressor, with Sri inducing SaPI1, the 80 α ORF15 inducing SaPIbov2, the recombinases inducing SaPI2 and the trimeric Duts inducing SaPIbov1. The trimeric Duts are particularly well characterised as the inducing phage protein of the SaPIbov1 island. As described in detail in section 1.4.4.4, the trimeric Duts interact with the Stl repressor of SaPIbov1 in a manner analogous to the signalling mechanism of eukaryotic G proteins^{304,305,312,368}. SaPI derepression by the trimeric Duts involves two catalytically conserved Dut motifs (motif IV and V) together with the unconserved region responsible for the high number of trimeric Dut allelic variants, named motif VI. Motifs IV and VI form the binding site for Stl, whilst the conserved, but highly flexible motif V stabilises the complex. Additionally, trimeric Dut SaPI induction also involves dUTP as a

second messenger molecule, which blocks the interaction with Stl when bound to the Dut³⁰⁴.

Recently, it has been identified that phages can encode dimeric Duts in place of trimeric Duts, with all phages encoding one or the other, but never both²⁸¹. Furthermore, the dimeric Dut encoded by Φ NM1 has also been characterised as an inducer of SaPIbov1³⁵⁹. That a phage encoding a dimeric Dut can still mobilise the SaPIbov1 island is striking; the four dimeric Dut structures that have been reported (*Trypanosoma cruzi*³⁵⁶, *Trypanosoma brucei*³⁵⁷, *Campylobacter jejuni*³⁵⁸, and *Leishmania major*³⁵⁵) indicate an all-helical structure completely different to the β -sheet folding of the 80 α and Φ 11-encoded trimeric Duts^{306,312}. This difference also extends to sequence level, with the catalytic motifs of the dimeric Duts completely divergent from those of the trimeric Duts. How are two completely different Dut proteins both interacting with the same Stl repressor to induce SaPIbov1?

The aims of this study were to further characterise the dimeric Duts of the staphylococcal bacteriophages. This work aimed to identify how prevalent the phage-encoded dimeric Duts are, and whether they display the divergent region allelic variation found in the trimeric Duts. The study aimed to establish the existence of any groups or families, and identify whether there was variation in the SaPI-inducing capabilities of any families found.

Particularly, we wanted to investigate the relationship between the different phage-encoded dimeric Duts and SaPIbov1. SaPIbov1 encodes the Stl repressor characterised as targeting the trimeric Duts and the Φ NM1-encoded dimeric Dut, leading to derepression of the island. In addition to this, we aimed to characterise the interaction of the dimeric Duts with SaPIbov5, an island encoding the same Stl as SaPIbov1. We expected that the interaction between the dimeric Duts and Stl_{SaPIbov5} would mirror the interaction with Stl_{SaPIbov1}, given that they are identical. The reasoning behind using both these islands for characterisation experiments was due to the different phage packaging mechanisms referred to in Sections 1.2.10.2 and 1.3.2.4. SaPIbov5 encodes a *cos* site, allowing it to utilise *cos* phage packaging systems, whereas SaPIbov1 is a prototypical *pac* island and can only be packaged by *pac* phages^{308,320}. Using both

islands allowed the characterisation of Duts encoded by both *pac* and *cos* phage types.

This study characterises the SaPI_{bov1}- and SaPI_{bov5}-inducing abilities of 6 allelic variants of phage-encoded dimeric Dut obtained from existing laboratory stocks. It establishes that not all dimeric Duts are capable of inducing the islands but, for those that do, induction depends on a protein:protein interaction between the Dut and the Stl. Furthermore, this study establishes the existence of 9 phage-encoded dimeric Dut groups and characterises the relationships between these groups and their SaPI-inducing capabilities.

3.2 Results

3.2.1 The phage encoded dimeric and trimeric Duts lack homology in sequence and structure

Regarding the differences between the dimeric and trimeric Duts found in different organisms, much more is known about the trimers, which are the most abundant in nature as a whole. For the phages specifically, there seems to be a more even divide between the numbers of dimeric and trimeric Duts encoded. The structure and function of the trimeric Duts has been extensively characterised, and the structures of the 80 α - and Φ 11-encoded trimeric Duts have both been solved^{305,306,312}. In contrast and as previously mentioned, only four dimeric Dut structures have been characterised and none of these structures represent phage-encoded Duts³⁵⁵⁻³⁵⁸. To confirm that the differences seen between the dimeric and trimeric Duts of other organisms are also found in the phage-encoded Duts, the sequences and structures of prototypical phage-encoded dimeric and trimeric Duts were compared.

To analyse differences between the sequences of the phage-encoded Duts, the 80 α -encoded trimeric Dut and Φ NM1-encoded dimeric Dut were compared. The 80 α trimeric Dut is a known SaPI_{bov1}-inducer and has been extensively studied^{280,304,305}, whilst the dimeric Dut encoded by Φ NM1 is the only dimeric Dut shown to induce SaPI_{bov1}³⁵⁹. *In silico* PRALINE-generated alignment of the protein sequences of these two phage-encoded Duts showed a low level of conservation throughout (Fig 3.1). Indeed, the 80 α trimeric Dut and Φ NM1

dimeric Dut feature only a 10.1% sequence similarity, indicating they are highly divergent.

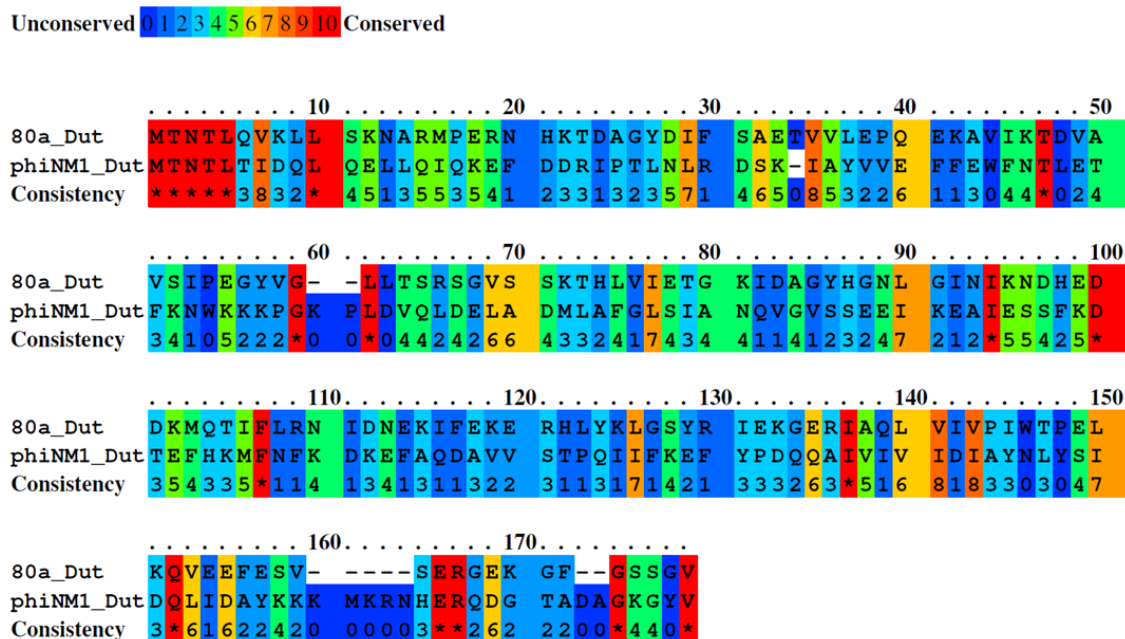
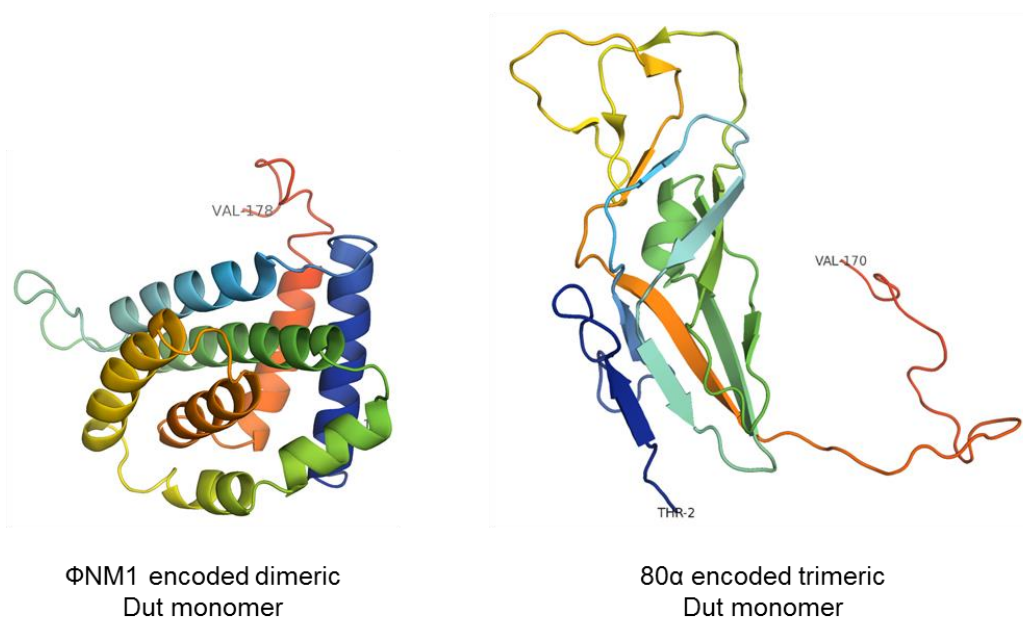


Figure 3.1 The dimeric and trimeric Duts have largely different protein sequences. A protein sequence alignment of the Φ NM1 encoded dimeric Dut and the 80 α encoded trimeric Dut. Colours indicate relative sequence conservation at each position, with red being most conserved and blue being least (adapted from alignment generated by PRALINE)^{284,285}.

Having confirmed the differences in protein sequence, the logical step was to confirm any structural differences. As previously described, whilst the structure of the 80 α -encoded Dut has been characterised³⁰⁴, no phage-encoded dimeric Dut structures have been solved. Therefore, in order to compare the 80 α and Φ NM1 Dut structures, the I-TASSER server was used to generate a predicted structure for the Φ NM1-encoded dimeric Dut (Fig 3.2A). This predicted structure contains 8 α -helices and a cartoon representative is shown in a side by side comparison with the 80 α -encoded trimeric Dut structure (PDB 3ZEZ)³⁰⁴. The structure of the 80 α Dut is made up of β -sheets and the complete divergence in structure with the predicted Φ NM1 Dut is clear. These differences were further confirmed by using the PSIPRED (Predict Secondary Structure) analysis with the PRALINE protein sequence alignment (Fig 3.2B). Here the α -helices of the dimeric Dut and the β -sheets of the trimeric Dut can be seen overlaid on the protein sequence, with no structural overlap between the two phage Dut types. As with the protein sequence, the structures of the phage-encoded dimeric and trimeric Duts are predicted to be completely divergent.

A**B**

	 10	20	30	40	50
(PRED) 80a_Dut	MTN	TLQVKIL	SKNARMPERN	HKTDAGYDIF	SAETVVLEPQ EKAVIKTDVA
(PRED) phiNM1_Dut	MTNT	LTIDQL	QELLQIQKEF	DDRIPTLNLR	DSK-IAYVVE FFEWFNTLET
	 60	70	80	90	100
(PRED) 80a_Dut	VSIPEGYVG-	-LLTSRSGVS	SKTHLVIETG	KIDAGYHGNI	GINIKNDHED
(PRED) phiNM1_Dut	EKNWKKKPGK	P	LDVQLDELA	DMLAFGLSIA	NQVGVSSEEI KEATESSFKD
	 110	120	130	140	150
(PRED) 80a_Dut	DKMQTIFLRN	IDNEKIFEKE	RHLYKLGSYR	IEKGERIAQL	VIVPIWTPEL
(PRED) phiNM1_Dut	TEFHKMFNFK	DKEFAQDAVV	STPQIIFKEF	YPDQQAIVIV	IDIAYNLYSI
	 160	170			
(PRED) 80a_Dut	KQVEEFESV-	---	SERGEK	GF--GSSGV	
(PRED) phiNM1_Dut	DQLIDAYKKK	MKEN	HEEQDG	TADAGKGYV	

HELIX (H) STRAND (E)

Figure 3.2 The dimeric and trimeric Duts have completely different predicted structures. (A) Cartoon representation of 3D structure of the ΦNM1 encoded dimeric Dut monomer (left) generated by I-Tasser and a cartoon representation of the 3D structure of the 80α trimeric Dut monomer (right, PDB 3ZEZ)³⁶⁹⁻³⁷¹. Both monomers are coloured by α-helices (ΦNM1 Dut) and β-sheets (80α Dut). I-TASSER templates and confidence values can be found in Appendix 2. **(B)** The predicted secondary structure of the ΦNM1 and 80α Duts based on a PRALINE protein sequence alignment with PSIPRED analyses^{392,393}. Red blocks of highlighted text indicate α-helices and blue blocks indicate β-sheets.

3.2.2 Characterising the different families of dimeric Duts

Based on the discovery that some phages encode dimeric in place of trimeric Duts, NCBI BLAST was utilised to identify additional phage-encoded dimeric Duts^{281,394}. Starting from the ΦNM1 Dut protein sequence, a total of 59 potential phage-encoded dimeric Duts were found (Appendix 3A). Some of these sequences were annotated as staphylococcal Duts in the NCBI database, however a number of staphylococcal species, including *S. aureus*, have been identified as

lacking any *dut* gene³⁴⁵. Taking this into account, Duts annotated as *S. aureus* would most likely originate from a prophage in the strain; as such, these have been included in the subsequent analyses. A PRALINE-generated protein alignment of the 59 sequences showed differences between the Duts suggesting allelic variation (Appendix 3B), so a distance-based phylogenetic tree was constructed to classify this (Fig 3.3).

A CLC workbench generated alignment of the full protein sequences of the 59 phage-encoded dimeric Duts (Appendix 3C) was used to construct the radial phylogenetic tree shown in Fig 3.3, using the neighbour joining method. Four phage-encoded trimeric Duts (80α, Φ11, ΦSLT, and Φ12/85, Appendix 3A) were included as outliers and to show their relationship to the dimeric Duts (Fig 3.3 - inset). The phage-encoded dimeric Duts divided into 9 distinct groups, with the trimeric Duts clustering distantly.

To further investigate the 9 groups, a dimeric Dut was chosen from each group as a representative to be taken forward for further analysis. These representatives were chosen based on the existence of phages encoding the Duts in laboratory stocks, thus allowing their subsequent cloning and study (excepting HKU10-03, ΦStauST398-3, and ΦMR25 - there were no appropriate phage stocks for each of these groups). As such, the representatives of the 9 groups taken forward for further analyses were: ΦStauST398-3, Φ55, ΦDII, ΦMR25, ΦDI, ΦO46, HKU10-03, ΦO11, and ΦNM1.

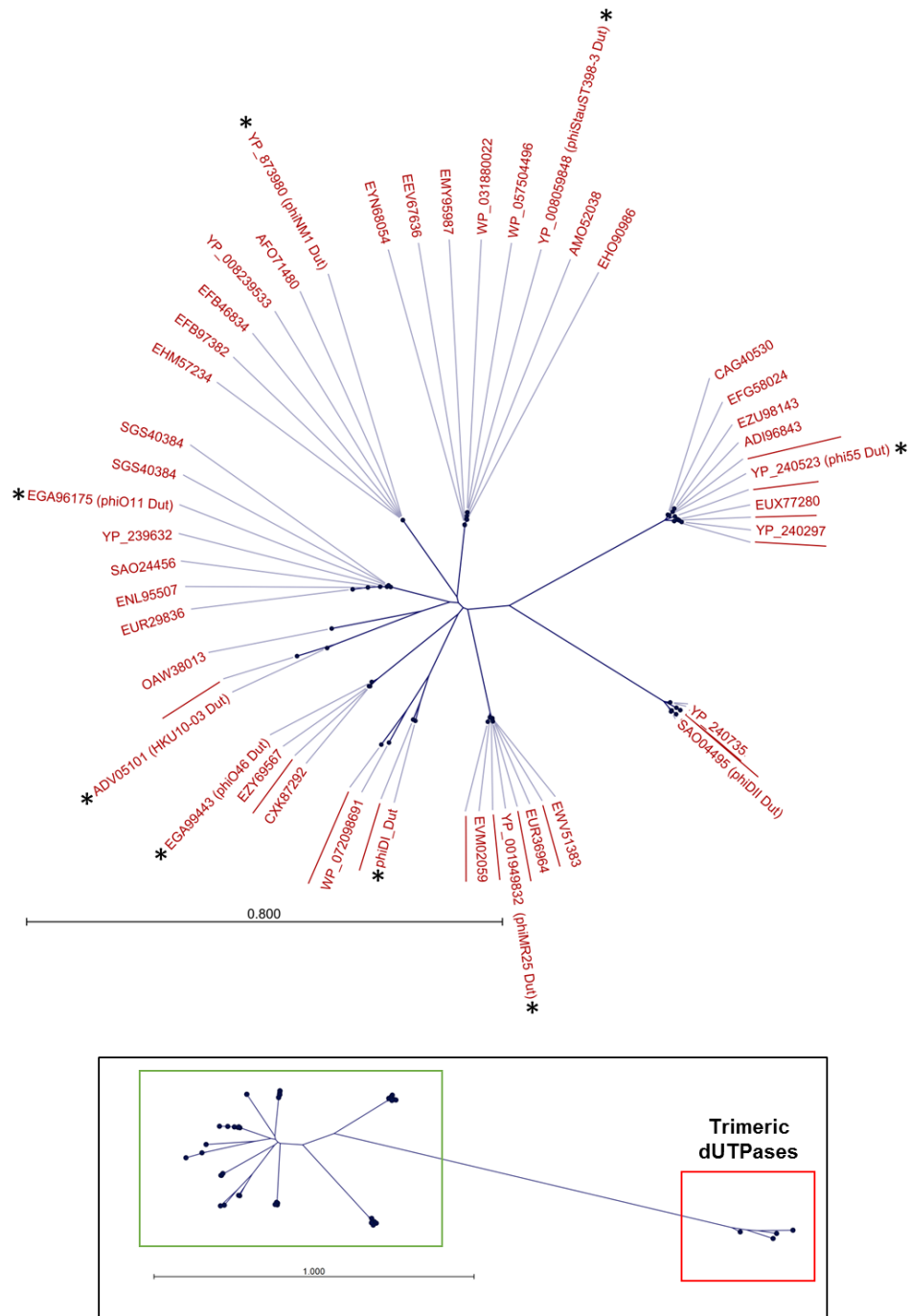


Figure 3.3 Radial phylogenetic tree of identified *Staphylococcus* phage dimeric Duts. Radial tree showing the arrangement of the dimeric Duts into 9 different groups, constructed using an alignment of protein sequences of the dimeric Duts from staphylococcal phages. An NCBI BLAST of the known SaPIbov1-inducing Φ NM1 encoded dimeric Dut (full sequence) was used to select 59 dimeric Dut sequences that were annotated as staphylococcal phage Duts or as staphylococcal Duts. The full sequences of these Duts were used for protein alignment and the tree. *Staphylococcus aureus* does not encode a genomic Dut so any such annotations would also be phagic. Four trimeric staphylococcal phage Duts were also included in the alignment and tree to provide outliers and show the distance of these proteins from the dimeric Duts. All the included sequences and their accession numbers as well as the CLC alignment used to create this tree are included in Appendix 3. The small boxed section below shows the full tree, including the trimeric Dut outliers (red box), which were hidden for clarity in the more detailed tree above. This detailed tree consists of the area highlighted with the green box in the full depiction. Representatives of each group that were used in subsequent analyses are highlighted with a *. The tree was created using the following parameters; algorithm = Neighbour Joining, distance measure = Jukes-Cantor, bootstrap = 100 Replicates. Proteins are named by NCBI accession number with the exception of the group representatives, where the name of the corresponding phage is included in brackets (Appendix 3).

Previous observations of the phage-encoded trimeric Duts indicated the existence of allelic variation, leading to Duts with differing affinities for binding the Stl repressor³⁰⁴. This allelic variation was due to a central unconserved region, now named motif VI. Based on these observations and the initial differences identified between the 59 dimeric Duts, a protein alignment of the phage-encoded dimeric group representatives was performed to establish the levels of sequence conservation and identify the location of any divergent regions similar to the trimeric Dut motif VI (Fig 3.4).

The PRALINE-generated protein sequence alignment showed that the allelic variation between the 9 dimeric Dut groups was primarily due to a central unconserved region, similar to that observed in the trimeric Duts (Fig 3.4). This unconserved region is variable in both size and sequence. Like with the trimeric Duts, the unconserved region is flanked by two conserved regions, which contain the five proposed catalytic motifs present in all dimeric Duts³⁵⁸. It is interesting that although the phage-encoded trimeric and dimeric Duts display divergent sequences and structures, both exhibit a variable central region flanked by two conserved regions containing 5 catalytic motifs.

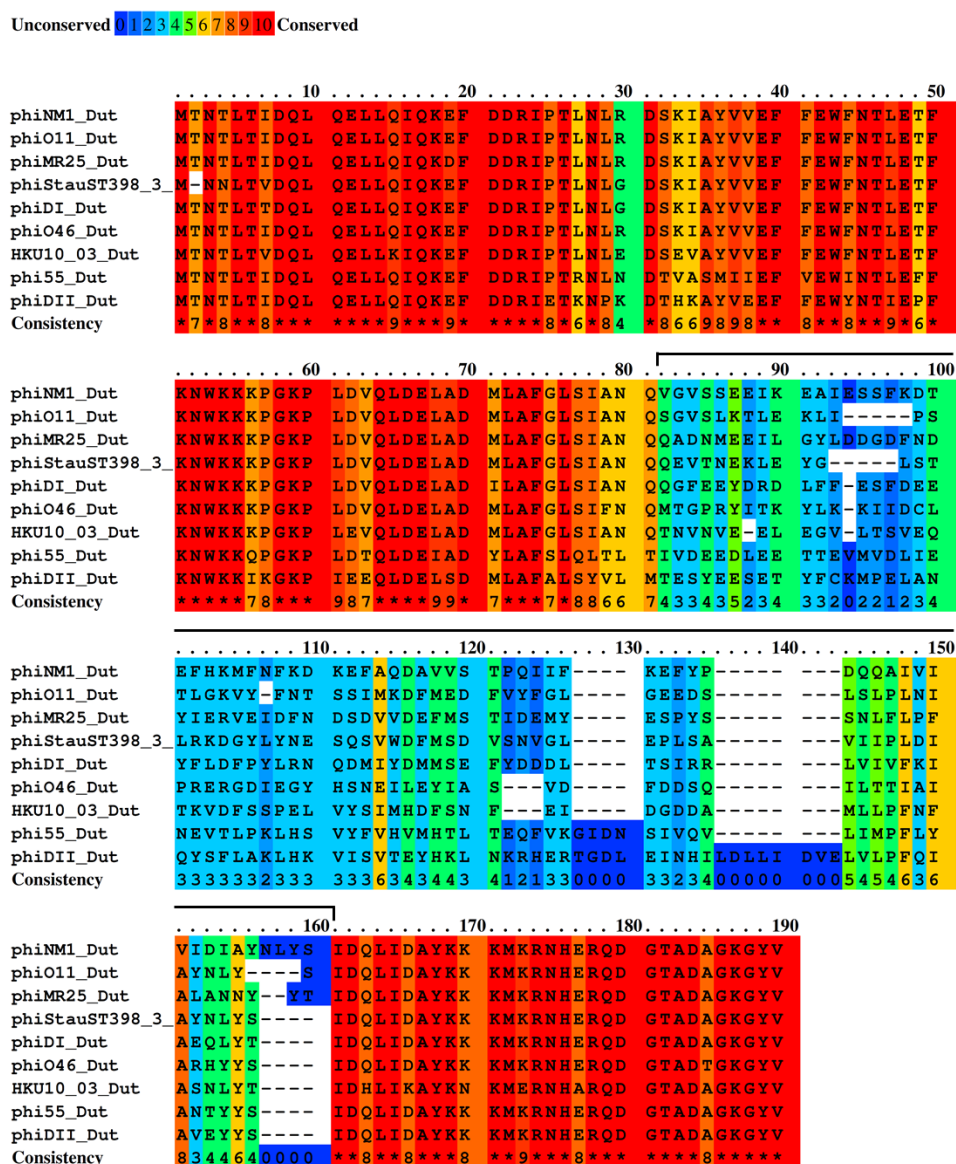


Figure 3.4 Alignment of *Staphylococcus aureus* phage encoded dimeric Duts shows allelic variation. Alignment of dimeric Duts from *S. aureus* phages Φ NM1, Φ DI, Φ DII, Φ O11, Φ O46, Φ MR25, StauST398-3, HKU10_03 and Φ 55 are shown representing the nine different dimeric Dut families identified. The unconserved domain is indicated by the black bracket. Pairwise identities that are conserved across all Duts are indicated by the * in the 'Consistency' row, otherwise the number of identical residues are numbered. Colours indicate relative sequence conservation at each position, with red being most conserved and blue being least (adapted from alignment generated by PRALINE)^{284,285}.

The trimeric Duts of different organisms, for example those from *S. aureus* phage 80 α , *M. tuberculosis*, and *R. norvegicus*, have been shown to exhibit similar structural conformations over the conserved motifs, while the unconserved motif VI exhibits varying conformations in each different organism³¹². As previously mentioned, no phage-encoded dimeric Dut structures have been solved to date, although the structure of the *C. jejuni* bacterial Dut has been³⁵⁸. To investigate how the newly identified phage-encoded dimeric

Duts compare structurally to the previously established dimeric Dut structures, I-TASSER structure predictions were constructed for each of the 9 dimeric Dut group representatives (Fig 3.5). The templates and confidence values for these predictions can be found in Appendix 2. All predicted structures are shown in cartoon representation, with the closest structural analog, the *C. jejuni* dimeric Dut (PDB 1W2Y), included for comparison.

The predicted structures of the phage-encoded dimeric Duts are shorter than the structure of the *C. jejuni* Dut, containing 7 or 8 α -helices in comparison to the 11 helices of the *C. jejuni* Dut. The phage-encoded Duts lack the $\alpha 2$ and the C-terminal helices $\alpha 10-11$ seen in the *C. jejuni* Dut. However, in a parallelism to the trimeric Duts, the helices contained within the conserved regions of the phage-encoded dimeric Duts ($\alpha 1-3$, and $\alpha 7/8$) show a similar conformation throughout the 9 representatives, and also with the corresponding *C. jejuni* Dut helices ($\alpha 1$, $\alpha 3-4$, and $\alpha 9$, respectively).

Given that the phage-encoded dimeric Dut cartoon representatives are based on predictions and that their unconserved domain is poorly defined, it is difficult to make any judgements about the structural conformations in this region. The unconserved region of the *C. jejuni* Dut ($\alpha 5-7$) appears to differ from the phage-encoded Duts ($\alpha 4-6/7$), although this would need to be confirmed by solving the structure of a phage-encoded dimeric Dut.

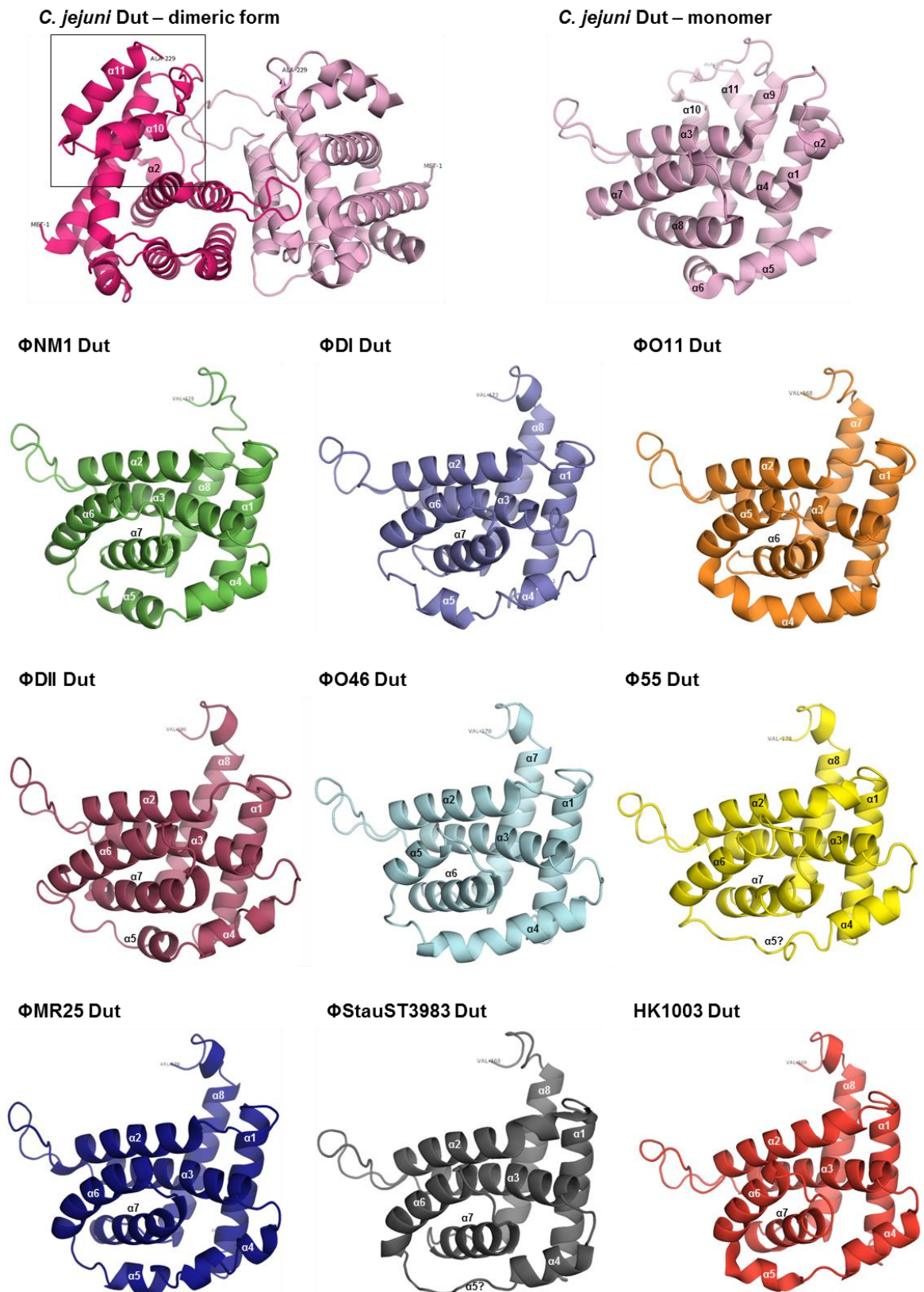


Figure 3.5 Comparison of the predicted structures of each phage-encoded dimeric Dut group representative and the solved structure of the *Campylobacter jejuni* – encoded dimeric Dut. Cartoon representations of the monomer 3D structures of each representative of the different phage-encoded dimeric Dut groups as generated by I-TASSER^{369–371}. The closest structural analog, the *C. jejuni* encoded dimeric Dut (PDB 1W2Y), is shown both in dimeric and in individual monomer form for comparison. I-TASSER templates and confidence values can be found in Appendix 2.

These structural predictions imply that the conserved motifs identified in the dimeric Dut sequences translate to a conserved structural conformation, whilst the unconserved regions are areas of potential structural variance. Again, this mirrors the trimeric Duts where the unconserved motif VI displays completely different configurations in different organisms, whilst the conserved motifs are fairly consistent³¹². Furthermore, the motif VI of different phage-encoded trimeric Duts also shows some differences in conformation³⁰⁶.

3.2.3 Induction of SaPIs by phage-encoded dimeric Duts

The variation of motif VI within the phage-encoded trimeric Duts contributes to these Duts displaying different levels of affinity for binding the Stl repressor of SaPIbov1³⁰⁴. Although the sequences and structures of the phage-encoded dimeric Duts are completely divergent to the trimeric, this study has identified apparent similarities in the manner of allelic variation, with both phage-encoded Dut types containing a unconserved region flanked by conserved domains. This implied that the variations in Stl binding affinity seen in the phage-encoded trimeric Duts could potentially be mirrored in the dimeric Duts.

To test this hypothesis, the capabilities of the different phage-encoded dimeric Duts in inducing SaPIbov1 and SaPIbov5 were analysed. Six different phage-encoded dimeric Duts were available from the laboratory stocks, those encoded by Φ NM1, Φ DI, Φ O11, Φ DII, Φ O46, and Φ 55 respectively. These were 3xFLAG-tagged and cloned into the pCN51 expression vector under a *Pcad* promoter. The cloned Duts were introduced into SaPIbov1 and SaPIbov5 positive strains and the induction of these islands following Dut expression was tested by Southern blot (Fig 3.6). As previously mentioned, SaPIbov1 and SaPIbov5 encode an identical Stl repressor, so it was expected that the induction capabilities of the Duts would be consistent for both islands. This was the case, with the dimeric Duts encoded by Φ NM1, Φ DI, Φ O11 and Φ O46 capable of inducing both SaPIbov1 and SaPIbov5. The dimeric Duts encoded by Φ DII and Φ 55 were incapable of inducing either island. This is in contrast to previous SaPI-induction results using trimeric Duts, where all three tested trimeric Duts induced SaPIbov1 albeit with varying degrees of efficiency.

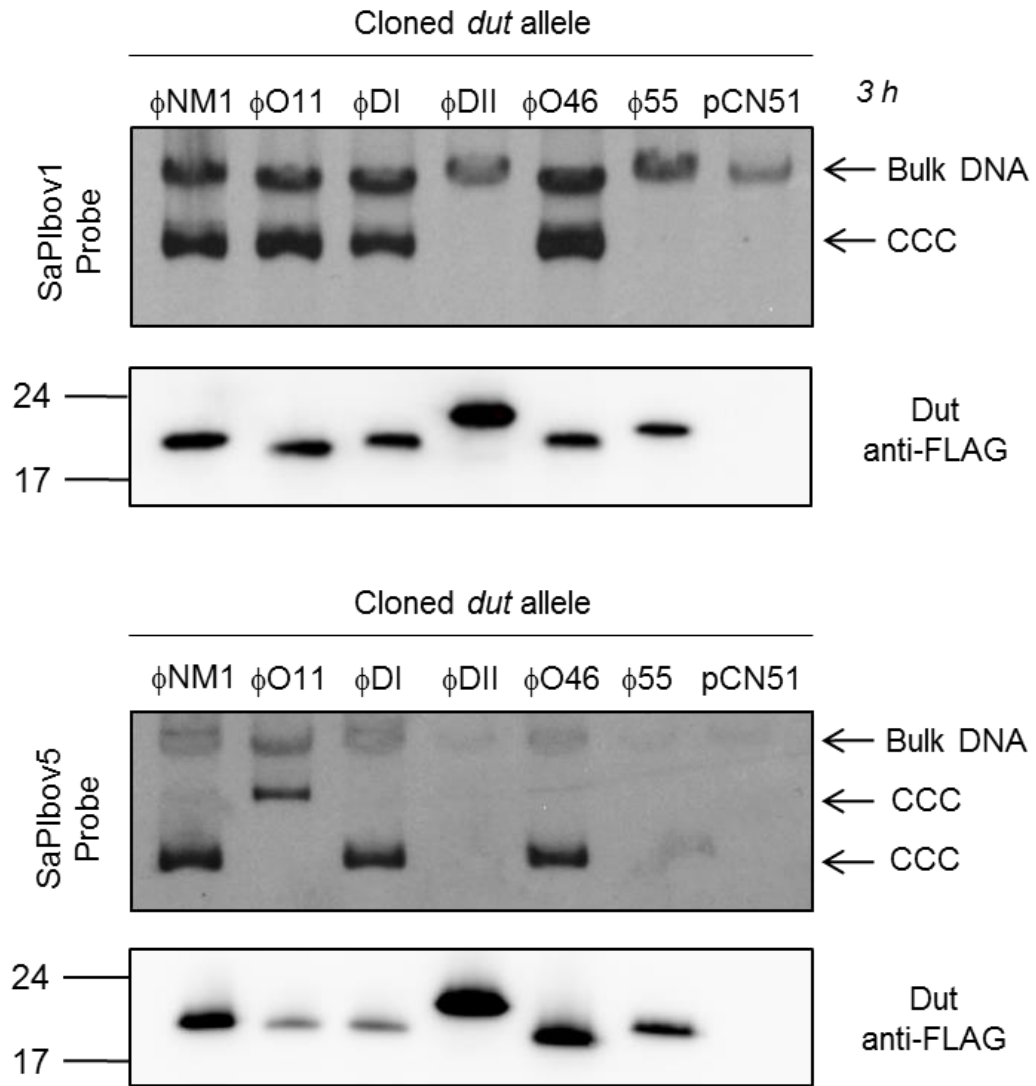


Figure 3.6 SaPIbov1 and SaPIbov5 replication is induced by different dimeric Dut variants. SaPIbov1 and SaPIbov5 excision and replication following induction of cloned ϕ NM1, ϕ O11, ϕ DI, ϕ DII, ϕ O46, and ϕ 55 *dut* genes. Strain JP6774 containing SaPIbov1 and strain JP16140 containing SaPIbov5 were complemented with the different pCN51 plasmids expressing the 3xFLAG-tagged dimeric Duts or the empty pCN51 control. Samples were isolated 3 hours after induction with 5 μ M CdCl₂ and Southern blots were performed using a probe for the SaPIbov1 and SaPIbov5 integrase (Table 2.2). The upper band is 'bulk' DNA, including chromosomal, phage, and replicating SaPI. CCC indicates covalently closed circular SaPI DNA. In these experiments, as no helper phage was present, the excised and replicating SaPI DNA appears as part of the bulk DNA or as CCC molecules, rather than the linear monomers that are seen following helper phage-mediated induction and packaging. The upper band represents bulk DNA, whilst the lower band signifies the SaPI band. The lower panels below each Southern are western blots probed with antibody to the 3xFLAG-tag carried by the Dut proteins at the N-terminal.

Western blots, using anti-FLAG antibodies to probe against the Dut N-terminal 3xFLAG-tags, found that all the Duts were expressing, although protein expression levels were slightly variable. This could mean that the Duts have different levels of stability. However, as the non-inducing Duts consistently

expressed at the same or higher levels than those of the inducing Duts, this variability could not be the cause of the differences in SaPI-induction.

Given that the differences in induction capability could not be attributed to varying expression levels and that the phage-encoded dimeric Duts are fairly well conserved in sequence, the only obvious difference between the inducing and non-inducing Duts seems to be the central unconserved region. This mirrors the previous trimeric Dut results, although there are some interesting differences. Whilst trimeric allelic variation led to differences in StI-binding affinity, all the tested phage-encoded trimeric Duts induce SaPI_{bov1}. However, only three trimeric Duts have been tested for SaPI_{bov1} inducing capability, so there could be some non-inducing trimeric phage-encoded Duts that have not been identified yet. In the case of the phage-encoded dimeric Duts, the differences are more extreme and it seems allelic variation has led to the creation of non-inducing dimeric Duts.

Following on from this, and to more closely examine the relationships between the 9 different phage-encoded dimeric Duts, a phylogenetic tree of the group representatives was constructed (Fig 3.7). Based on a CLC workbench generated protein sequence alignment (Appendix 4), this neighbour joining tree shows the Jukes-Cantor measure of distance between the different phage-encoded dimeric Dut representatives and includes 4 example phage-encoded trimeric Duts as outliers.

As expected, the trimeric Duts group together, distant to the dimers (branch length = 1.402). Using the known SaPI-inducing capabilities of the 6 cloned dimeric Duts it was established that the non-inducing dimeric Duts (Φ DII and Φ 55) seem to sit away from the inducing dimeric Duts, although they themselves are not any more closely related. On the other hand, the known inducing dimeric Duts (Φ NM1, Φ DI, Φ O11 and Φ O46 Duts) seem to form a cluster. This implied that the inducing Duts are more closely related to one another than to either of the non-inducing Duts. The next step would be to identify the SaPI_{bov1} and SaPI_{bov5} inducing capabilities of the remaining three groups (representatives HKU10-03, Φ StauST398-3, and Φ MR25) to establish the consistency of this clustering and whether all the non-inducers sit outwith this group.

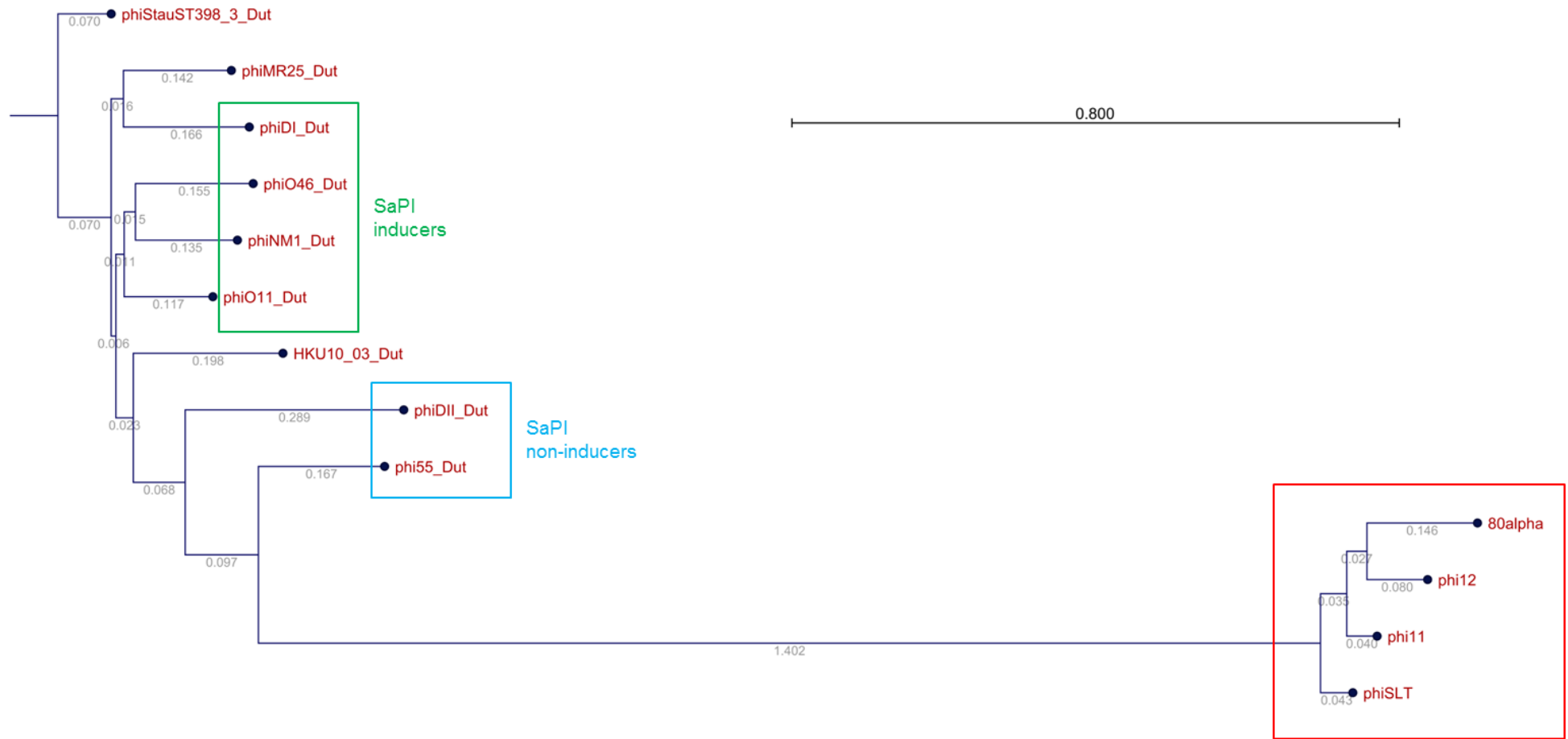


Figure 3.7 Phylogram of the representative Dut protein sequences of the 9 different dimeric Dut families. Phylogenetic tree showing the arrangement of the 9 representatives of the different phage-encoded dimeric Dut groups, constructed using an alignment of the full protein sequences. Four trimeric staphylococcal phage Duts were also included in the alignment and tree to provide outliers, these are highlighted in the red box. Each Dut is named according to the encoding phage. The green box indicates known SaPI-inducers and the blue box indicates known non-inducers. The tree was created using the following parameters; algorithm = Neighbour Joining, distance measure = Jukes-Cantor, bootstrap = 100 Replicates. The CLC alignment used to construct this tree is included in Appendix 4.

Using a CLC workbench generated protein sequence alignment of the phage-encoded dimeric Dut representatives alone (Appendix 5), pairwise comparisons were performed to further investigate the relationships between the different groups (Fig 3.8). Comparisons of the percentage identity and number of differences including gaps (Fig 3.8A, upper and lower, respectively) both indicate that the non-inducing Φ DII and Φ 55 Duts have the lowest percentage identity and the highest number of sequence differences, in relation to the other Duts. Similarly, the non-inducing (Φ DII and Φ 55) Duts display the largest number of gaps and the highest Jukes-Cantor corrected distance measures when compared to the other phage-encoded dimeric Duts (Fig 3.8B, upper and lower respectively).

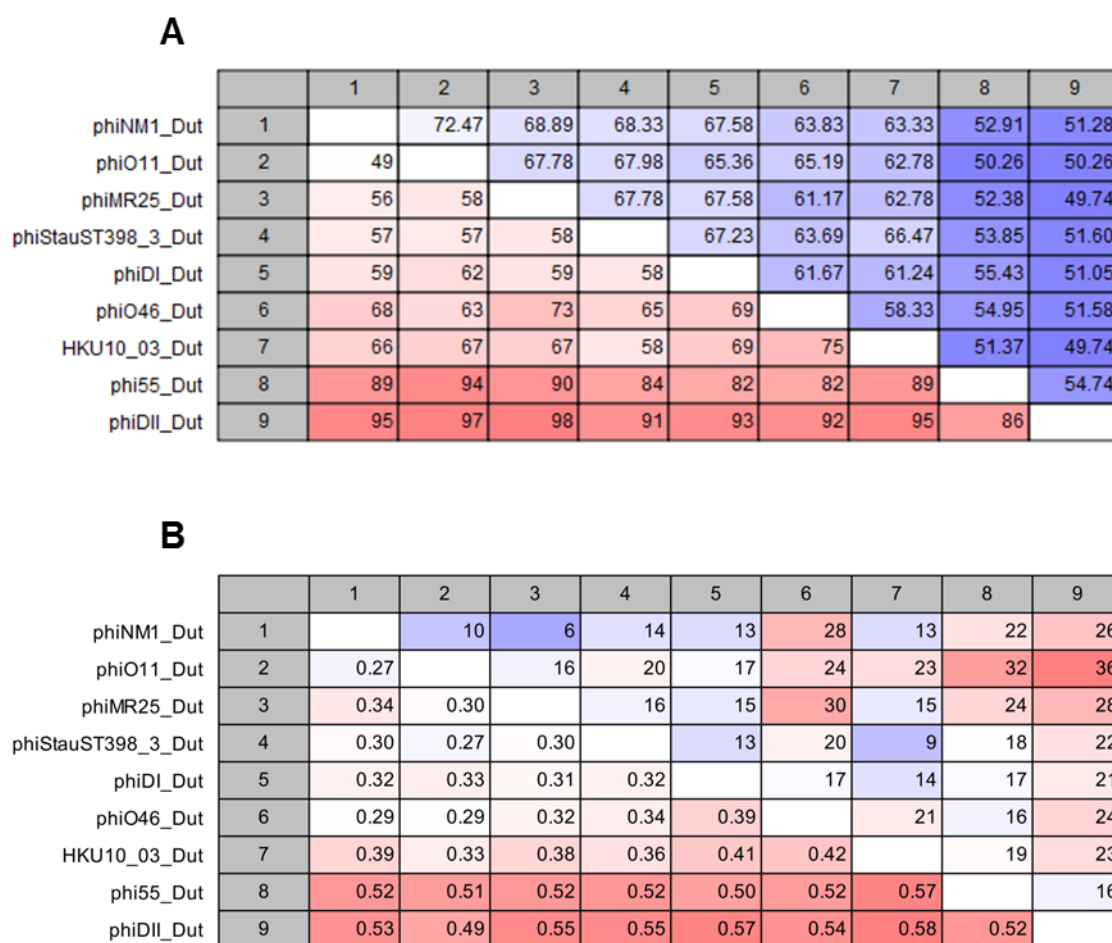


Figure 3.8 Pairwise comparisons of the representative Duts for each of the 9 families. These pairwise comparisons were constructed using the protein sequence alignment of the listed phage-encoded Duts (shown in Appendix 5). The gradient of blue to red represents a sliding scale, with dark blue as minimum, white as median, and dark red as maximum. **(A) Showing the percentage identity above and the differences below.** The upper triangle shows the percentage of identical residues in alignment positions to overlapping alignment positions between the two sequences. The lower triangle shows the number of alignment positions where one sequence is different from the other, including gap differences **(B) Showing gap numbers above and the distances below.** The upper triangle shows the number of alignment positions where one sequence has a gap and the other does not (gaps). The lower triangle shows the Jukes-Cantor corrected distance between two sequences (distance).

This indication that the two non-inducing phage-encoded dimeric Duts (Φ DII and Φ 55) are the most divergent in sequence supports the phylogenetic tree data indicating they cluster more distantly from the rest of the dimeric Duts. It would be logical that the two phage-encoded dimeric Duts that cannot induce the SaPIs contain more sequence differences than the SaPI-inducing dimeric Duts, the assumption being that some of these sequence differences contribute to the loss of induction ability. This could, therefore, represent the evolving strategy of the two non-inducers.

3.2.4 SaPI induction is Dut specific and interferes with phage transfer

Having identified the SaPI-inducing capacities of the 6 representatives of the phage-encoded dimeric Duts, the next step was to clearly confirm this capacity in the phage itself. Furthermore, confirmation of dimeric Dut SaPI induction using viable phages would establish whether induction was solely dependent on the Dut proteins or involved another protein. Two SaPI_{bov1}-inducing phages were carried forward for these experiments based on their availability and ability to infect and form plaques on the RN4220 laboratory strain.

The SaPI-inducing Φ NM1 and Φ DI encoded Duts were selected and inframe deletions of the *dut* gene were constructed in the equivalent prophage genomes. These phages were chosen as dimeric Dut encoding prototypes of the two phage packaging systems; Φ DI uses the *cos* packaging mechanism, whilst Φ NM1 is a fairly well characterised *pac* (headful) packaging phage. The constructed *dut* knock-out mutants and equivalent wild-type phages were introduced into strains containing the appropriate SaPI, based on packaging type: SaPI_{bov1} for Φ NM1 and SaPI_{bov5} for Φ DI. The ability of the phages to induce the SaPI cycle was assessed by Southern blot following SOS response induction of the phages by MC. As expected, the wild-type phages, but not their equivalent *dut* mutants induced the SaPIs (Fig 3.9).

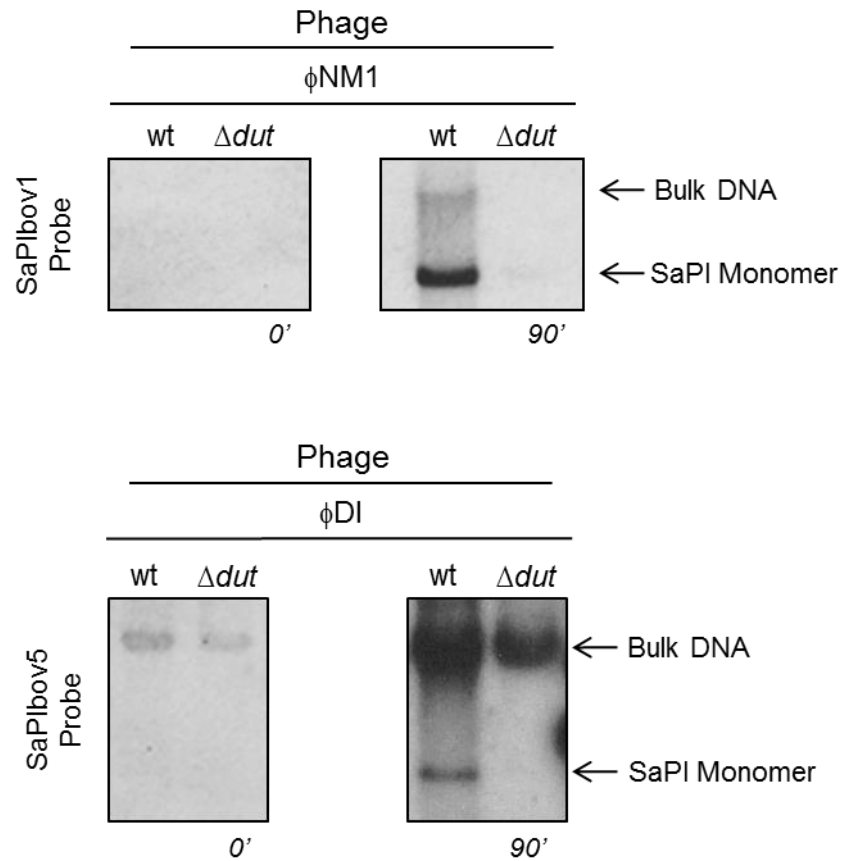


Figure 3.9 SaPIbov1 excision and replication following induction of Φ NM1 and SaPIbov5 replication following induction of Φ DI. Wild-type phages or those with an inframe deletion of the *Dut* gene are shown. Southern blots were performed using strains with SaPIbov1 or SaPIbov5 and respective phages. Samples were isolated at 0' or 90' minutes following induction with mitomycin C. Southern blots were performed using a probe for the SaPIbov1/5 integrase. The upper band represents bulk DNA, whilst the lower band signifies the SaPI linear monomers released from phage heads.

The SaPIs used in these experiments carry an antibiotic resistance marker for selection; the SaPIbov1 carries a *tetM* marker, whilst the SaPIbov5 carries an *ermC* marker. This enabled lysates generated from the strains used in Fig 3.9 to be used in transfer studies. These transfer studies analysed the lysates for the presence of infectious phage and SaPI particles, and quantified the levels of those particles present. As shown in Table 3.1, the *dut* mutants had a significantly lower level of SaPI transfer in comparison to the levels of wild-type phage SaPI transfer. Indeed, whilst the wild-type Φ NM1 transferred SaPIbov1 and SaPIbov5 at 10^7 and 10^6 CFU ml⁻¹ respectively, the *dut* mutant had very low levels of SaPI transfer (<10). Similarly, the wild-type Φ DI transferred SaPIbov5 at a level at 10^4 ml⁻¹, whilst the *dut* mutant had very low SaPIbov5 transfer levels (<10). The *cos* Φ DI is not capable of transferring the *pac* SaPIbov1, hence the low transfer level even in the wild-type (<10).

Table 3.1 The transfer of SaPIbov1 and SaPIbov5 by dimeric Duts Φ D1 and Φ NM1^a.

	No island	SaPIbov1			SaPIbov5		
Phage	Φ titre ^b	Rpl ^c	SaPI titre ^d	Titre ^b	Rpl ^c	SaPI titre ^d	Φ titre ^b
Φ D1	7.91×10^4	+	<10	4.29×10^4	+	3.65×10^5	4.63×10^2
Φ D1 Δ <i>dut</i>	1.06×10^5	-	<10	7.40×10^4	-	<10	3.14×10^5
Φ NM1	3.62×10^9	+	4.47×10^7	3.56×10^7	+	6.93×10^6	6.50×10^8
Φ NM1 Δ <i>dut</i>	4.20×10^9	-	<10	1.35×10^9	-	<10	2.22×10^9

^aThe means of the results of three independent experiments are presented. Variation was within 5% in all cases.

^bPlaques ml⁻¹ of lysate, using RN4220 as indicator.

^cRpl, replication as determined by Southern blot.

^dTransductants ml⁻¹ of lysate, using RN4220 as recipient.

As with previous studies using the SaPIs, the presence of SaPIbov1 and SaPIbov5 interfered with phage transfer, leading to reduced phage titres in the SaPI-positive wild-type strains of 10^1 - 10^2 PFU ml⁻¹. This was true for all the wild-type strains except the wild-type Φ DI SaPIbov1-positive strain where, as expected for a phage that cannot transfer the SaPI, no large reduction in phage titre was seen. This reduction in phage titre was alleviated in the *dut* mutant strains, with phage titres similar to those of the SaPI-negative wild-type phage strains. Again, this was expected based on the hypothesis that without the *dut* gene present the phages would not induce the SaPIs.

Similar patterns of interference were seen when phages were transferred into recipient strains containing the islands (Table 3.2). Here the wild-type phage strains Φ NM1 and Φ DI without any islands were transferred to recipients containing SaPIbov1 and SaPIbov5, respectively. Compared to the wild-type titres into a recipient without an island, wild-type titres to recipients containing the islands were reduced by $\sim 10^5$ PFU ml⁻¹ indicating high levels of interference. As anticipated and supporting the hypothesis that the Duts are the sole SaPIbov1 and SaPIbov5 inducers in these strains, the *dut* mutants escaped this interference, with titres at the same levels as the island-negative wild-type strains.

Table 3.2 Interference of SaPIbov1 and SaPIbov5 with phage Φ D1 and Φ NM1 titres^a.

	Recipient strains		
	No island	SaPIbov1	SaPIbov5
Phage	Φ titre ^b	Φ titre ^c	Φ titre ^d
Φ D1	3.37×10^5		<10
Φ D1 Δ <i>dut</i>	4.63×10^5		4.00×10^5
Φ NM1	1.30×10^{10}	1.15×10^5	
Φ NM1 Δ <i>dut</i>	5.64×10^{10}	1.09×10^{10}	

^aThe means of the results of three independent experiments are presented. Variation was within 5% in all cases.

^bPlaques ml⁻¹ of lysate, using RN4220 as indicator.

^cPlaques ml⁻¹ of lysate, using RN4220 SaPIbov1 as indicator.

^dPlaques ml⁻¹ of lysate, using RN4220 SaPIbov5 as indicator.

Given the high levels of interference the SaPIs impose on phage transfer, the phage can be evolved by titration on to a recipient strain containing the interfering SaPI. Evolved phages that have overcome SaPI interference to grow on the recipient can be sequenced for genes affected by spontaneous mutations, an indicator that they could be the SaPI-inducing phage protein. With phages Φ NM1 and Φ DI, plaques that could grow in the presence of SaPIbov1 and SaPIbov5 in the recipient strain, respectively, were sequenced. As expected, the sequenced *dut* genes in these evolved strains contained mutations leading to truncations or changed amino acids in conserved motifs (Appendix 6).

These experiments confirm that expression of the phage-encoded dimeric Dut by the phage is responsible for the induction of SaPIbov1 and SaPIbov5, with *dut* knock outs unable to induce the island. Furthermore, they establish that both SaPIbov1 and SaPIbov5 interfere with the transfer of the inducing phage and that *dut* mutants avoid both SaPI transfer and SaPI interference. The results also establish that, although SaPIbov5 can be packaged by both *pac* and *cos* phages, a *cos* packaging phage such as Φ DI cannot transfer the *pac* island, SaPIbov1. These results further establish the phage-encoded dimeric Duts as the SaPI inducers of SaPIbov1 and SaPIbov5.

3.2.5 SaPI induction is triggered by a direct interaction between the dimeric Dut and the Stl repressor

Having established a subset of the phage-encoded dimeric Duts as inducers of SaPI_{bov1} and SaPI_{bov5}, the method by which these Duts induced the islands was yet to be elucidated. The phage-encoded trimeric Duts have been shown to induce SaPI_{bov1} by a direct interaction with the Stl repressor, which removes the Stl repression of the intergenic region between the *stl* and *str*, allowing transcription of the *str*, *xis* and the downstream genes^{304,305}. A similar mechanism has been suggested for the Φ NM1-encoded dimeric Dut³⁵⁹.

To confirm whether this was the case for all tested phage-encoded dimeric Duts with SaPI-inducing capacity, the plasmid pJP674 was utilised. This plasmid carries a β -lactamase reporter gene fused to *xis*, downstream of the encoded Stl-repressed *str* promoter and *str* (Fig 3.10). By introducing this plasmid into strains expressing the different Dut clones (pCN51), the consequences of Dut expression with regard to gene expression of the Stl-repressed genes could be analysed.

Figure 3.10 shows the levels of β -lactamase expression in the presence or absence of an inducing concentration of CdCl₂ (5 μ M) for the different phage-encoded dimeric Duts. The β -lactamase expression levels can be interpreted as a surrogate for the expression of the Stl-repressed *str* and *xis* genes, following derepression of Stl from the intergenic region. As expected, the SaPI-inducing dimeric Duts (Φ NM1, Φ DI, Φ O11 and Φ O46 Duts) show a significant increase in β -lactamase levels following induction with CdCl₂, whilst the non-inducing Duts (Φ DII and Φ 55 Duts) show no difference between non-induced and induced β -lactamase levels.

Whilst this experiment does not show the method of removal of Stl from the intergenic region, it does suggest that the dimeric Duts capable of SaPI mobilization induce the SaPI cycle by releasing transcription of Stl-repressed genes. In this the phage-encoded dimeric Duts show parallels to the unrelated trimeric Dut proteins.

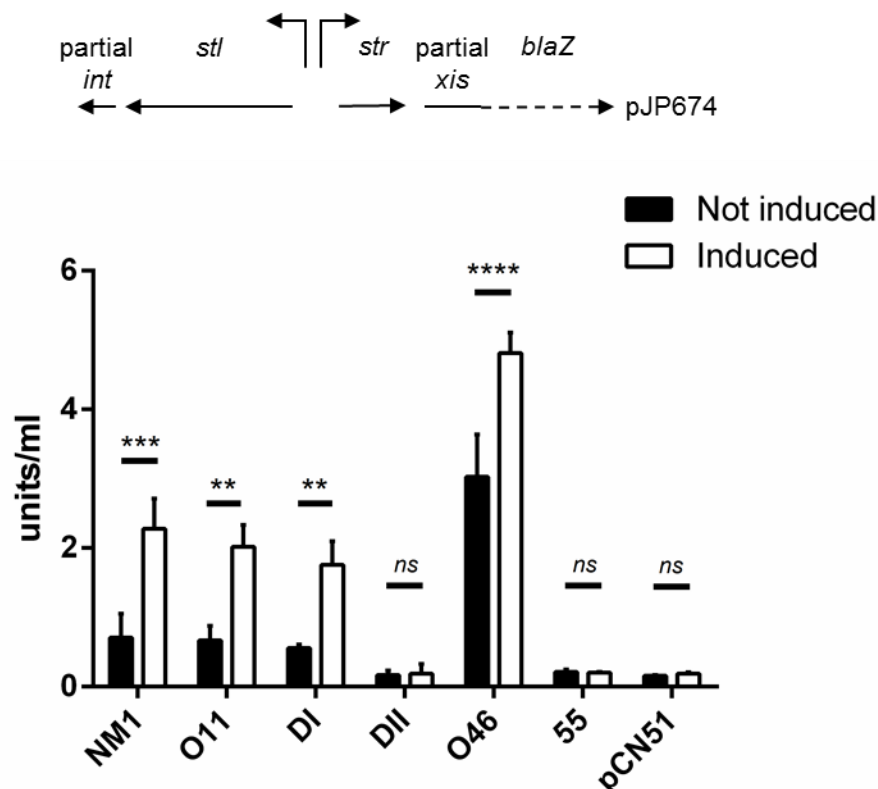


Figure 3.10 Protein expression of Dut variants leading to the release of Stl repression of downstream genes. The diagram represents a schematic of a *blaZ* transcriptional fusion generated in pJP674. β -lactamase assays were performed on strains containing pJP674 together with pCN51-derived plasmids expressing the Φ NM1, Φ O11, Φ DI, Φ DII, Φ O46, and Φ 55 *dut* genes or the empty pCN51 control (JP15105). Samples were taken after 5 hours in the absence or following induction with 5 μ M CdCl_2 . All data is the result of three independent experiments. Error bars represent standard error of the mean. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences within rows. Adjusted *p* values were as follows; Φ NM1 = 0.0003***, Φ O11 = 0.0021**, Φ DI = 0.0075**, Φ DII = >0.9999^{ns}, Φ O46 = <0.0001****, Φ 55 = >0.9999^{ns}, pCN51 = >0.9999^{ns}.

To investigate whether the phage-encoded dimeric Duts directly bind to the SaPI_{bov1} Stl repressor, two SaPI-inducing Duts were chosen as representatives; the Φ NM1 Dut and Φ O11 Dut. These were cloned, untagged into an expression vector together with a His₆-tagged-Stl_{SaPI_{bov1}}, which was in a separate reading frame and had a separate ribosomal binding site (RBS). Controls in which the His-tags were removed were also constructed. Affinity chromatography was then used to try and co-purify the Stl and Duts, with the protein peaks run out in an SDS-PAGE gel (Fig. 3.11). Co-purifications of both Stl_{SaPI_{bov1}} with Dut _{Φ NM1} and Stl_{SaPI_{bov1}} with Dut _{Φ O11} were possible; indicating that in both cases the Stl and the SaPI-inducing dimeric Duts bind to one another. The identity of each of these bands was confirmed by amino acid sequencing and mass spectrometry. These results confirmed a direct interaction between the Stl repressor of SaPI_{bov1} and the phage-encoded dimeric Duts of Φ NM1 and Φ O11.

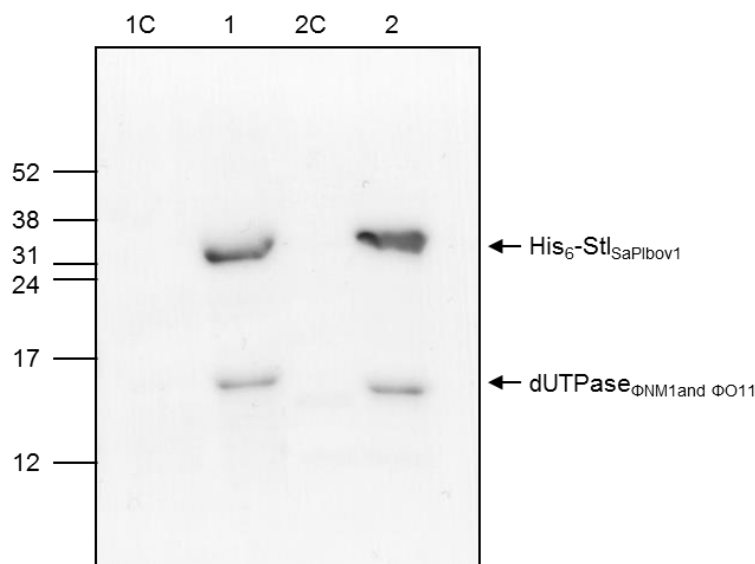


Figure 3.11 Affinity chromatography of the Φ NM1 and Φ O11 Duts for His-tagged SaPIbov1 Stl. Strains were induced with 10mM isopropyl- β -d-thiogalactoside (IPTG) and samples taken at 3 hours. Cells were disrupted and expressing proteins were applied to a Ni^{2+} column and eluted. Lane 1, elution fraction for His-Stl_{SaPIbov1} and Dut _{Φ NM1} (JP). Lane 1C, corresponding elution fraction for the control Stl_{SaPIbov1} and Dut _{Φ NM1} (JP, no His-tag). Lane 2, elution fraction for His-Stl_{SaPIbov1} and Dut _{Φ O11} (JP14832). Lane 2C, corresponding elution fraction for the control Stl_{SaPIbov1} and Dut _{Φ O11} (JP14833, no His-tag). Proteins were confirmed by Mass Spectrometry analysis by Rafael Ciges at the proteomics facility of SCSIE University of Valencia as described in section 2.5.3.

These results establish both that a direct interaction between the dimeric Duts and the Stl_{SaPIbov1} does occur and also imply that Dut expression leads to the derepression of downstream genes. Taken together, this suggests that the SaPI-inducing phage-encoded dimeric Duts achieve this induction by interacting with the Stl repressor to lift the repression of the *str*, *xis*, and downstream genes. This matches with the previous observations for the Φ NM1 dimeric Dut³⁵⁹ and with the established mechanism by which the phage-encoded trimeric Duts induce SaPIbov1^{304,305}. Again, even with their divergent sequence and structure, these results indicate a parallelism between the mechanism of SaPI-induction by the dimeric and trimeric Duts.

3.3 Discussion

The identification that some phages encode dimeric and not trimeric Duts and that the Φ NM1-encoded dimeric Dut was capable of inducing SaPIbov1, opened up the possibility that these phage-encoded dimeric Duts could represent a new subset of SaPIbov1 inducers. The molecular mechanism by which the trimeric Duts induce SaPIbov1 has been well characterised, involving a direct interaction

between the trimeric Duts and SaPI_{bov1} Stl repressor. This was, therefore, a potential template for establishing the method by which the dimeric Duts induce SaPI_{bov1}. However, given the complete difference in both sequence and structure of the known dimeric and trimeric Duts, the chances of any parallelism in SaPI induction mechanism between the two Dut types seemed remote.

Initially this study confirmed that the phage-encoded dimeric and trimeric Duts are completely divergent in both sequence and structure. This matches previous studies into the Dut types, which have established the trimeric Duts as the most numerous type 1 dUTPases consisting of mainly β -sheets, while the dimeric Duts have been labelled the type 2 dUTPases and are made up of α -helices.

Past phylogenetic classification of the type 2 dimeric Duts has separated these Duts into three branches, including one containing the predicted Duts in various Gram-positive bacteria and their phages³⁵⁸. As yet no structure has been solved for any of the Duts classified on this branch, although this will be addressed in a later chapter. Indeed, so far structures have only been solved for Duts from organisms on one branch: the branch for the Duts of trypanosomatids, *C. jejuni*, *M. degradans* and *C. psychrerythraea*³⁵⁸. The predicted structures of the different phage-encoded dimeric Duts shown here indicate that the Duts encoded by the staphylococcal phages represent a simplified version of dimeric Dut when compared to those with solved structures. The structures are predicted to maintain the conserved core of the dimeric Duts, whilst displaying a smaller size particularly in the C-terminal. The implications of this are discussed in a later chapter, in which the phage-encoded dimeric Dut structure is described in more detail (Section 4.2.1).

Previous studies into SaPI_{bov1} have focused on the trimeric Duts, with only Φ NM1 identified as a SaPI inducer encoding for a dimeric Dut. This study aimed to establish the prevalence of the dimeric Duts in staphylococcal phage populations and characterise the existence of any groups within this subset of the dimeric Duts. Investigations into the sequences of the staphylococcal phage dimeric Duts identified 9 different groups. On further analysis, it became apparent that these sub-groups were due to a central divergent region contributing to allelic variation. The phage-encoded dimeric Duts displayed highly conserved catalytic motifs, specific to the dimerics, together with an

internal region highly variable in sequence and length. This mirrored previous findings for the trimeric Duts; although each type of Dut encodes a specific set of conserved motifs, both contain a central divergent region flanked at either end by these conserved regions.

The surprising parallelisms shown between the phage-encoded dimeric and trimeric Duts extended to the impact of allelic variation on SaPI induction. Both types of Dut display an unconserved central region which seems to impact on the capacity of these Duts to induce SaPI_{bov1}. Furthermore, both induce SaPI_{bov1} mobilization by interacting with the Stl and removing this repressor from its DNA binding site. These findings could imply that dimeric and trimeric Duts present a similar mechanism of interaction with the Stl repressor. However, previous work indicates that the trimeric Duts interact with Stl as a trimer³⁰⁴, while it has been proposed that binding to the Stl repressor causes dimer disruption in the dimeric Φ NM1 Dut³⁵⁹. As such, there must be some particularities in the molecular mechanism of Stl recognition and binding that are specific for each type of Dut.

Uniquely, some phage-encoded dimeric Duts do not induce the SaPIs and these cluster more distantly from those that do, implying a divergence in evolution leading to two sub-groups: the SaPI-inducers and the non-inducers. It is intriguing that the SaPI-inducing group of dimeric Duts persist. As shown in this study, SaPI induction has a detrimental effect on phage transfer. Surely, if the phage could avoid this by encoding an alternative Dut, that would be preferable? Indeed, the need for the *S. aureus* phages to encode either a dimeric or trimeric Dut is unclear. All phages encode one or the other, implying the Duts are important for the phage cycle, but they can be lost in laboratory conditions with no obvious impairment on the phage²⁸⁰.

One hypothesis could be that the phage-encoded Duts are important for regulatory functions; interacting with another phage or bacterial protein to provide a functional advantage. Indeed, this has been previously shown for another phage-encoded SaPI-inducer, Sri. Sri is the phage-encoded inducer of SaPI1, but also acts as a regulator for the cellular helicase loader, DnaI^{313,395}. This regulatory function provides the phage with a functional advantage. The SaPIs could then take advantage of this, developing a Stl repressor mimicking the structure of one of these advantage carrying cellular partners. Molecular

mimicry is a well-known tactic of pathogens, particularly viruses, used to exploit cellular processes³⁹⁶. In this hypothetical, the SaPI acts as a phage pathogen, with the Stl repressor mimicking the as yet unknown partner for the phage-encoded Duts.

This leads to a classical evolutionary paradigm; the phage-encoded inducer protein interaction with the Stl repressor is highly detrimental to the phage, but this protein is too useful to be lost and so evolves rapidly, developing variable alleles to escape the interaction. In the case of the phage-encoded Duts, this has led to the unconserved central domain (motif VI) in both the dimeric and trimeric Duts, with the added option of substituting between the analogous dimeric and trimeric Dut types. This would explain the existence of the variety of dimeric and trimeric Duts, many with different affinities for binding the Stl repressor^{281,305}.

All this implies that the non-inducing phage-encoded dimeric Duts have been most successful at evolving to avoid an interaction with the Stl repressor, so why are these not ubiquitous? Because in this premise the Stl is mimicking a cellular protein targeted by the Dut to give a selective advantage for the phage, a reduction in affinity for the Stl could be taken as a reduction in affinity for the true cellular target. This competing affinity makes it more difficult for the phages to evade SaPI induction, the assumption being that the phages encoding these non-inducing dimeric Duts no longer gain the functional advantages that come with a greater affinity for the cellular target. Thus, the multiple Dut types encoded by the *S. aureus* phages and the allelic variation within these types represent the equilibrium between maintaining a selective advantage and avoiding SaPI induction. The ecological niche or the presence of SaPIs in a bacterial population would impact which route would gain more of an advantage for the phage.

3.4 Conclusions and future work

SaPIs were the first PICIs to be identified and within this group the SaPI_{bov1} is now the best characterised. The interaction that occurs between the phage and the SaPI_{bov1} has been elucidated to a molecular level, with the interaction between the phage-encoded trimeric Dut and the Stl_{SaPI_{bov1}} repressor well

understood. This chapter expands the group of SaPI_{bov1} and SaPI_{bov5} inducers to include a subset of the dimeric Duts and finds parallelisms between the dimeric and trimeric Dut inducers, in spite of their sequence and structural differences. Furthermore, it also identifies that the dimeric Duts depress the SaPIs through interaction with Stl, again with some similarities to the trimeric Duts. However, the structural differences between the dimeric and trimeric Duts mean that the exact molecular interaction of the dimeric Duts with the Stl repressor cannot be the same as for the trimeric. Identifying the mechanism of interaction between the dimeric Duts and the Stl repressor is necessary to understand how the Stl repressor interacts with two very different proteins, and this will be addressed in later chapters. It would also be beneficial to finish characterising the 9 phage-encoded dimeric Dut groups identified, by obtaining and testing the SaPI_{bov1} and SaPI_{bov5} inducing capabilities of representatives HKU10-03, ΦStauST398-3, and ΦMR25.

Chapter 4 Characterisation of the dimeric Duts interaction with SaPIbov1

Results from this section are included in:

Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer (2017) **Bowring J***, Neamah MM*, Donderis J*, Mir-Sanchis I, Alite C, J Ciges-Tomas R, Maiques E, Mehmedov I, Marina A, Penadés JR. *eLIFE*. DOI: 10.7554/eLife.26487

Convergent evolution involving dimeric and trimeric dUTPases in pathogenicity island mobilization (2017) Donderis J*, **Bowring J***, Maiques E*, Ciges-Tomas R, Alite C, Mehmedov I, Angeles Tormo-Mas M, Penadés JR, Marina A. *PLoS Pathogens*. DOI: 10.1371/journal.ppat.1006581

*Denotes joint first authorship

Disclaimer on the work performed:

The structures referred to in this chapter were solved by Dr Alberto Marina's laboratory group and the structural modelling data in this chapter was performed in conjunction with and under the guidance of Dr Alberto Marina.

4.1 Background

Bacteriophages are important MGEs that impact on bacterial evolution^{389,390}. They are inherently associated with the PICIs, which rely on bacteriophage for their induction and packaging^{59,270,283,292,305,325}. In turn, PICIs also impact on bacterial genetics and could affect the bacteriophage genome and evolution^{29,268}. Both PICIs and bacteriophage have been linked with the spread of virulence genes and play an important role in bacterial pathogenicity^{44,57,73,209,246}.

The best characterised bacteriophage-PICI pairing is the induction and transfer of SaPIbov1 by the 80 α phage. SaPIbov1 induction occurs following a direct interaction between the 80 α -encoded trimeric Dut (dUTPase) and the SaPIbov1 master repressor, Stl. The phage-encoded trimeric Duts have now been established as the prototypical inducers of SaPIbov1, and a number of phages encoding different versions of trimeric Dut have been shown to induce this island^{304-306,312,368}. However, a recent study found that the Φ NM1-encoded dimeric Dut also interacts with the Stl repressor of SaPIbov1 to induce the island³⁵⁹. The results in chapter 3 expand upon this, to identify a subset of phage-encoded dimeric Duts capable of inducing SaPIbov1.

The mechanism by which the phage-encoded trimeric Duts interact with the SaPIbov1 Stl repressor has been highlighted in section 1.4.4. In this interaction, the Stl binds to motifs IV and VI and the conserved but flexible motif V moves to stabilise the complex³⁰⁴. Since the motif V in this interaction comes from a different protomer than that of motifs IV and VI, it seems the phage-encoded trimeric Duts involve multiple subunits for binding Stl³⁰⁵. In contrast, it has been suggested that binding the Stl repressor leads to dimer disruption in the Φ NM1-encoded dimeric Dut³⁵⁹. Taken together with the complete divergence in sequence and structure between the dimeric and trimeric Duts, this implies the phage-encoded dimeric Duts must have their own distinct mechanism for binding the SaPIbov1 Stl repressor.

The aim of this study was to identify the mechanism by which the phage-encoded dimeric Duts interact with the SaPIbov1 Stl repressor. Since the dimeric Duts are so far removed in structure from the trimeric Duts, understanding how

the dimeric Duts bind the Stl repressor would develop our understanding of how the Stl repressor protein is capable of interacting with two such unrelated Dut proteins. This, in turn, will build on the understanding of how SaPIs achieve intra- and inter-generic transfer. Furthermore, since it is suspected that the Stl could be mimicking a separate, more beneficial protein partner for the phage-encoded Duts, understanding the mechanism of action by which the dimeric Dut interacts could help identify this ‘true’ partner.

4.2 Results

4.2.1 The structures of the Φ O11 and Φ DI encoded dimeric Duts

The dimeric Dut structures that been solved consist of the bacterial *C. jejuni* Dut and the parasitic *T. cruzi*, *T. brucei*, and *L. major* Duts³⁵⁵⁻³⁵⁸. No phage-encoded dimeric Dut structure has been solved up until this point. In order to gain insight into the molecular mechanism by which the phage-encoded dimeric Duts interact with Stl, the structural characterisation of two different SaPI-inducing phage-encoded dimeric Duts was completed^{310,373}.

The structures of the dimeric Φ O11 and Φ DI Duts in complex with the non-hydrolysable analog of dUTP, α,β -imido-dUTP (dUPNPP), and Mg^{2+} were solved by SAD, both to 2.1Å resolution (Appendix 7). For the structure of the Φ O11-encoded Dut, the crystal structure indicated 2 molecules in an asymmetric unit organised as a homodimer, whilst the Φ DI-encoded Dut crystal structure displayed 4 molecules in the asymmetric unit organised as two independent homodimers with identical conformation (Fig 4.1). The active centre of each protomer is occupied by one dUPNPP molecule and two Mg^{2+} atoms.

Both the Φ O11 and Φ DI Duts display a ‘compact’ or ‘reduced’ form when compared to the solved structures of dimeric Dut counterparts found in other organisms, which consist of 10-13 α -helices³⁵⁵⁻³⁵⁸. The Φ O11 Dut is made up of just seven α -helices per protomer (Fig 4.1A, right-hand panels; α 1, residues 7-23; α 2, 29-47; α 3 61-82; α 4, 86-98; α 5, 104-108; α 6, 110-121 and α 7, 127-141), whilst the Φ DI Dut consists of eight α -helices per protomer (Fig 4.1A, left-hand panel and Fig 4.1C; α 1, residues 7-23; α 2, 29-47; α 3 61-82; α 4, 86-98; α 5, 104-108; α 6, 110-121; α 7, 127-141 and α 8, 144-163). Still, both Φ O11 and Φ DI Dut

protomers present the characteristic dimeric Dut structural core that makes up the active site for the nucleotide ($\alpha 1$ - $\alpha 3$ and $\alpha 7$ in $\Phi O11$, $\alpha 1$ - $\alpha 3$ and $\alpha 8$ in ΦDI), which is consistent with the active site of the other, larger dimeric Duts³⁵⁵⁻³⁵⁷.

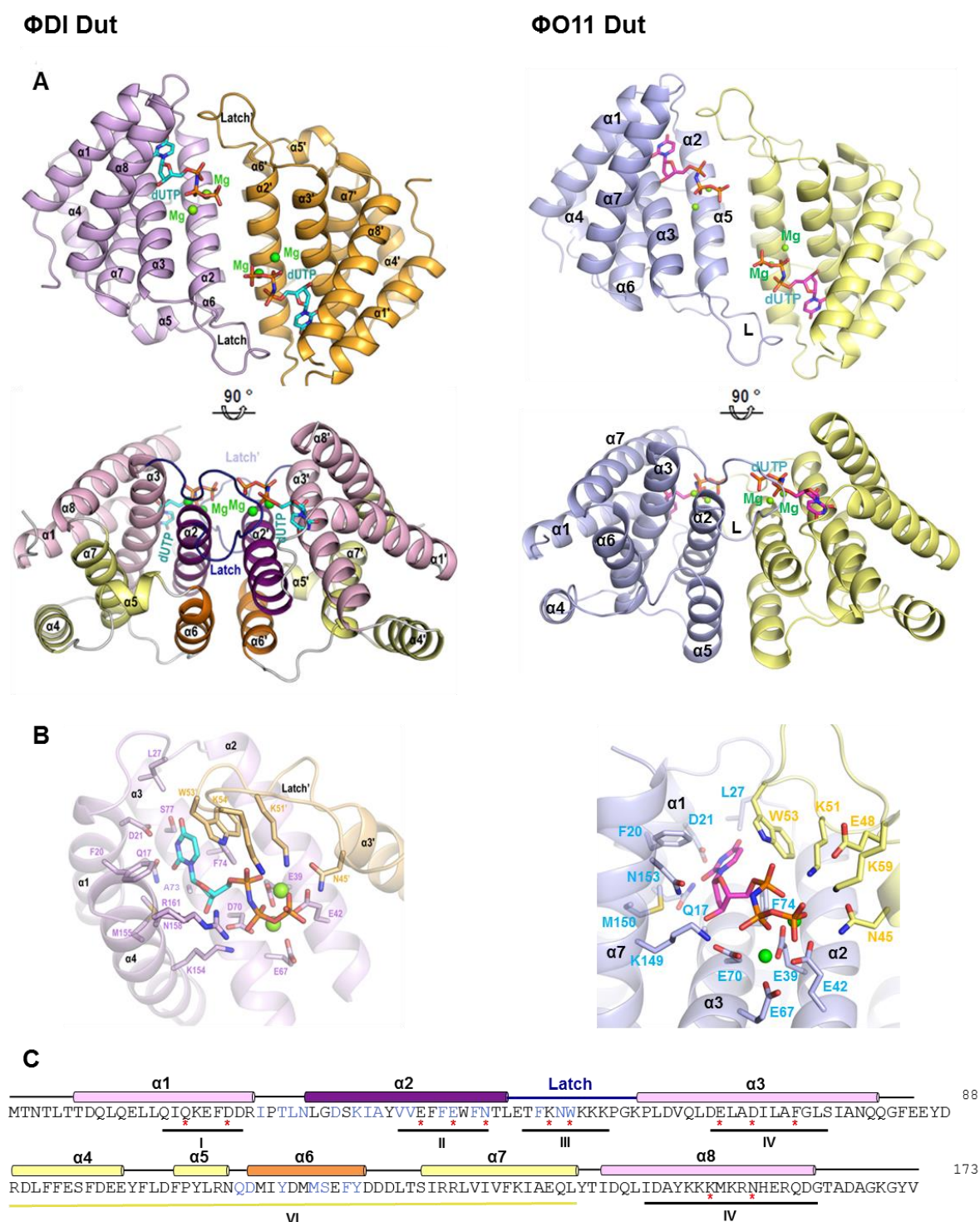


Figure 4.1 Structures of the ΦDI and $\Phi O11$ encoded dimeric Duts. **A** Cartoon representations of the ΦDI Dut dimer (left, protomer in pink and orange) and the $\Phi O11$ Dut dimer (right, protomer in blue and yellow). The secondary structural elements are numbered and labelled in order from N to C terminus. For the ΦDI dimer (left) the ' after the helix label indicates elements from the second protomer. One molecule of dUMPNNP (dUTP analog, labelled as dUTP) and two Mg ions, represented in stick and sphere respectively, occupy the active centre of each protomer. Two

orthogonal views of the dimers are shown. For the Φ DI Dut orthogonal view (left, bottom) the conserved and variable (motif VI) regions are coloured in pink and yellow, respectively. The structural elements exploited to dimerize (helices $\alpha 2$ and $\alpha 6$, and the latch) are highlighted in dark tones. **L**, Latch. **B** Close-up views of the Φ DI Dut (left) and Φ O11 Dut (right) active sites. The substrate dUPNPP is represented in stick with carbon atoms in cyan (Φ DI) or magenta (Φ O11). The residues interacting with the substrate are labelled and represented in stick with carbon atoms coloured according to the protomer to which they correspond. Nitrogen, oxygen, phosphorus atoms are coloured in dark blue, red and orange, respectively. The Mg ions are represented as green spheres. **C** Sequence of the Φ DI Dut, with residues that interact with the substrate highlighted with red stars and residues that interact across the dimer interface in blue text. The locations of the five dimeric Dut conserved motifs and the unconserved motif VI are indicated. Structural elements are shown above the sequence coloured as panel A (Φ DI left, lower).

Again, similarly to the established dimeric Dut structures, the active sites of the Φ O11 and Φ DI dimers are oriented to the same molecule face. This orientation forms a long channel that accommodates two dUPNPP molecules. The differences between the larger bacterial and parasitic dimeric Duts and the more compact phage-encoded dimeric Duts are mainly localised to the C-terminal and the unconserved central region, named motif VI to mimic the trimeric Duts. The C-terminals of the phage dimeric Duts lack the additional helices found in the other, larger dimeric Duts. The divergent motif VI is variable in size and sequence between the phage-encoded dimeric Duts. Other dimeric Duts also encode a divergent motif VI, and these again differ from one another in size and sequence. Placed on the opposite molecule face to the active sites, the motif VI encompasses three helices in the Φ O11 Dut ($\alpha 4$ -6, residues 83-138) and four helices in the Φ DI Dut ($\alpha 4$ -7, residues 86-141).

Previous work into the structural conformations of the bacterial and parasitic dimeric Duts identified conformational movements, allowing the protomers to be divided into a rigid and flexible domain. The rigid domain is linked to dimerization, whilst the flexible mobile domain undergoes conformational changes from the “open” ligand-free conformation to the ligand-bound or “closed” conformation, as modulated by the substrate^{356,357}. In the Φ DI-encoded dimeric Dut the shortened C-terminal region means that the mobile domain is reduced to helices $\alpha 1$ and $\alpha 8$ ($\alpha 1$ and $\alpha 7$ in Φ O11 Dut). The Φ DI Dut rigid domain contains helices $\alpha 2$ and $\alpha 6$ ($\alpha 2$ and $\alpha 5$ in Φ O11 Dut), which interact with the reciprocal helix in the opposite protomer to mediate dimerization (Fig 4.1A, see left, lower panel colour-coding). This interaction is further stabilised by the connecting loops of helices $\alpha 1$ - $\alpha 2$ and $\alpha 2$ - $\alpha 3$. The connecting $\alpha 2$ - $\alpha 3$ loop has been labelled the ‘latch’ (or L) and contains catalytic residues which project on to the neighbouring protomer active site (Fig 4.1B).

The truncation of the C-terminal in the two phage-encoded Duts means that several residues observed interacting with the dUTP phosphates in the other, longer dimeric Duts are not present in the phage Duts. Instead, some residues form novel interactions specific to the *S. aureus* phagic Duts that are not found in the previously defined conserved motifs. In the phage-encoded Duts the Asn45 (N45) or Lys59 (K59) of the neighbouring protomer interacts with the dUTP phosphate, an interaction not seen in the other dimeric Dut crystallography studies (Fig 4.1B). The involvement of the neighbouring protomer residues in the dUTP interaction is enabled by the protomer-protomer conformation of the phage-encoded dimeric Duts, which display a more parallel orientation of the dimerization mediating helices (Φ DI, α 2 and α 6; Φ O11, α 2 and α 5). This reorientation also reduces the dimerization interface in comparison to the bacterial and parasitic encoded Duts ($\sim 1240 \text{ \AA}^2$ to $\sim 1600\text{-}1900 \text{ \AA}^2$, respectively).

One more phage-encoded dimeric Dut particularity is seen in the sugar moiety binding site. Whilst in the previously characterised dimeric Dut structures the deoxyribose was inserted between the aromatic rings of a conserved His-Phe site in motif IV, in the phagic dimeric Duts the His is substituted to an Ala (Φ DI Dut A73, Fig 4.1B left panel). This expands the sugar binding pocket, allowing for the accommodation of other nucleotides, thus explaining the dCTPase and dITPase activity shown by some phage-encoded dimeric Duts³⁷³. This matches previous findings that other members of the all- α NTP pyrophosphohydrolases superfamily, to which the dimeric Duts belong, can hydrolyse alternative deoxynucleotides³⁷².

Structural predictions of the phage-encoded dimeric Duts and their comparison with the 80 α -encoded trimeric Dut indicated that the dimeric Duts did not contain any of the trimeric motifs involved in Stl interaction (motifs IV-VI, Sections 3.2.1-3.2.2). The structural characterisation of the Φ O11 and Φ DI dimeric Duts confirms this. However, there was the potential that the Stl recognition site was generated spatially by the disposition of specific residues from the motifs IV-VI, and that these residues could be mirrored in the dimeric Duts. To eliminate this possibility the active sites of both the trimeric 80 α and the dimeric Φ O11 Duts were superimposed for spatial comparison, as the trimeric motifs IV-VI are present in this region of the trimeric Duts (Fig 4.2). No similarities were seen between the different dimeric and trimeric Dut active

sites, with the disposition of the residues surrounding the nucleotides also completely different. This implies that the SaPIbov1 Stl repressor uses different domains or ways to recognise two different sites for Dut interaction, one for the dimers and another for the trimers.

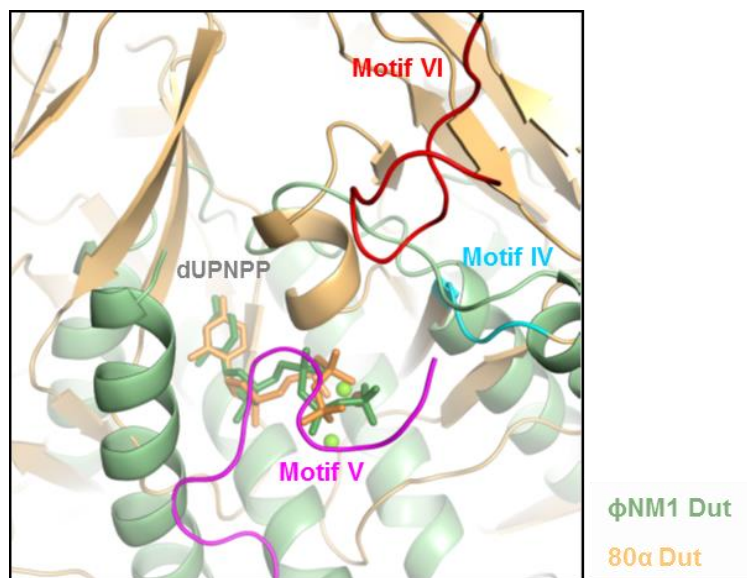


Figure 4.2 The active sites of the phage encoded dimeric and trimeric Duts do not share any overlapping residues. Superimposition of dUPNPP molecules in the active sites of the dimeric $\Phi O11$ Dut (green tones) and the trimeric 80α Dut (orange tones). The bound nucleotide molecules acquire different conformations (stick representation), including the disposition of the Mg ions (sphere representation), and the spatial arrangement of the structural elements confirming each active centre is essentially different. No structural element equivalent to the Stl binding motifs of 80α (labelled and coloured in cyan, magenta and red for motif IV, V and VI, respectively) is observed in $\Phi O11$.

4.2.2 The role of the dimeric Duts unconserved extra domain (motif VI) in SaPIbov1 induction

Given that the interaction of the SaPIbov1 Stl repressor was likely to occur by a different mechanism for the dimeric Duts than that seen in the trimeric Duts, this opened up the possibility that the interaction could occur away from the active centre. Taking into account that some dimeric Duts have escaped Stl interaction and that the most obvious difference between these non-inducers and the inducing SaPIs is the divergent motif VI, one hypothesis was that the motif VI could be involved in the SaPI-binding mechanism.

To test this hypothesis, various chimeric mutants were made between the SaPI-inducing ΦDI -encoded dimeric Dut and the non-inducing $\Phi 55$ Dut. These

consisted of exchanges between the different helices making up the motif VI, as displayed in Fig 4.3A. The helices that compose motif VI, $\alpha 4$ - $\alpha 7$, were exchanged in the following combinations: $\alpha 4$ only, $\alpha 4$ - $\alpha 5$ - $\alpha 6$, $\alpha 4$ - $\alpha 5$ - $\alpha 6$ - $\alpha 7$, $\alpha 4$ - $\alpha 5$, $\alpha 5$ - $\alpha 6$, and $\alpha 7$ only. These exchanges were performed in both directions, meaning the Φ DI Dut was mutated to contain the $\Phi 55$ helices and vice versa.

The chimeric Duts together with the wild-type Φ DI and $\Phi 55$ Duts were cloned into the pCN51 expression vector under a *Pcad* promoter. The cloned Duts were introduced into a SaPIbov1 positive strain and the induction of this island following Dut expression was tested by Southern blot (Fig 4.3). The different exchanges between the helices within motif VI did not affect the SaPI-inducing capacity of the Duts; the Φ DI Dut still induced SaPIbov1, whilst the $\Phi 55$ Dut did not. However, the different Φ DI Dut mutants did seem to be more effective SaPI-inducers, with SaPIbov1 induction occurring without any CdCl_2 present to promote Dut expression. This implies that the minimal expression of the Φ DI Dut mutants from basal or leaky pCN51 replication was sufficient to induce the island, whilst this was not the case for the Φ DI wild-type Dut which only induced SaPIbov1 in the presence of CdCl_2 .

These results imply that the divergent motif VI may have some role in the affinity of the Dut for the Stl. However, they also indicate that the motif VI is not responsible for SaPI-inducing capabilities. Therefore, the mechanism by which the phage-encoded dimeric Duts bind the Stl repressor must involve additional regions of the protein.

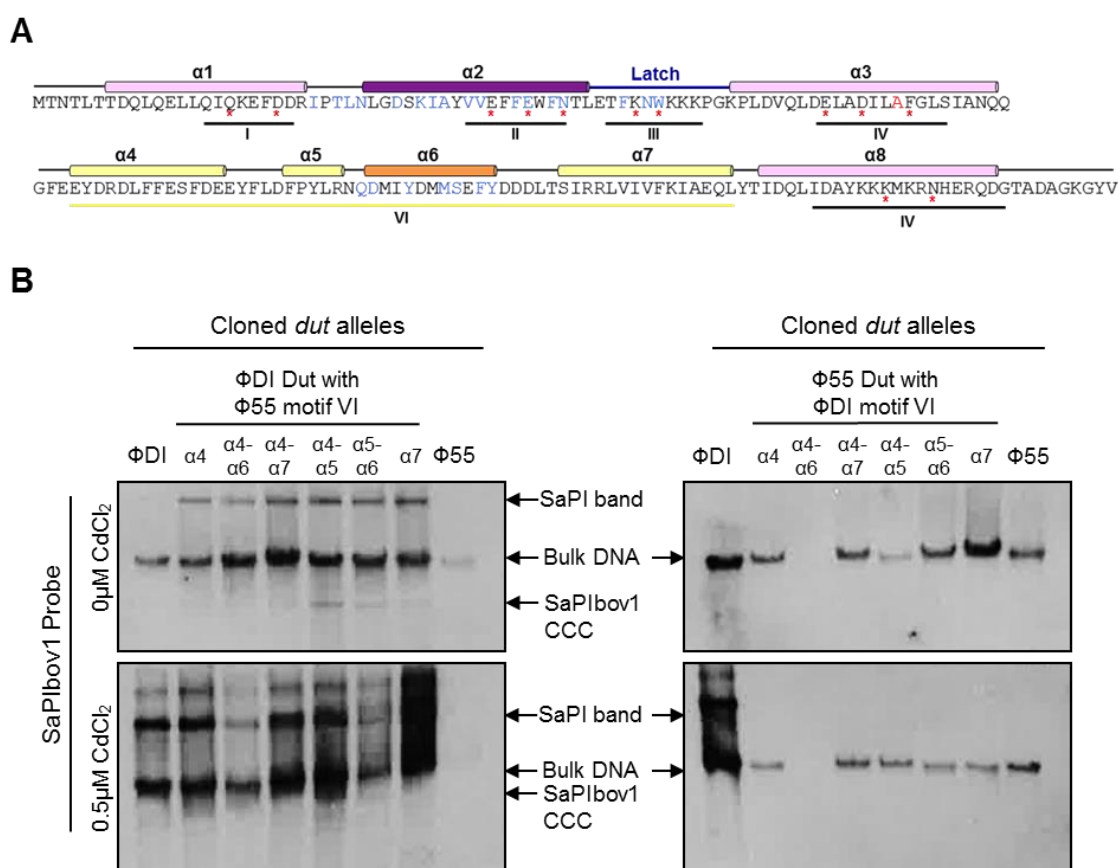


Figure 4.3 The divergent motif VI is not responsible for the SaPIbov1 inducing capabilities of the dimeric Duts. **A** Sequence of the Φ DI Dut, with residues that interact with the substrate highlighted with red stars and residues that interact across the dimer interface in blue text. The locations of the five dimeric Dut conserved motifs and the unconserved motif VI are indicated. Structural elements are shown above the sequence with the conserved and variable (motif VI) regions coloured in pink and yellow tones, respectively. The structural elements exploited to dimerize (helices $\alpha 2$ and $\alpha 6$, and the latch) are highlighted in darker tones. **B** SaPIbov1 excision and replication following the induction of cloned Φ DI and $\Phi 55$ wt *dut* genes and chimeric Φ DI and $\Phi 55$ *dut* genes containing helices from motif VI of the $\Phi 55$ Dut (left-hand panels) or chimeric $\Phi 55$ *dut* genes containing helices from motif VI of the Φ DI Dut (right-hand panels). The helices exchanged in each chimeric mutant are indicated above the blot and their relative position in the Φ DI Dut sequence can be seen in panel A. Strain JP6774 containing SaPIbov1 was complemented with the different pCN51 plasmids expressing the wild-type and mutant dimeric Duts. Samples were isolated 3 hours post-induction with or without 0.5 μ M CdCl_2 and Southern blots were performed using a probe for the SaPIbov1 integrase (Table 2.2). Bulk DNA is labelled as such and includes chromosomal, phage, and replicating SaPI DNA. In these experiments, as no helper phage was present, the excised and replicating SaPI DNA appears as part of the bulk DNA or as CCC molecules, rather than the linear monomers that are seen following helper phage-mediated induction and packaging. The upper band labelled ‘SaPI band’ represents oligomeric forms of the CCC molecules. The middle band represents bulk DNA, whilst the lower band signifies the SaPIbov1 CCC. CCC indicates covalently closed circular SaPI DNA.

4.2.3 The role of the dimeric Duts enzymatic function in SaPIbov1 induction

It has been demonstrated that dUTP inhibits the interaction of the SaPIbov1 Stl repressor with the *S. aureus* phage-encoded trimeric Duts^{304,368}. In the trimeric Duts, the binding of dUTP causes a conformational change in which the

conserved C-terminal P-loop (trimeric motif V) moves to cover the active site once the nucleotide has been accommodated^{305,353}. This blocks the binding site of the Stl, thus blocking the Dut-Stl interaction. Stl-induced dUTPase activity inhibition was also identified in the Φ NM1-encoded dimeric Dut³⁵⁹. Furthermore, Stl has been shown to inhibit or abolish dUTPase activity in the SaPI-inducing Φ O11 and Φ DI Duts, whilst having no effect on the dUTPase activity in the non-inducing Φ 55 and Φ DII Duts³⁷³. Additionally, the presence of dUTP has also been shown to inhibit the binding of Stl to the Φ DI and Φ O11 Duts³⁷³.

The previously characterised dimeric Duts have been shown to change conformation upon binding dUTP, alternating between an 'open' (dUTP-free) and a 'closed' (dUTP-bound) conformation. Based on this parallelism with the trimeric Duts, it was hypothesised that the dUTP binding region of the phage-encoded dimeric Duts could be involved in Stl binding, and that the enzymatic function and induced conformational change could therefore play a role in the interaction with Stl.

To test this possibility Φ DI Dut mutants were designed that would be unable to bind the dUTP nucleotide. Based on the Φ DI Dut structure, the residue Ala73 was chosen and mutated to the hydrophobic residue Leu (A73L). As discussed in Section 4.2.1, this Ala73 residue occupies the bottom of the nucleotide binding pocket and is responsible for deoxyribose discrimination. As such, substituting the simple Ala for a bulky hydrophobic residue (Leu) should eliminate the nucleotide access and make the Dut enzymatically non-functional. Additional mutations aiming to abolish the nucleotide binding ability of the Φ DI Dut included mutating the conserved residue Gln17, which interacts with the dUTP substrate, to Met and removing the helix α 8, which is involved in the active centre, by the truncation mutation Lys152.

The wild-type Φ DI Dut, together with the Q17M, A73L and K152* mutants were cloned into the pCN51 vector under a *Pcad* promoter. The wild-type Φ 55 Dut was included as a known non-inducer. These cloned Duts were introduced into a SaPIbov1 positive strain and the induction of this island following Dut induction was tested using a Southern blot (Fig 4.4A). All the Φ DI Dut mutations had some impact on the Duts SaPIbov1 inducing ability, with SaPIbov1 induction reduced in the Q17M mutant and completely abolished in the A73L and K152* mutants.

As a single point mutation that completely abolished the SaPIbov1 inducing capability of the Φ DI Dut, the Φ DI Dut A73L mutant was taken forward for further experimentation. An additional Southern blot using the cloned wild-type and A73L Φ DI Duts was performed, including the empty vector as a control (Fig 4.4B). Again, the A73L mutant was not able to induce SaPIbov1, confirming the previous result.

To analyse the consequences of Φ DI A73L Dut expression with regard to gene expression of the Stl-repressed genes, the plasmid pJP674 was introduced into strains expressing the different Φ DI Dut clones (pCN51, wt and A73L). This plasmid carries a β -lactamase reporter gene fused to *xis*, downstream of the encoded Stl-repressed *str* promoter, *str*, and *stl* (Fig 4.4C). The levels of β -lactamase expression with or without Dut overexpression were interpreted as a substitute for the expression of the Stl-repressed *str* and *xis* genes. As previously seen, the CdCl₂-induced Φ DI wild-type Dut displayed a significant increase in β -lactamase levels compared to non-induced levels. However, there was no significant difference between the β -lactamase levels of the non-induced and induced Φ DI A73L Dut, implying the mutant could not trigger expression of the downstream SaPI genes. Furthermore, there were significant differences between the β -lactamase levels of the induced Φ DI wild-type Dut when compared to the induced Φ DI A73L Dut, supporting the hypothesis that this mutant could not bind Stl to trigger expression of the downstream genes.

To confirm these effects were also seen in the phage, a recombinant phage containing the A73L mutation was created. The Φ DI^{A73LDut} and Φ DI^{WT} phages were introduced into a SaPIbov5 positive strain. As previously mentioned, Φ DI uses the *cos* packaging mechanism and cannot package SaPIbov1. SaPIbov5, however, carries the *cos* sites necessary for packaging by Φ DI and contains an identical Stl to SaPIbov1. The ability of the phages to induce the SaPIbov5 cycle was assessed by Southern blot following SOS response induction of the phages by MC. As with the previous results, Φ DI^{WT} induced the SaPIbov5 as normal, whilst the Φ DI^{A73L} Dut mutant displayed reduced, if not completely abolished, SaPIbov5 induction (Fig 4.4D).

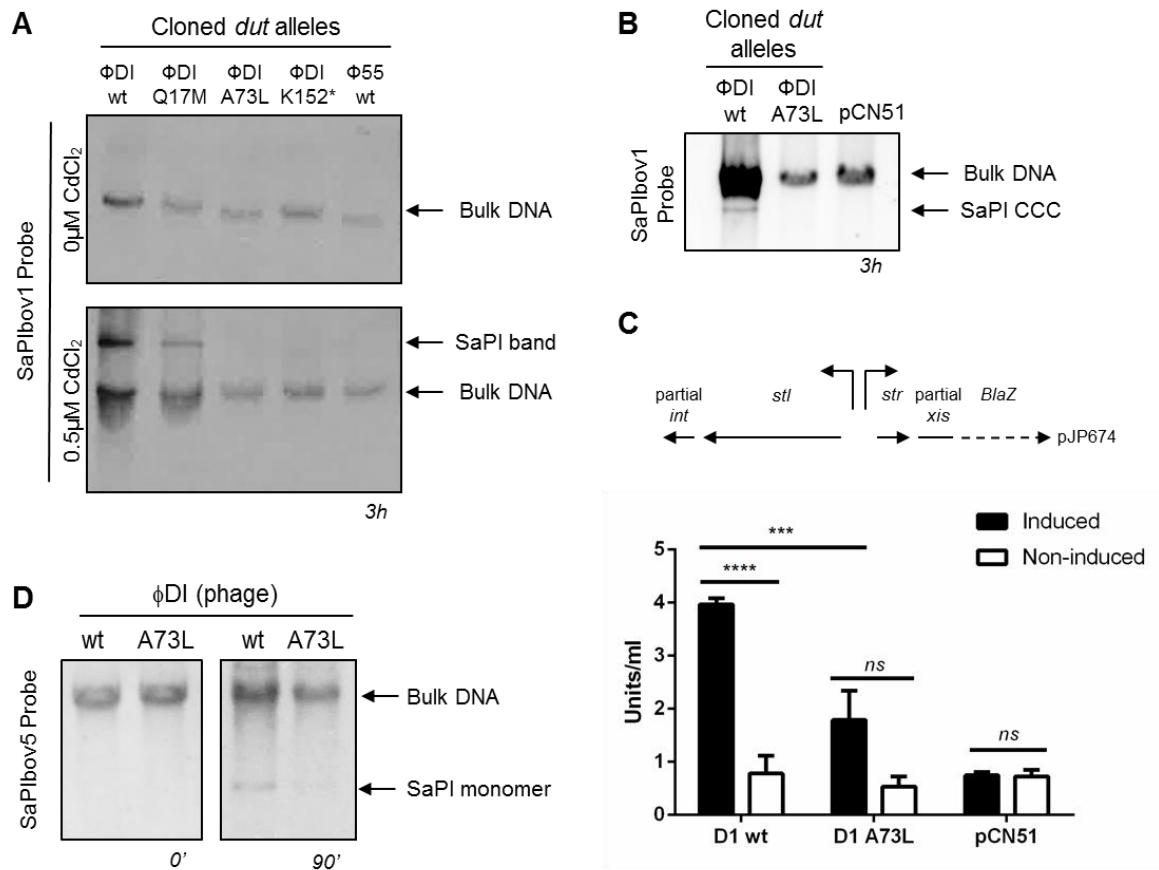


Figure 4.4 The ΦDI A73L Dut mutant is incapable of inducing SaPIbov1. **A** Mutations affecting the highly conserved regions of the ΦDI dimeric Dut involved in the enzymatic function decrease or abolish SaPIbov1 induction. SaPIbov1 excision and replication following induction of cloned ΦDI *dut* genes, either wild-type, point mutants (Q17M or A73L), or truncation mutant (K152*). In this case a known SaPIbov1 non-inducer, the cloned Φ55 Dut, was used as a control. Strain JP6774 containing SaPIbov1 was complemented with the different pCN51 plasmids expressing the wild-type and mutant dimeric Duts. Samples were isolated 3 hours post-induction with or without 0.5 μM CdCl₂ and Southern blots were performed using a probe for the SaPIbov1 integrase (Table 2.2). Bulk DNA is labelled as such and includes chromosomal, phage, and replicating SaPI DNA. In these experiments, as no helper phage was present, the excised and replicating SaPI DNA appears as part of the bulk DNA or as CCC molecules, rather than the linear monomers that are seen following helper phage-mediated induction and packaging. The upper band labelled 'SaPI band' represents oligomeric forms of the CCC molecules. The lower band represents bulk DNA. **B** SaPIbov1 excision and replication following induction of cloned wild-type and A73L ΦDI *dut* genes. Strain JP6774 containing SaPIbov1 was complemented with the different pCN51 plasmids expressing the dimeric Duts or the empty pCN51 control. Samples were isolated 3 hours post-induction with 0.5 μM CdCl₂ and Southern blots were performed using a probe for the SaPIbov1 integrase (Table 2.2). The upper band represents bulk DNA, whilst the lower band signifies the SaPI CCC band. CCC indicates covalently closed circular SaPI DNA. **C** The diagram represents a schematic of a *blaZ* transcriptional fusion generated in pJP674. β-lactamase assays were performed on strains containing pJP674 together with pCN51-derived plasmids expressing the ΦDI wild-type and A73L *dut* genes or the empty pCN51 control (JP15105). Samples were taken after 5 hours in the absence or following induction with 5 μM CdCl₂. All data is the result of three independent experiments. Error bars represent standard error of the mean. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences. Adjusted *p* values were as follows; ΦDI wt = <0.0001****, ΦDI A73L = 0.0602^{ns}, pCN51 = >0.9999^{ns}, ΦDI wt^{induced} vs. ΦDI A73L^{induced} = 0.0008***. *ns*, not significant. **D** Southern blots were performed using strains positive for SaPIbov1 and either phage ΦDI^{wt} or phage ΦDI^{A73L}. Samples were isolated at 0' or 90' minutes following induction with mitomycin C. Southern blots were performed using a probe for the SaPIbov5 integrase. The upper band represents bulk DNA, whilst the lower band signifies the SaPI linear monomers released from phage heads.

These results suggest that the enzymatic function of the phage-encoded dimeric Duts is important for SaPI interaction. The A73L mutant prevents the access of dUTP to the active centre binding site, making it enzymatically non-functional. This mutant was unable to induce the SaPIbov1 and SaPIbov5 islands, implying that the active site integrity, hydrolysis of dUTP and associated conformational changes are important for the mechanism of interaction with the Stl repressor.

4.2.4 The role of the dimerization interface of the dimeric Duts

Section 4.2.2 established that, whilst the phage-encoded dimeric Dut divergent motif VI may play a role in the affinity of the Dut for Stl, it did not seem to be responsible for SaPI-induction capability. This suggested that another region within the dimeric Duts is involved in the interaction with Stl.

Previous to the characterisation of the dimeric Φ DI and Φ O11 Dut structures, a protein sequence alignment of two inducing dimeric Duts (Φ NM1 and Φ DI) and two non-inducers (Φ DII and Φ 55) was performed to identify any overlooked areas of divergence. This revealed an additional area of the Dut proteins containing conserved amino acid residue differences between the inducers and non-inducers, separate to the divergent motif VI previously analysed (Fig 4.5A). Based on this, a chimeric mutant was designed in which the area containing these residues was exchanged between the inducing Φ DI Dut and the non-inducing Φ 55 Dut. The Φ DI and Φ 55 wild-type Duts and chimeras exchanging the Ile24-Gln82 region of the Dut (Φ DI Dut ^{Φ 55I24-Q82} and Φ 55 Dut ^{Φ DI24-Q82}) were cloned into the pCN51 expression vector under a *Pcad* promoter. The cloned Duts were introduced into a SaPIbov1 positive strain and the induction of these islands following Dut expression was tested by Southern blot (Fig 4.5B).

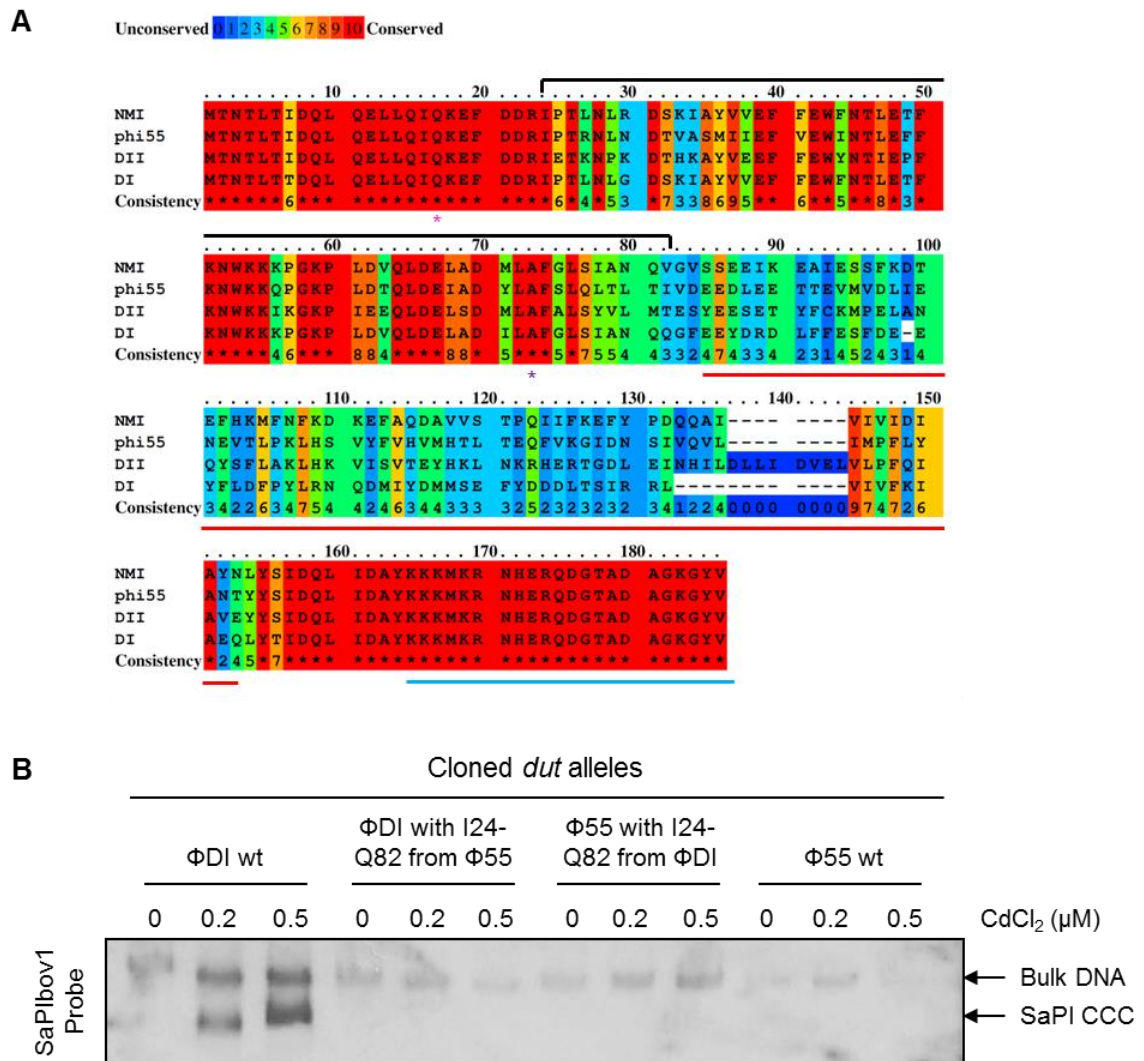


Figure 4.5 A second region containing sequence divergence impacts on SaPIbov1 induction capability. **A** Alignment of the dimeric Duts from *Staphylococcus aureus* phages ΦNM1, ΦDI, ΦDII, and Φ55 are shown, representing two inducers of SaPIbov1 and two non-inducers. The region identified as having conserved differences between the inducers and non-inducers, separate to those previously analysed, is indicated by the black bracket. The divergent motif VI that was analysed in the original chimeric mutants (Fig 4.3) is indicated by a red bracket beneath the sequence. The Q17M and A73L point mutants are indicated by the pink and purple asterisks, respectively. The K152* truncation mutant is indicated by the blue bracket beneath the sequence. Residue colours indicate relative sequence conservation at each position, with red being most conserved and blue being least (adapted from alignment generated by PRALINE)^{284,285}. **B** SaPIbov1 excision and replication following induction of cloned wild-type and chimeric ΦDI and Φ55 *dut* genes. Strain JP6774 containing SaPIbov1 was complemented with the different pCN51 plasmids expressing the wild-type dimeric Duts, chimeric Duts in which the highlighted region of interest had been exchanged, or the empty pCN51 control. Samples were isolated 3 hours post-induction with or without CdCl₂ and Southern blots were performed using a probe for the SaPIbov1 integrase (Table 2.2). Bulk DNA is labelled as such and includes chromosomal, phage, and replicating SaPI DNA. In these experiments, as no helper phage was present, the excised and replicating SaPI DNA appears as part of the bulk DNA or as CCC molecules, rather than the linear monomers that are seen following helper phage-mediated induction and packaging. The upper band represents bulk DNA, whilst the lower band signifies the SaPI CCC band. CCC indicates covalently closed circular SaPI DNA.

Interestingly only the Φ DI wild-type dut was capable of inducing SaPIbov1, with the chimeric Φ DI Dut showing no SaPIbov1 induction. As previously seen, the Φ 55 wild-type was not capable of inducing SaPIbov1 and the chimeric interchange introducing the Φ DI I24-Q82 region was not sufficient to gain induction ability. These results imply one of two situations; either the interchange of the I24-Q82 region creates misfolded proteins that are non-functional and cannot bind Stl, or the region is important for Stl interaction, but not solely responsible, so the interchange causes a loss of Stl interaction for the inducing Dut but is not sufficient for the development of Stl-binding in the non-inducer.

The characterisation of the structures of the Φ DI and Φ O11 phage-encoded dimeric Duts allowed further investigation of this new region of interest. The region exchanged in the I24-Q82 chimeric mutations was identified within the structures and found to make up part of the dimerization interface of helices α 2 and the latch, as well as including helix α 3. Interestingly, previous studies into SaPIbov1 induction by the Φ NM1-encoded Dut had suggested that the binding of Stl disrupts the Φ NM1 Dut dimer³⁵⁹.

The combination of these previous findings and the preliminary data from the I24-Q82 mutants led to the development of a hypothesis; the Dut dimer is disrupted by Stl and the Stl binds to the exposed dimerization interface. An examination of the dimerization interface established that it is made up of helices α 2 and α 6 in the Φ DI and Φ 55 Duts, whilst helices α 2 and α 5 form the interface in the Φ O11 Dut. Figure 4.6A displays a superimposition of the Φ DI, Φ O11 and Φ 55 Duts in three orthogonal views and the dimerization interface is highlighted throughout. The Fig 4.6A right, upper and left, lower panels indicate helices α 2 and α 5/6 to show how they interact with the corresponding helix in the opposite protomer. Close-up views of the α 2 and α 5/6 helices indicate that residue 41 on helix α 2 interacts with the equivalent residue on the neighbouring protomer (Fig 4.6A, boxes). Thus, residue 41 was identified as an optimum candidate for mutation, since a single point mutation here would have a greater impact on the dimerization interface.

To test the hypothesis that the dimerization interface is important for Stl interaction, the residue 41 was mutated in a number of different dimeric Duts: Φ DI^{wt}, Φ DI^{A73L}, Φ DII^{wt}, and Φ 55^{wt} Duts. Residue 41 in the Φ DI and Φ DII Duts is

Phe41, whilst in $\Phi 55$ it is Val41. These amino acids are both non-polar and hydrophobic, so different mutations were designed that were predicted to have varying effects: Ala was chosen to remove the Phe aromatic side chain and encourage Dut monomers without changing the dimerization interface charge; Cys was chosen as it does not affect the charge of the dimerization surface and makes the protein principally a monomer in reductant conditions like the intracellular environment, whilst forming a stable dimer in oxidant conditions; Asp was chosen to encourage Dut monomers and introduce a negative charge to the dimerization interface. Based on these predicted properties the Ala and Cys mutations were expected to increase SaPIbov1 induction and the Asp mutation was expected to abolish induction. All three mutations (Phe41Ala, Phe41Cys, and Phe41Asp) were introduced into the Φ DI wt and Φ DI A73L Duts, whilst Phe41Cys was introduced to the Φ DII Dut and Phe41Ala and Phe41Cys to the $\Phi 55$ Dut. The reasoning behind this was that the Φ DII and $\Phi 55$ Duts are non-inducers, so introducing a mutation to abolish induction capability was pointless.

The Φ DI, Φ DII and $\Phi 55$ wt Duts, together with the Φ DI A73L mutant and all the residue 41 point mutation Duts were 3xFLAG-tagged and cloned into the pCN51 expression vector under a *Pcad* promoter. The cloned Duts were introduced into a SaPIbov1 positive strain and the induction of these islands following Dut expression was tested by Southern blot (Fig 4.6B). As predicted, the Φ DI F41D mutation abolished SaPIbov1 induction in the wild-type and the F41D/A73L double mutant displayed no SaPIbov1 induction, as with the Φ DI A73L Dut. Interestingly, the F41C and F41A mutations not only increased the SaPIbov1 induction levels for the Φ DI wild-type, but restored induction capability in the Φ DI A73L Dut mutant. However, the F41C and F41A mutants did not convey SaPI induction to the non-inducing Φ DII and $\Phi 55$ Duts, although the activity of these proteins needs to be tested to ensure the phenotype is not an artefact of protein inactivity.

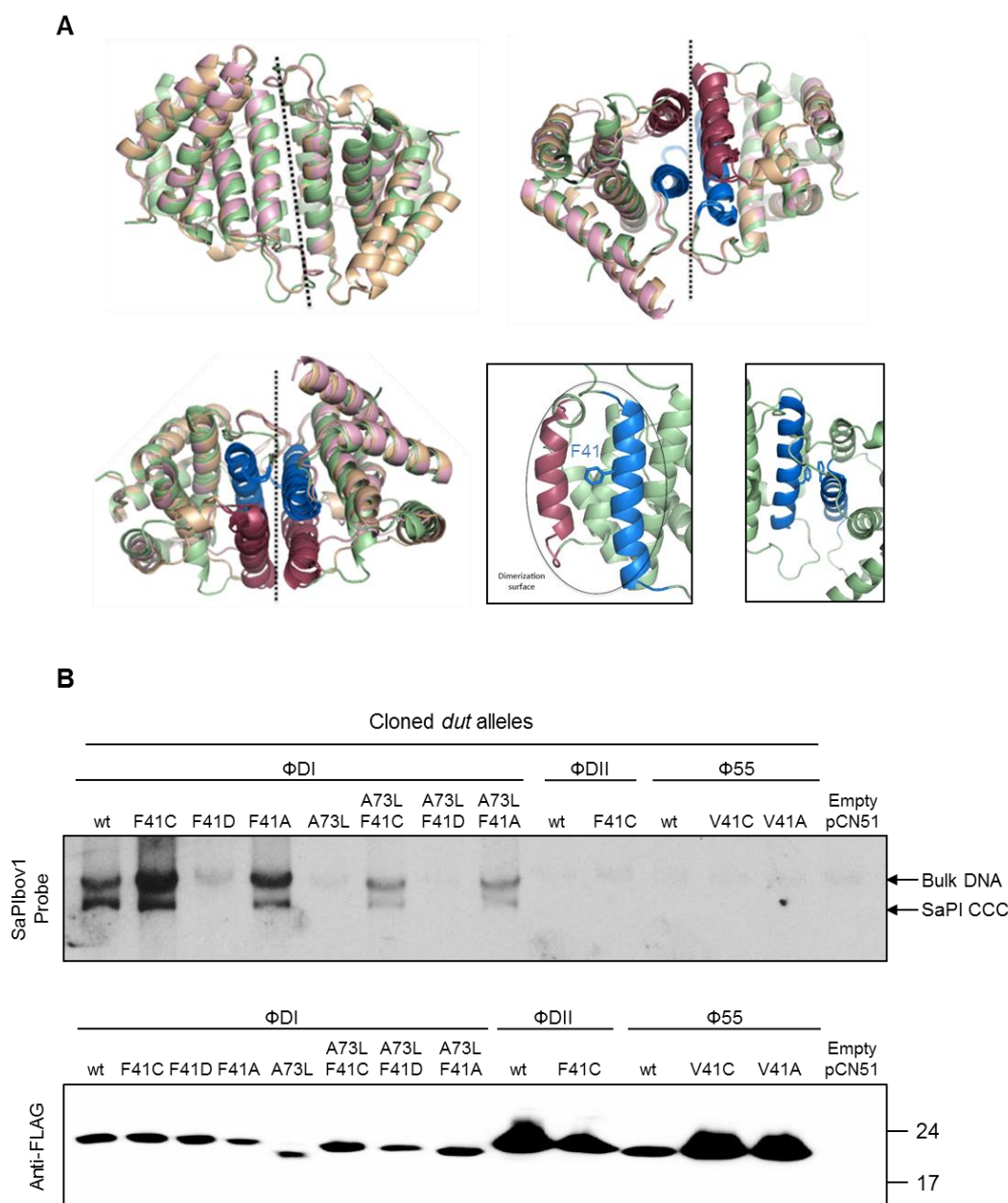


Figure 4.6 The central residue of the dimerization interface plays a role in SaPIbov1 induction. **A** The dimerization surface of the different dimeric Duts. Showing the structure of three different dimeric Duts, the inducing ΦDI (orange) and ΦOII (pink) Duts and the non-inducing Φ55 (green). The dimerization interface is indicated by the black dotted line. Three orthogonal views of the dimers are shown. In the upper, right and lower, left orthogonal views the two helices at the dimerization interface, α2 and α5/6 (ΦOII Dut = α5, ΦDI and Φ55 Dut = α6), are highlighted in blue and red, respectively. The left inset box highlights the positioning of residue 41 in the ΦDI Dut at the centre of the dimerization interface on helix α6. The right-hand box shows residue 41 of each ΦDI Dut protomer, which interact together. **B** SaPIbov1 excision and replication following induction of cloned ΦDI, ΦDII and Φ55 *dut* genes (wild-type and point mutant). Strain JP6774 containing SaPIbov1 was complemented with the different pCN51 plasmids expressing the wild-type and point mutant dimeric Duts, or the empty pCN51 control. Samples were isolated 3 hours post-induction with 5 μM CdCl₂ and Southern blots were performed using a probe for the SaPIbov1 integrase (Table 2.2). Bulk DNA is labelled as such and includes chromosomal, phage, and replicating SaPI DNA. In these experiments, as no helper phage was present, the excised and replicating SaPI DNA appears as part of the bulk DNA or as CCC molecules, rather than the linear monomers that are seen following helper phage-mediated induction and packaging. The upper band represents bulk DNA, whilst the lower band signifies the SaPI CCC band. CCC indicates covalently closed circular SaPI DNA. The lower panel below the Southern blot represents a western blot probed with antibody to the 3xFLAG-tag carried by the Dut proteins at the N-terminal.

Western blots, using anti-FLAG antibodies to probe against the Dut N-terminal 3xFLAG-tags, found that all the Duts were expressing, although protein expression levels were slightly variable. However, as the non-inducing Duts consistently expressed at higher levels than those of the inducing Duts, this variability was unlikely be the cause of the differences in SaPI-induction.

These results clearly indicate a role for the dimerization interface in the phage-encoded dimeric Duts mechanism for binding Stl. Mutations in residue 41 of the interface promoting a monomer conformation in the Duts without affecting the surface charge returned SaPIbov1 induction to the Φ DI A73L Dut that had lost this ability. However, without establishing a Φ DII or Φ 55 Dut mutant capable of inducing SaPIbov1 it remains unclear whether there is another region involved in this interaction. One hypothesis as to why the residue 41 mutations were not able to confer SaPI induction capabilities to the Φ DII and Φ 55 Duts was that these Duts may have a greater affinity for the neighbouring protomer, making it harder to disrupt the dimer. If this was the case, a single point mutation may not impact the more stable dimer formation.

To test this hypothesis, one method would be to exchange the helices that make up the dimerization interface (α 2 and α 5/6) between an inducing and non-inducing Dut. However, when this was designed using the Φ DI and Φ 55 dimeric Duts, structural predictions of the new protein sequences made it apparent that the newly introduced helices may interfere with other helices within the protomer (α 3 and α 8, Appendix 8 and 9). This interference would likely make the protein non-functional and a subtler tactic was required. One method would be to alter only the residues of helices α 2 and α 6 that interact with the neighbouring protomer and are thus involved in the dimerization interface, leaving the rest of each helix unchanged. Figure 4.7A indicates the differences between the dimerization interface residues of the Φ DI and Φ 55 Duts and displays a structural comparison.

Gene synthesis was used to construct chimeric mutants of the Φ DI and Φ 55 Duts in which only the residues highlighted in Fig 4.7A were interchanged. These chimeras and the corresponding wild-type Duts were cloned into the pCN51 expression vector under a *Pcad* promoter. The cloned Duts were introduced into a SaPIbov1 positive strain and the induction of this island following Dut

expression was tested by Southern blot (Fig 4.7B). Again, although the mutations abolished the SaPIbov1 inducing capability of the Φ DI Dut, they did not confer SaPI-induction ability to the Φ 55 Dut.

These results support the hypothesis that the dimerization interface plays a role in the phage-encoded dimeric Dut interaction with the Stl repressor. However, they also imply that another region is involved in this interaction and that a combination of these interaction factors would be required to convey SaPI-inducing capabilities to one of the non-inducing Duts (Φ DII and Φ 55).

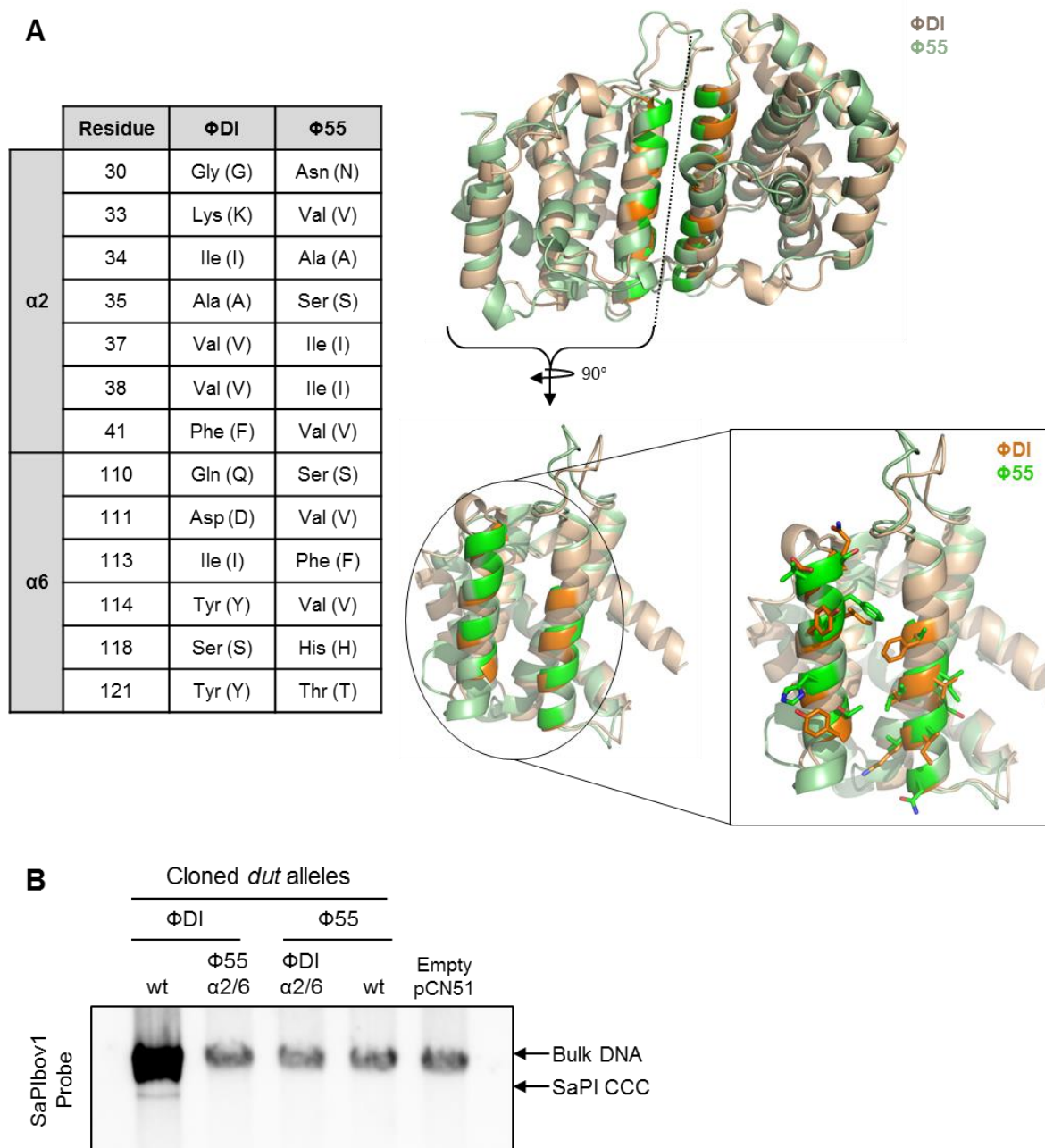


Figure 4.7 The dimerization interface plays a role in SaPI induction, but is not solely responsible for the induction of SaPIbov1. **A** The dimerization surface of the ΦDI (brown/orange) and the non-inducing Φ55 (green) Duts. Showing the structures of the inducing ΦDI (brown/orange) and the non-inducing Φ55 (green) Duts. The dimerization interface is indicated by the black dotted line. The lower, left circle shows the face-on interfaces of a single protomer of each Dut type (ΦDI, orange and Φ55, green) and the lower, right box shows the side-chains of each Dut type in stick, coloured according to the Dut with which they correspond. The left hand table lists the residues for which the ΦDI and Φ55 Dut encode different amino acids, displays the residue for each Dut, and indicates which helix this residue is found on. **B** SaPIbov1 excision and replication following induction of cloned ΦDI and Φ55 *dut* genes (wild-type and chimeric for the dimerization interface residues highlighted in A). Strain JP6774 containing SaPIbov1 was complemented with the different pCN51 plasmids expressing wild-type and chimeric dimeric Duts, or the empty pCN51 control. Samples were isolated 3 hours post-induction with 5 μM CdCl₂ and Southern blots were performed using a probe for the SaPIbov1 integrase (Table 2.2). Bulk DNA is labelled as such and includes chromosomal, phage, and replicating SaPI DNA. In these experiments, as no helper phage was present, the excised and replicating SaPI DNA appears as part of the bulk DNA or as CCC molecules, rather than the linear monomers that are seen following helper phage-mediated induction and packaging. The upper band represents bulk DNA, whilst the lower band signifies the SaPI CCC band. CCC indicates covalently closed circular SaPI DNA.

4.3 Discussion

SaPIs, and more broadly PICIs, are highly effective subcellular parasites^{279,281}. Understanding their interaction with bacteriophage inducers is crucial for establishing how they achieve widespread transfer in bacterial populations. The interaction of the phage-encoded trimeric Duts and the SaPIbov1 Stl repressor has been well characterised, but this Stl repressor has been shown to interact with the structurally divergent dimeric Duts as well^{304,310,359}. How this repressor interacts with the phage-encoded dimeric Duts is not well understood. This chapter aimed to elucidate the mechanism of interaction between the inducing dimeric Duts and the SaPIbov1 Stl repressor, which in turn would help clarify the method by which the Stl repressor is targeting two structurally unrelated proteins.

This study characterizes the structure of two different SaPI-inducing, phage-encoded dimeric Duts, in the first instance of a phage-encoded dimeric Dut structure being solved³¹⁰. These structural characterisations confirm the presence of 5 conserved dimeric Dut motifs in the phage Dut, equivalent to those previously identified except for a truncated C-terminal motif V. Indeed these structural characterisations indicate that the phage-encoded dimeric Duts represent a reduced version compared with the other established dimers, with 7-8 α -helices instead of 10-13. The active centre of the phage-encoded dimeric Duts was further characterized, with no overlap in the spatial disposition of dUTP-binding residues found between the phage dimeric and trimeric Duts. This indicates that, although there may be some general parallelisms between the Stl-binding mechanisms of the two Dut types, the Stl most likely recognises different residues for each Dut type.

Based on the parallelisms in the positioning and variation of the motif VI of the phage-encoded dimeric and trimeric Duts, as identified in Chapter 3, this study aimed to characterise the role of the dimeric motif VI in SaPI-induction. In the trimers, the motif VI is one of two motifs to which the Stl binds (motif IV and VI) before the interaction is stabilised by motif V. It seems that the parallelism of the positioning of the divergent dimeric and trimeric motif VI does not carry through to a parallel role in SaPI-induction. Exchange of the inducing Φ DI Dut motif VI with that of the non-inducing Φ 55 Dut did not affect the SaPI-inducing

capabilities of these two Duts, the Φ DI Dut continued to induce, whilst the Φ 55 could not. However, there did seem to be some impact on the amount of induction, particularly without induction of the Duts by CdCl_2 . With these samples, some of the Φ DI Dut mutants encoding helices from the Φ 55 Dut motif VI induced the island without overexpression of the vector, implying low levels of the protein were sufficient for induction, which was not the case for the Φ DI wt. One possibility for this could be that the phage-encoded dimeric Dut motif VI contributes to the affinity of the Dut for the StI, but more investigation into this motif is needed to draw any firm conclusions.

It has been shown that dUTP inhibits the interaction of the SaPIbov1 StI repressor with the phage-encoded trimeric Duts^{304,368}. This is because when dUTP binds to the trimeric Duts this induces a conformational change in which the conserved C-terminal P-loop (motif V) covers the active site, blocking the access of the StI repressor to its binding site^{304,368}. Additional data from collaborators found similarities with the dimeric Duts; Native-PAGE analysis identified that the Φ DI and Φ O11 Duts interaction with StI was reduced in the presence of the non-hydrolysable dUTP analog, dUPNPP, but not in the presence of the product, dUMP³⁷³.

Given the parallelisms of both the dimeric and trimeric Duts StI-binding capabilities being reduced in the presence of dUTP, one hypothesis was that the dimeric Duts could undergo a similar conformational change to that seen in the trimeric Duts. This concept was further supported by the identification that the previously characterised dimeric Duts can have 'open' (ligand-free) or 'closed' (ligand-bound) conformations^{355,356}. However, comparisons of the structures of ligand-bound and ligand-free (*apo*) Φ DI Duts found minimal conformational changes, with any differences between the two forms limited to the C-terminal $\alpha 8$ helix (motif V), which seems more stable when the ligand is bound³⁷³. Although this is a much lesser effect than that seen in the trimeric Duts, again both Dut types show conformational changes in the motif V when the ligand is bound.

To further test the hypothesis that the dimeric motif V conformational change is important for SaPI-induction, the Φ DI A73L Dut mutant was generated to look at whether a dimeric Dut mutant that was unable to bind dUTP could interact with

Stl to induce the SaPIs. A similar test had been performed with the trimeric Duts, in which the trimeric 80 α Dut mutant Y81I that could not bind the dUTP nucleotide was unable to induce SaPIbov1³⁰⁵. The dimeric Φ DI Dut^{A73L} mutant was confirmed to be inactive as a dUTPase, but the protein was well folded and displayed a dimeric arrangement³⁷³. The results in this chapter confirmed that the A73L dimeric Dut mutant was unable to induce SaPIbov1, although initially it was unclear why. However, investigation into the effect of the mutant on Dut motif V conformation found that the A73L mutant was more stable than both the *apo* and dUPNPP-bound wt Duts³⁷³. This suggests that stabilization of the active centre and catalytic elements inhibits the dimeric Dut-Stl interaction. Such stabilisation occurs in the wt phage-encoded dimeric Duts when dUTP binds, albeit to a lesser extent than in the A73L mutant. This explains why dUTP is an inhibitor of the dimeric Dut:Stl interaction. Therefore, although it is still unclear as to the mechanism of Stl interaction with the dimeric Duts, the dUTP seems to be involved as a signalling regulator for both Dut types.

Although there appear to be some similarities in the mechanism of Stl-interaction for the dimeric and trimeric Duts, there are also clear differences. The conformational change shown by the dimeric Dut motif V on binding dUTP is subtle in comparison to the crucial role the dUTP plays in stabilizing the P-loop motif V of the trimeric Duts. The most likely scenario is that the regulation of the dimeric and trimeric Dut Stl-interaction mechanisms is similar, both involving dUTP as a regulatory switch turning off Stl-interaction, whilst the molecular interaction with the Stl itself differs between the two Dut types.

The previously characterized non-phagic dimeric Duts encoded several phosphate interacting residues on the mobile C-terminal portion of the protein, which is missing in the truncated phage-encoded dimeric Duts^{355,356}. The phage-encoded dimeric Duts have substituted novel catalytic residues from the adjacent subunit, requiring protomer-protomer reorientation to a more parallel conformation³⁷³. This has reduced the size of the dimerization interface in the phage-encoded dimeric Duts, which could lead to reduced stability of the dimer. It has been suggested that the Φ NM1 dimeric Dut binds to Stl as a monomer³⁵⁹. This led to the speculation that the Stl has taken advantage of the low dimer stability to bind the dimeric Dut in its monomeric form.

The results in Section 4.2.4 suggest that the dimerization interface is vital for the Φ DI dimeric Dut Stl-interaction. A single point mutation in helix $\alpha 2$ of the dimerization interface was sufficient to abolish SaPI induction. However, even completely altering the dimerization interface of the non-inducing $\Phi 55$ Dut to match that of an inducing Dut was not sufficient to convey SaPI-inducing capabilities to this Dut. Inactivity of the $\Phi 55$ Dut needs to be ruled out, but this implies that another region of the phage-encoded dimeric Dut is involved in Stl-interaction. It is tempting to speculate that this could be the stabilising ‘latch’ loop between helices $\alpha 2$ - $\alpha 3$, which interacts with the opposite protomer and includes catalytic residues involved in the neighbouring protomers active centre. Indeed, one appealing hypothesis is that the differences in capacity for SaPI-induction between the different dimeric Duts as documented in section 3.2.3 are due to differences in dimer stability.

If dimer stability is the dictator of whether Stl-interaction occurs, this would explain why the dUTP-bound Φ DI Dut exhibited reduced interaction with Stl. The dUTP-mediated conformational change stabilises the dimer when bound, making it harder for dimer disruption to occur. Furthermore, although it may initially appear counterintuitive that the A73L mutant prevents Stl interaction with the dimerization interface, this mutant was shown to be even more stable than the dUPNPP-bound form, promoting a compact conformation that could be stabilising the dimer further³⁷³.

4.4 Conclusions and future work

Establishing the molecular mechanism by which the SaPI Stl repressors interact with their phage-encoded inducers is crucial if the method by which SaPIs achieve efficient inter- and intra-generic transfer is to be understood. This study establishes a potential scenario for the molecular mechanism by which the phage-encoded dimeric Duts interact with the SaPIbov1 Stl repressor, involving the dimerization interface as the binding site for Stl. To fully confirm this hypothesis the molecular characterization of the interaction of Stl with the different *S. aureus* phage-encoded dimeric Dut allelic variants is necessary. Furthermore, the creation of a non-inducing dimeric Dut mutant that is able to interact with the Stl to induce the SaPI would further support this hypothesis and realistically will require knowledge of the interacting structures. Understanding

the Stl repressor itself would also help establish the methods of SaPI-induction by the phage Duts; whether the Stl contains multiple domains for its interaction with both the dimeric and trimeric Duts, as seems most likely, or whether this is achieved through another means. This is addressed in Chapter 6.

Chapter 5 Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts

Results from this section are included in:

Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer (2017) **Bowring J***, Neamah MM*, Donderis J*, Mir-Sanchis I, Alite C, J Ciges-Tomas R, Maiques E, Mehmedov I, Marina A, Penadés JR. *eLIFE*. DOI: 10.7554/eLife.26487

*Denotes joint first authorship

Disclaimer on the work performed:

The SaPI2 work referred to in the Discussion section (5.3) was performed by MM Neamah.

5.1 Background

PICIs and bacteriophages have been shown to impact on one another's evolution, with phages evolving to escape PICI induction and PICIs evolving to better target and interfere with helper phage lifecycles²⁸¹. On top of this, PICIs transfer clinically relevant toxins to bacteria^{29,297}, whilst phages impact on bacterial evolution and pathogenicity^{87,391}. The scale of this impact is a point of interest. Phages have a relatively specific host range, indicating they cannot transfer between different bacterial hosts particularly easily³⁹⁷. However, the mechanisms of the phage life cycle are conserved among phages with different host ranges, implying there could be fairly conserved proteins between these phages.

Investigation into the induction of SaPIbov1 has shown that the SaPIbov1 Stl repressor is capable of targeting both *S. aureus* phage-encoded dimeric and trimeric Duts for its induction, unrelated proteins performing the same function for the phage. The most likely explanation for this is that the Stl repressor is targeting the phage process by encoding multiple domains to target each Dut type. If, as the dimeric and trimeric Dut results suggest, SaPIs target conserved phage processes over specific proteins this could allow them to target multiple phages with bacterial hosts from different species, potentially leading to inter-specific SaPI transfer.

Indeed, it has been demonstrated that SaPIs can be inter- and intra-generically transferred in various laboratory conditions. For example, a previous study simulated the conditions of bovine mastitis by introducing *S. aureus* and *L. monocytogenes* strains into milk and detected the transfer of both SaPI1 and SaPIbov1 to *L. monocytogenes* from the *S. aureus* strain²⁸⁸. Furthermore, another study indicated inter-specific transfer of SaPIbov2 from *S. aureus* to a number of *S. epidermidis* strains and a *S. xylosus* strain under laboratory conditions²⁶⁸. More recently, the ability to confer inter- and intra-specific transfer of SaPIs has been extended to the *cos* packaging type phages, having previously only been associated with *pac* phages³²⁵.

Intriguingly, *in silico* analysis has identified SaPI-like elements in Enterococci, Lactococci, and Streptococci families, whilst *in vitro* testing confirmed the

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts existence of an *Enterococcus faecalis* phage-inducible island (EfCIV583)²⁷⁷. These islands within different families appear to be the result of convergent evolution rather than co-ancestral divergence, but with the SaPI-like elements appearing numerous within these species and genera there is an implication that intra- and inter-specific transfer could be contributing to the spread of PICIs. However, this is only a hypothesis and, with most research in this area being done under laboratory conditions, the actual impact of intra- and inter-specific transfer of PICIs in nature remains unsolved.

The aim of this study was to identify whether, by targeting phage processes that are often conserved between species, the SaPI-like elements can be successfully spread and maintained in nature. As the SaPIbov1 Stl repressor has been characterised as targeting the functionally similar, but structurally divergent dimeric and trimeric Duts, one hypothesis was that if SaPIbov1-related repressors could be found in other staphylococcal species they might also target the phage dUTPase process for their induction. A number of SaPIbov1 Stl homologues within different staphylococcal species were identified, with differing levels of divergence from the SaPIbov1 repressor. Using SaPIbov1-inducing phage-encoded dimeric and trimeric Duts, the ability of these Duts to induce the different SaPIbov1 Stl homologues was tested. Interestingly, this study confirms that the staphylococcal phage-encoded dimeric and trimeric Duts are capable of interacting with divergent SaPIbov1 Stl homologues from other staphylococcal species.

Establishing whether SaPI targeting of conserved phage processes enables transfer between species could increase understanding of how SaPIs achieve intra- and inter-generic transfer. Furthermore, by using SaPIbov1 Stl repressor homologues identified in clinical and veterinary isolates this study could help to establish the extent to which such transfer occurs in nature.

5.2 Results

5.2.1 Identifying SaPIbov1 Stl repressor homologues in different staphylococcal species

A number of studies investigating the transfer of SaPIs outside of *S. aureus* strains have been published in the past ten years^{268,288,325}. However, the majority of these were performed under laboratory conditions and focused on the ability of known SaPIs to transfer to other species and genera. New PICIs have also been identified in other genera using *in silico* analyses, and the novel EfCIV583 island from *E. faecalis* has been confirmed as functional²⁷⁷. Currently, the number and host range of the PICIs and the breadth of their transfer in nature are unknown.

The previous chapters have established that SaPIbov1 can be induced by the structurally unrelated dimeric and trimeric Duts through a direct interaction with the Stl repressor. This targeting of a conserved phage process rather than a specific protein led to a new hypothesis; if there were other, unidentified islands encoding homologues to the SaPIbov1 Stl repressor then potentially they would also target this conserved process.

In order to test this hypothesis, *in silico* analysis was performed to identify the existence of any staphylococcal PICIs encoding homologues of the SaPIbov1 Stl. NCBI BLAST was used to identify non-*S. aureus* proteins with homology to the SaPIbov1 Stl repressor and the results of this search are shown in Table 5.1. The analysis returned six SaPIbov1 Stl homologues from different staphylococcal species; *Staphylococcus sciuri*, *Staphylococcus hominis*, *Staphylococcus haemolyticus*, *Staphylococcus lugdenensis*, *S. saprophyticus*, and *Staphylococcus simulans*. These different homologues had a varying percentage identity with the SaPIbov1 Stl, ranging from 41-99% (Table 5.1).

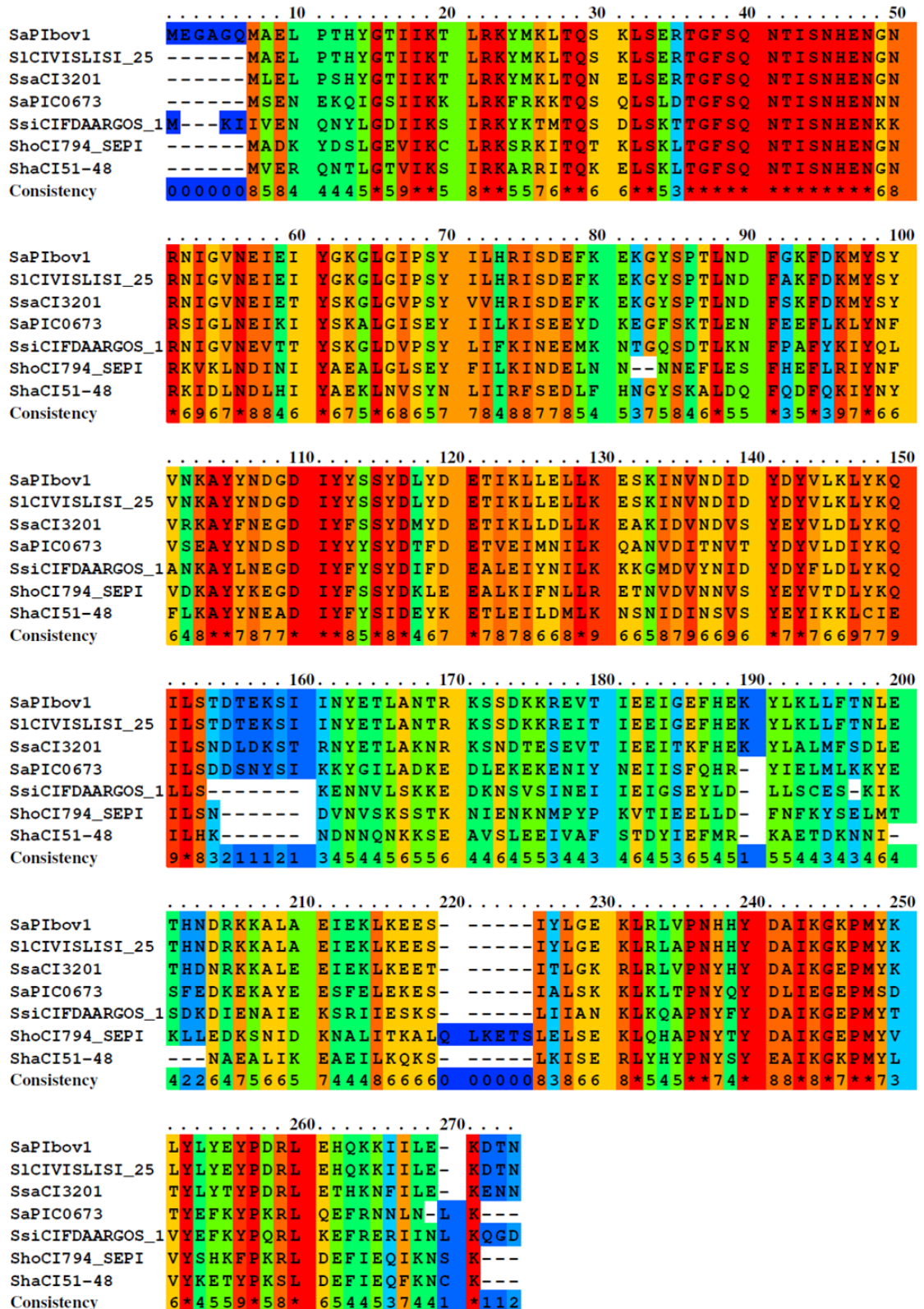
Table 5.1 Phage-inducible islands containing a Stl-like repressor homologous to the SaPIbov1 Stl repressor, as identified in this study.

PICI	Strain	Accession number (Stl protein)	% Identity with SaPIbov1 Stl
SaPIbov1	<i>Staphylococcus aureus</i> RF122	AAG29617	100
SaPIC0673	<i>Staphylococcus sciuri</i> C0673*	EZX20759	49
ShoCI794_SEPI	<i>Staphylococcus hominis</i> 794_SEPI	WP_049379208	45
ShaCI51-48	<i>Staphylococcus haemolyticus</i> 51-48	WP_053031052	41
SICIVISLISI_25	<i>Staphylococcus lugdunensis</i> VISLISI_25	ARJ14897	99
SsaCI3201	<i>Staphylococcus saprophyticus</i> 3201	OEK45527	79
SsiCIFDAARGOS_124	<i>Staphylococcus simulans</i> FDAARGOS_124	AMG96720	45

Accession numbers and % identities are for Stl-like repressor proteins only. Stl % identities were determined with BLASTP. Asterisk (*) indicates strain that is mislabelled as *S. aureus* in NCBI database³⁹⁸.

To analyse the differences in divergence between the SaPIbov1 Stl and homologues, the proteins sequences of these repressors were compared. A PRALINE-generated alignment of the protein sequences of the SaPIbov1 Stl homologues listed in Table 5.1 indicates some areas of conservation and other areas of divergence (Fig 5.1). The region between residues 154 and 210 seems the most unconserved, whilst the section between residues 36 to 48 is identical for all homologues and forms part of a wider helix-turn-helix domain encompassing residues 18-72. The central region between residues 100 and 150 also seems fairly well conserved. These regions of similarity could indicate conserved domains, and the existence of the helix-turn-helix domain makes sense in the context of the Stl repressors role of repressing the island by binding the PICI *stl-str* intergenic DNA. In fact, in support of this idea the *stl-str* intergenic regions of different SaPIs seem to contain a conserved motif (Appendix 10). This also correlates with other SaPI islands that show conservation in this region, but are otherwise unconserved³⁹⁹. Interestingly, the SICIVISLISI_25 homologue from *S. lugdunensis* is almost completely identical to the SaPIbov1 Stl, with only 3 residue differences and 6 gaps at the N-terminal.

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved



Having confirmed the presence of similarities and differences in the protein sequences, the next step was to more closely examine the relationships between the different Stl homologues, as well as the SaPIbov1 Stl repressor. To achieve this, a phylogenetic tree of the SaPIbov1 Stl and homologues was constructed based on a CLC workbench generated protein sequence alignment (Appendix 11). This neighbour joining tree shows the Jukes-Cantor measure of distance between the different SaPIbov1 Stl repressor homologues and includes the divergent SaPI2 Stl as an outlier (Fig 5.2).

As expected from the protein sequence identity percentages, the SaPIbov1 Stl and the SICIVISLISI_25 homologue group together. The SsaCI3201 homologue also groups near to these two proteins, again matching the identity percentages. The SaPIC0673 homologue grouped together with the SsiCIFDAARGOS_124 homologue, whilst the ShoCI794_SEPI and ShaCI51-48 homologues grouped together. As would be predicted for a divergent Stl repressor from another SaPI, the SaPI2 Stl repressor outlier sits away from the SaPIbov1 Stl and homologues (branch length = 1.255).

Using the same CLC workbench protein sequence alignment excluding the SaPI2 Stl (Appendix 11), pairwise comparisons were performed to further investigate the relationships between the different homologues (Fig 5.3). Comparisons of the percentage identity and the number of gaps (Fig 5.3A, upper and lower, respectively) both indicate that the SaPIbov1 Stl, the SICIVISLISI_25 homologue, and the SsaCI3201 homologue have the highest percentage identity and the lowest number of gaps in relation to the other homologues. The other homologues are fairly consistent in their divergence, with percentage identities between these proteins ranging between 35-50%. Similarly, the SaPIbov1 Stl, SICIVISLISI_25 homologue, and SsaCI3201 homologue have the lowest Jukes-Cantor corrected distance measures and the lowest number of differences including gaps when compared to the other homologues (Fig 5.3B, upper and lower, respectively). Again, the other homologues have higher, fairly similar Jukes-Cantor corrected distance measures and differences. Overall, the SaPIbov1 Stl and the SICIVISLISI_25 homologue are highly similar, correlating with the relationships shown in the phylogenetic tree and the protein alignment.

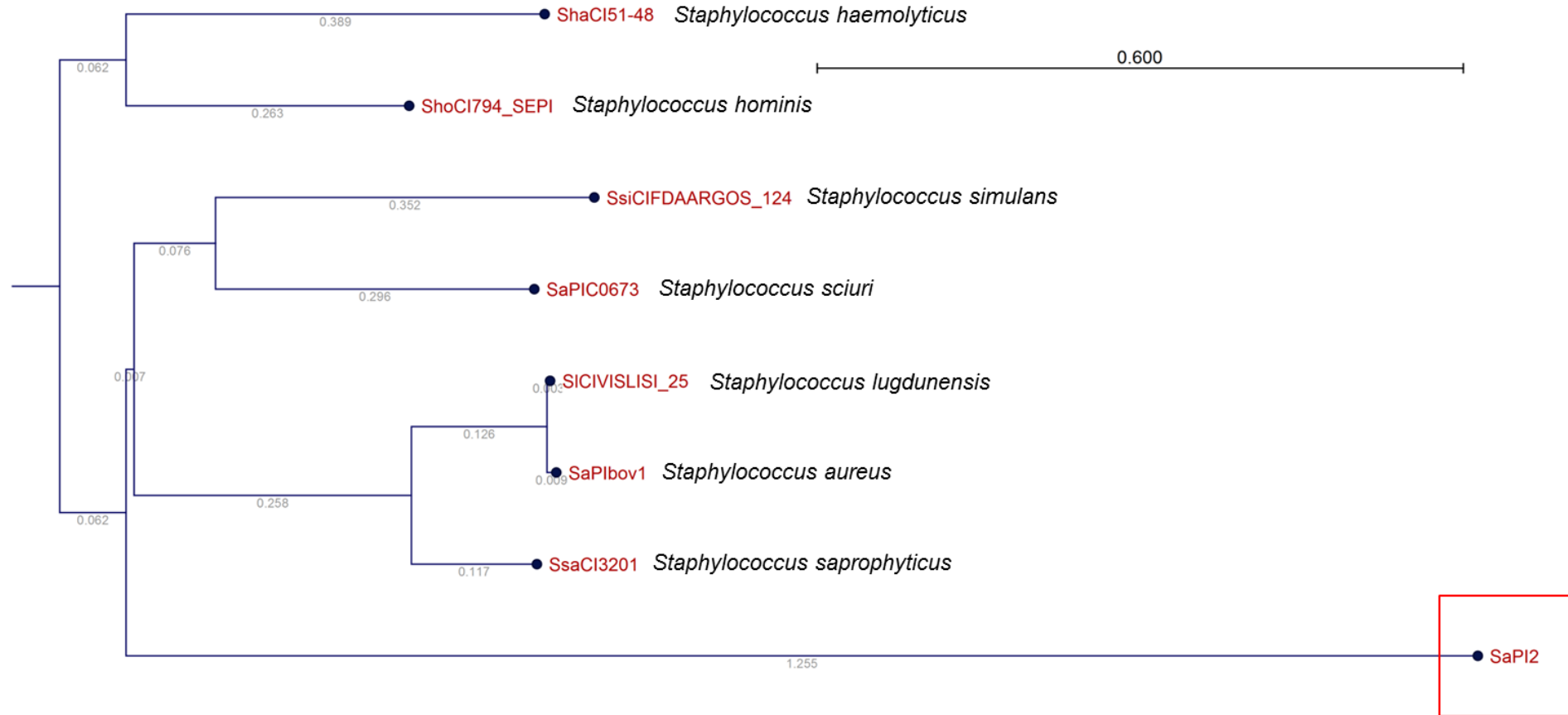


Figure 5.2 Phylogram of the SaPIbov1 StI repressor and different homologues from various staphylococcal species. Phylogenetic tree showing the arrangement of the SaPIbov1 StI repressor homologues detailed in Table 5.1, constructed using an alignment of the StI protein sequences. The PICI name is detailed in red text, with the species in black text alongside. The SaPI2 StI repressor (*Staphylococcus aureus*) was also included in the alignment and tree to provide an outlier (red box). The tree was created using the following parameters; algorithm = Neighbour Joining, distance measure = Jukes-Cantor, bootstrap = 100 Replicates. The CLC alignment used to construct this tree is included in Appendix 11.

A

		1	2	3	4	5	6	7
SaPIbov1	1		96.63	76.78	42.59	38.15	45.19	41.85
SICIMSLISI_25	2	6		77.78	43.56	38.64	46.04	42.70
SsaCI3201	3	6	0		44.32	40.91	46.04	42.70
ShoCI794_SEPI	4	17	11	11		50.39	44.53	40.45
ShaCI51-48	5	25	19	19	12		36.98	39.08
SaPIC0673	6	16	12	12	17	25		49.25
SsiCIFDAARGOS_124	7	17	17	17	22	18	19	

B

		1	2	3	4	5	6	7
SaPIbov1	1		0.01	0.24	0.81	0.89	0.75	0.83
SICIMSLISI_25	2	9		0.25	0.81	0.91	0.75	0.81
SsaCI3201	3	62	58		0.79	0.84	0.75	0.81
ShoCI794_SEPI	4	155	149	147		0.65	0.76	0.84
ShaCI51-48	5	167	162	156	128		0.93	0.90
SaPIC0673	6	148	143	143	147	167		0.65
SsiCIFDAARGOS_124	7	157	153	153	159	159	135	

Figure 5.3 Pairwise comparisons of the SaPIbov1 StI repressor and 6 StI homologues. These pairwise comparisons were constructed using the protein sequence alignment of the StI repressor homologues (listed in Table 5.1 and shown in Appendix 11). The gradient of blue to red represents a sliding scale, with dark blue as minimum, white as median, and dark red as maximum. **(A) Showing the percentage identity above and the number of gaps below.** The upper triangle shows the percentage of identical residues in alignment positions to overlapping alignment positions between the two sequences. The lower triangle shows the number of alignment positions where one sequence has a gap and the other does not (gaps). **(B) Showing the distances above and the number of differences below.** The upper triangle shows the Jukes-Cantor corrected distance between two sequences (distance). The lower triangle shows the number of alignment positions where one sequence is different from the other, including gap differences (differences).

5.2.2 Pirating conserved phage processes supports inter-specific SaPI transfer

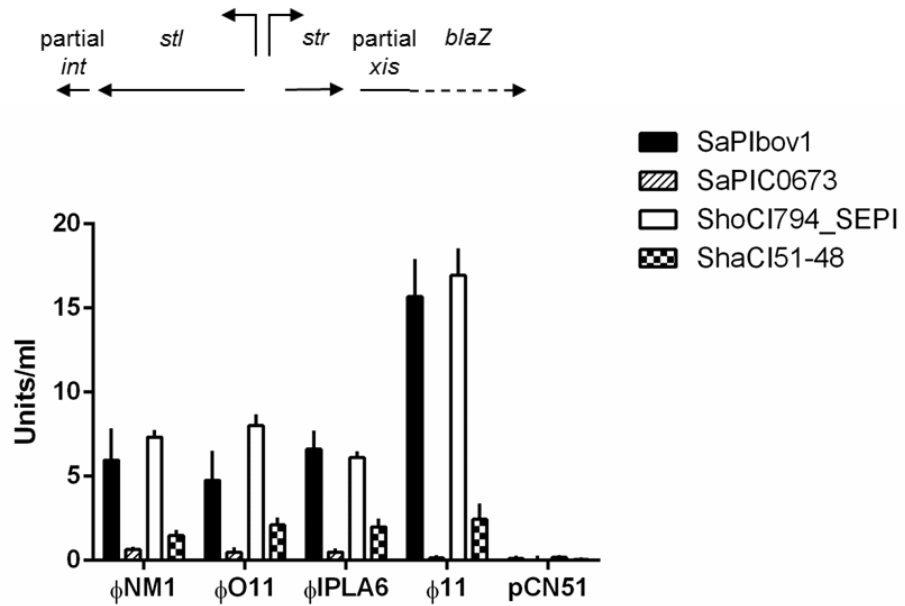
Having identified six homologues of the SaPIbov1 Stl and establishing their protein sequence similarities and differences, the next step was to look at whether the SaPIbov1-inducing staphylococcal phage-encoded dimeric and trimeric Duts were capable of interacting with these SaPIbov1 Stl homologues. To achieve this, plasmids carrying a β -lactamase reporter gene fused to what would traditionally be the Stl-repressed *str* promoter, *str* and *xis* genes were created. By introducing this plasmid into strains expressing different staphylococcal phage-encoded dimeric and trimeric Dut clones (pCN51), the consequences of Dut expression with regard to gene expression of the Stl-repressed genes could be analysed.

Of the six homologues identified, *blaZ*-transcriptional fusion plasmids were constructed for SaPIC0673, ShoCI794_SEPI, and ShaCI51-48. These were selected due to their lower level of homology with the SaPIbov1 Stl; if these could interact with *S. aureus* phage Duts then that would address the hypothesis better than if highly similar homologues were used. Dimeric Duts encoded by phages ϕ NM1 and ϕ O11 and trimeric Duts encoded by phages ϕ IPLA6 and ϕ 11 were cloned into pCN51 under control of a cadmium promoter. The *S. epidermidis* ϕ IPLA6 phage-encoded trimeric Dut was included to establish whether this Dut from a different phage species could interact with the SaPIbov1 Stl and homologues. This would support the idea that following inter-species transfer the PICI can be maintained in the new recipient species. Strains containing all the different combinations of fusion plasmids and pCN51s were obtained, including the SaPIbov1 Stl wt (pJP674) and the empty pCN51 as a control.

Figure 5.4 shows the levels of β -lactamase expression in the presence of an inducing concentration of CdCl_2 ($5\mu\text{M}$) for the different combinations of pCN51 Dut clones and fusion plasmid Stl homologues. The β -lactamase expression levels can be interpreted as a surrogate for the expression of the Stl-repressed *str* and *xis* (or equivalent) genes following derepression of Stl from the intergenic region.

As previously seen, expression of the ϕ NM1, ϕ O11 and ϕ 11 Duts in the presence of the SaPIbov1 wt plasmid all led to significantly higher levels of β -lactamase levels when compared to the empty pCN51 control. Interestingly, induction of the *S. epidermidis* ϕ IPLA6 Dut with the SaPIbov1 wt plasmid also led to a significant increase in β -lactamase levels, comparable to the levels seen following ϕ NM1 and ϕ O11 Dut induction. This indicates that the Duts from other phage species are also able to induce PICIs, supporting the idea that following inter-specific transfer the islands can be maintained.

The β -lactamase levels for the plasmid encoding the SaPIC0673 homologue (*S. sciuri*) were not significantly different to the empty pCN51 control levels, indicating none of the chosen Duts were capable of derepressing translation of the downstream genes. The ShaCI51-48 homologue (*S. haemolyticus*) plasmid showed low levels of β -lactamase following induction of the different Dut clones, but these were still significantly higher than those of the empty vector control. Interestingly, the plasmid encoding the ShoCI794_SEPI homologue (*S. hominis*) showed similar β -lactamase levels to the SaPIbov1 wt plasmid, with the same pattern of expression for the different Dut inducers. A summary of the different β -lactamase levels for the different StI homologues and the different Duts can be seen in Fig 5.4B.

A**B**

PICIs	Phage Duts			
	Dimeric		Trimeric	
	φNM1	φO11	φIPLA6	φ11
	<i>S. aureus</i>	<i>S. aureus</i>	<i>S. epidermidis</i>	<i>S. aureus</i>
SaPIbov1	++	++	++	+++
SaPIC0673	-	-	-	-
ShoCI794_SEPI	++	++	++	+++
ShaCI51-48	+	+	+	+

Figure 5.4 Protein expression of trimeric and dimeric Dut variants leading to the release of Stl repression of downstream genes in SaPIbov1 and SaPIbov1 Stl repressor homologue regulatory modules. (A) The diagram represents a schematic of the *blaZ* transcriptional fusion plasmids generated for each of the four islands. β -lactamase assays were performed on strains containing the different SaPIbov1 Stl homologues together with pCN51-derived plasmids expressing the dimeric Duts encoded by φNM1 or φO11 and the trimeric Duts encoded by φ11 or φIPLA6 as well as the empty pCN51 control. Samples were taken after 5 hours following induction with 5 μ M CdCl₂. All data is the result of 5 independent experiments. Error bars represent standard error of the mean. **(B)** Levels of induction were based on the calculated units/ml with the following ranges: (-): no induction, <1 Units/ml; (+): low induction, 1-5 Units/ml; (++) : moderate induction, 5-10 Units/ml; (+++) : high induction, >10 Units/ml. A 2-way ANOVA with Holm-Sidak's multiple comparisons test was performed to compare mean differences within rows. The differences observed among the +, ++ or +++ samples, compared with the control, were in all cases statistically significant ($p < 0.05$). The p -values for the different Stl regulatory module homologues when comparing the phage Dut-induced values to the pCN51 control were as follows: SaPIbov1, φNM1= <0.0001 ^{****}, φO11= <0.0001 ^{****}, φIPLA6= <0.0001 ^{****}, φ11= <0.0001 ^{****}; SaPIC0673, φNM1=0.8776^{ns}, φO11=0.8973^{ns}, φIPLA6=0.8973^{ns}, φ11=0.8973^{ns}; ShoCI794_SepI, φNM1= <0.0001 ^{****}, φO11= <0.0001 ^{****}, φIPLA6= <0.0001 ^{****}, φ11= <0.0001 ^{****}; ShaCI51-48, φNM1=0.0483^{*}, φO11=0.0162^{*}, φIPLA6=0.0168^{*}, φ11=0.0055^{**}.

These results confirm that some, but not all of the SaPIbov1 Stl homologues identified in different staphylococcal species are capable of being induced by both dimeric and trimeric staphylococcal phage-encoded Duts. The SaPIC0673 homologue did not show any increase in the translation of the downstream genes, as indicated by an increase in β -lactamase levels. However, both the ShoCI794_SEPI and ShaCI51-48 homologues displayed an increase in downstream translation in the presence of both dimeric and trimeric Dut expression. This implies these two repressor homologues were capable of interacting with the dimeric and trimeric Duts, whilst the SaPIC0673 repressor homologue was incapable of interacting with any of these Duts. Taken together this would imply that the SaPIbov1 Stl and the ShoCI794_SEPI and ShaCI51-48 homologues have a common origin and all target variations of the phage Duts, but that the SaPIC0673 has evolved to target a different, though potentially similar, conserved phage process.

Interestingly, the ShoCI794_SEPI and ShaCI51-48 homologues were grouped together in the phylogenetic tree referred to earlier in the Chapter (Fig 5.2). The SaPIC0673 homologue, however, was grouped separately with the untested SsiCIFDAARGOS_124 homologue. It would be useful to construct a fusion plasmid for the SsiCIFDAARGOS_124 homologue to see whether this repressor is capable of interacting with the staphylococcal phage Duts. If not, this would further support the theory that this grouping represents Stl homologues targeting a related, but separate, phage mechanism.

Based on the Dut variation seen within the *S. aureus* phages (Chapter 3), it would be expected that PICIs within other species would come into contact with phages encoding highly variable Duts. These other staphylococcal phage-encoded Duts could potentially be quite divergent from those in *S. aureus* phages. As such, the PICI repressors would evolve to interact with those Duts encoded by phage present in the surrounding environment and diverge from the point of common origin. This evolution could reduce the capacity of the PICI repressors in other species to interact with *S. aureus* phage Duts. This seems to be the case for the ShaCI51-48 homologue, which has low β -lactamase levels implying the *S. aureus* phage Duts are not that efficient at interacting with this repressor homologue. Conversely, the ShoCI794_SEPI homologue has a similar phenotype to

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts

SaPIbov1, implying the presence of phage Duts in the environment that are potentially homologous to the *S. aureus* phage-encoded Duts.

In an effort to confirm whether or not the ShoCI794_SEPI and ShaCI51-48 homologues were targeting Duts encoded by phages infecting their host species (*S. hominis* and *S. haemolyticus*, respectively) species-specific prophages documented in NCBI or papers were identified and the presence or absence of a *dut* gene was determined. This also allowed investigation into the presence or absence of Duts in *S. sciuri* prophage, as *S. sciuri* encodes the SaPIC0673 homologue that could not interact with the *S. aureus* phage Duts. Prophages identified for each species and any *dut* genes they encoded are documented in Table 5.2.

Table 5.2 Known prophages in the staphylococcal species encoding SaPIbov1 StI homologues and the presence or absence of a phage-encoded Dut in that phage genome.

Bacterial host species (PICl)	Prophages	Accession number (Phage genome)	Dut	Accession number (Dut protein)	Ref
<i>S. sciuri</i> (SaPIC0673)	φ575	KY389063.1	None	N/A	398
	φ879	KY389064.1	None	N/A	398
<i>S. hominis</i> (ShoCI794_SEPI)	φStB12	JN700520.2	Dimeric Dut	AFD22247.1	400,401
	φStB27	JN700519.1	None	N/A	400,401
<i>S. haemolyticus</i> (ShaCI51-48)	φShaem1	AP006716 (2353153 - 2396293)	Trimeric Dut	WP_011276624.1	400

Accession numbers are for the full phage genome (or where this does not exist the bacterial host genome with phage genome range labelled) or for the Dut protein alone, as indicated by the headings. Ref, Reference papers for the phages.

Interestingly, the identified *S. sciuri* prophages did not encode any *dut* genes, whilst both the *S. hominis* φStB12 and *S. haemolyticus* φShaem1 did. PRALINE-generated alignments of the protein sequences of the φStB12-encoded dimeric Dut with the *S. aureus* φNM1 dimeric Dut and of the φShaem1-encoded trimeric Dut with that of *S. aureus* φ11 indicated some sequence conservation (Fig. 5.5A and B). Indeed, the dimeric Dut alignment indicated that the φStB12 and φNM1-

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts encoded dimeric Duts were highly conserved at the N- and C-terminal, with a central divergent region reminiscent of the *S. aureus* phage-encoded dimeric Dut alignments in Chapter 3. However, there are some residue variations in the conserved domains of this alignment that are not seen when comparing two *S. aureus* dimeric Duts.

These results seem to correlate with the β -lactamase data from Fig 5.4, with the Dut from ϕ StB12 the best conserved in comparison to the ϕ NM1 dimeric Dut. This prophage was identified in *S. hominis*, which encodes the ShoCI794_SEPI that exhibited a similar β -lactamase phenotype to the SaPIbov1 wt. On the other hand, the ϕ Shaem1-encoded trimeric Dut shares some conserved residues with the ϕ 11 Dut, but to a lesser degree of conservation. This prophage was identified in *S. haemolyticus*, which encodes the ShaCI51-48 homologue that displayed reduced levels of β -lactamase production, implying lower levels of interaction with the *S. aureus* phage Duts. That these prophages were identified in the same bacterial species encoding the related SaPIbov1 Stl homologues hints at the possibility that the ϕ StB12- and ϕ Shaem1-encoded Duts could be the inducers of ShoCI794_SEPI and ShaCI51-48 islands, respectively.

Further investigation into the genomes of the identified *S. sciuri* phages ϕ 575 and ϕ 879 identified one putative NTP pyrophosphohydrolase MazG family protein in each phage (Table 5.2, MazG accession numbers: AQN32201.1 and AQN32255.1, respectively). The MazG family proteins act as non-specific nucleoside triphosphate pyrophosphatases (NTP-PPases) and belong to a superfamily of α -helical NTP pyrophosphohydrolases, which also includes the dimeric Duts^{372,402,403}. It has been suggested that this superfamily share a MazG-like subunit as a common ancestor³⁷². A PRALINE-generated alignment indicated that the *S. sciuri* phage-encoded MazG proteins are divergent in protein sequence from the *S. aureus* ϕ NM1-encoded dimeric Dut (Fig 5.5C). As such, these *S. sciuri* phage-encoded MazG family proteins may be distantly related to the dimeric Duts, but are unconserved in sequence. This could explain why the SaPIC0673 repressor displays homology to the SaPIbov1 Stl, but could not interact with the staphylococcal Duts. To test this hypothesis, it would be interesting to see whether these MazG proteins could induce the SaPIC0673 island.

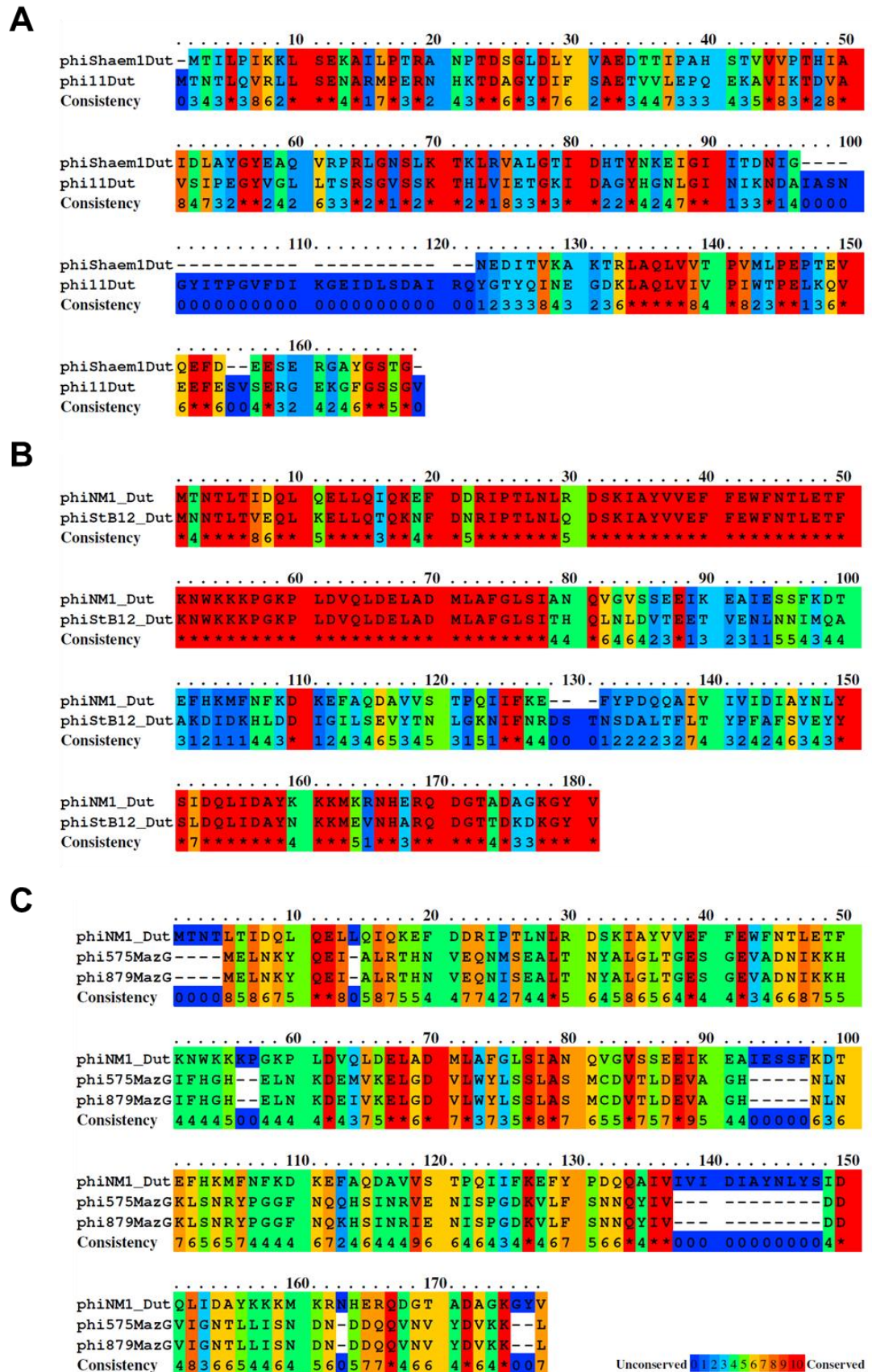


Figure 5.5 Protein alignment of (A) ϕ 11- and ϕ Shaem1-encoded trimeric Duts (B) ϕ NM1- and ϕ StB12-encoded dimeric Duts (C) ϕ NM1-encoded dimeric Dut and ϕ 575- and ϕ 879-encoded MazG family proteins. Details of different Duts displayed in Table 5.2 and MazG proteins in text. Colours indicate relative sequence conservation at each position, with red being most conserved and blue being least (adapted from alignment generated by PRALINE)^{284,285}.

5.3 Discussion

Understanding how PICIs spread in nature and the extent to which this occurs would clarify the significance of PICIs in intra- and inter-specific pathogenicity gene transfer. Previous studies have mainly used known SaPIs under laboratory conditions to look at whether they can transfer to other genera and species of bacteria, but have also detected the presence of PICIs in a range of bacterial genera and species^{268,277,288,325}.

The previous chapters in this study have established that the SaPIbov1 Stl repressor is capable of interacting with both dimeric and trimeric *S. aureus* phage Duts for island induction. One explanation for this is that the SaPIbov1 repressor has evolved to target the conserved phage dUTPase process, rather than individual phage proteins. In this way, the SaPI repressor could have diverged to encode different domains enabling the Stl to interact with the two structurally unconserved Dut types, increasing chances of phage-mediated induction. Extending this hypothesis, and taking into account previous papers showing SaPI transfer to other genera and species, the ability of the SaPI Stl to target multiple, divergent proteins performing the same conserved phage process could enable them to target a broader range of phages. If a SaPI repressor was capable of targeting Duts encoded by phages that infect other staphylococcal species, then this could facilitate inter-specific SaPI transfer.

This study aimed to establish whether the potential Stl repressor targeting of conserved phage processes occurs in nature and if it has enabled PICIs to transfer to other species, before further evolving to target the phages in this new environment. To do this, the Stl repressor of the best characterised SaPI, SaPIbov1, was used to search for homologues in other species. This chapter aimed to identify and characterise SaPIbov1 Stl homologues from other staphylococcal species and establish their relationship with SaPIbov1-inducing *S. aureus* phage-encoded dimeric and trimeric Duts.

This study uncovered the presence of previously unidentified SaPIbov1 Stl repressor homologues in multiple staphylococcal species. The relationships between the SaPIbov1 Stl and these different homologues were investigated through their protein sequences, finding a range of conservation between the

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts

SaPIbov1 Stl and homologues. Three of the more divergent SaPIbov1 Stl homologues were carried forward for *in vitro* characterisation of their interaction with the *S. aureus* phage-encoded dimeric and trimeric Duts.

β -lactamase experiments between plasmids encoding the SaPIC0673, ShoCI794_SEPI, and ShaCI51-48 homologues and the dimeric and trimeric *S. aureus* phage Duts indicated that only the ShoCI794_SEPI and ShaCI51-48 homologues were capable of interacting with these Duts. The SaPIC0673 homologue actually had a slightly higher percentage identity to the SaPIbov1 Stl than the other two homologues, but did not show any downstream translation of the *blaZ* gene, implying no interaction with the Duts. These results suggested that the SaPIbov1 Stl and the ShoCI794_SEPI and ShaCI51-48 homologues shared a common ancestor and have maintained similar regions for Dut interaction, whilst the SaPIC0673 homologue may well have diverged in a different niche and no longer shared the similarities necessary for Dut interaction.

Investigation into the prophages documented in the same species that encode the SaPIbov1 Stl homologues identified the presence or absence of phage-encoded *dut* genes. Correlating with the β -lactamase experiments, Dut-encoding prophage were identified in *S. hominis* and *S. haemolyticus* (encoding ShoCI794_SEPI and ShaCI51-48, respectively), but not in *S. sciuri* (encoding SaPIC0673). Interestingly, prophages in the *S. sciuri* were found to encode putative MazG family proteins. The MazG family of proteins have been bracketed in the same α -helical NTP pyrophosphohydrolases superfamily as the dimeric Duts³⁷².

Referring back to Fig 5.2, the SaPIC0673 homologue was grouped together with the SsiCIFDAARGOS_124. Preliminary investigations into the proteins of the *S. simulans* FDAARGOS_124 strain identified 3 MazG-like proteins and no Dut proteins, although these proteins were not clearly linked to any prophages. However, this could potentially indicate that these two homologues grouped together as MazG targeting repressors. If this is the case, it would explain why both the ShoCI794_SEPI and ShaCI51-48 homologues interacted with the Duts, whilst the SaPIC0673 did not. Furthermore, it opens up the possibility that while the SaPIbov1 Stl repressor and homologues have evolved to target the conserved

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts phage-encoded dUTPase process, others may have diverged to target the MazG family.

On top of this, work completed in tandem to the SaPIbov1 Stl/Dut studies described here established a similar pattern for the SaPI2 Stl and phage-encoded recombinases³¹⁰. These studies identified the phage 80 α -encoded Sak recombinase as the inducer of SaPI2. In parallelism with the SaPIbov1 results, it was established that *S. aureus* phage-encoded recombinases from the Sak4, Erf, and RedB families, which are unrelated in protein sequence, were also capable of inducing SaPI2. These different SaPI2-inducing recombinases were all predicted to have different structures implying that, like the SaPIbov1 Stl, the SaPI2 Stl has evolved to target multiple unrelated proteins performing the same function for the phage.

Interestingly, SaPI2 Stl homologues were identified in a number of different staphylococci (*Staphylococcus argenteus*, *Staphylococcus caprae*, *S. lugdunensis*, *S. epidermidis*, *S. haemolyticus*, *S. simulans*, *S. xylosus*, and *Staphylococcus capitis*), as well as in PICI from other Gram-positive bacteria, including *Bacillus decisifrondis* or *Streptococcus pyogenes*³¹⁰. As with SaPIbov1 and SICIVISLISI_25, some of these SaPI2 Stl homologues were identical to the SaPI2 Stl wt (SarCISJTUF21285 - *S. argenteus*, ScCIM23864:W1- *S. caprae*, and SICIFDAARGOS_141 - *S. lugdunensis*).

A β -lactamase dual plasmid experiment was performed using the SaPI2 Stl wt and the ShaCI137133 (*S. haemolyticus*), SeCINIHLM095 (*S. epidermidis*) and SsCIUMC-CNS- (*S. simulans*) homologues together with the different *S. aureus* phage-encoded recombinases³¹⁰. All the SaPI2 homologues were induced by at least one recombinase, however there was variation between which recombinase could interact with the different homologues. For example, the 80 α -encoded Sak induced SaPI2, SeCINIHLM095 and ShaCI137133, but not SsCIUMC-CNS-990. The ϕ 52A-encoded Sak4 recombinase induced SaPI2 and SsCIUMC-CNS-990, whilst the ϕ SLT Erf recombinase induced SaPI2 and SeCINIHLM095. The ϕ N315-encoded RedB recombinase only induced SaPI2, unable to interact with any of the homologues. This supports the SaPIbov1 data indicating that different Stl repressors have evolved domains to interact with different phage-encoded

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts inducers that are structurally unrelated proteins, but that perform the same function for the phage.

That both SaPI2 and SaPIbov1 Stl repressors had identical homologues encoded in different staphylococcal species strongly supports the idea of ancestral elements being horizontally transferred between species. The expectation would be that these elements would then evolve independently to adapt to their new host. The number of homologues found for both the SaPIbov1 and SaPI2 implies that pirating conserved phage mechanisms could allow this inter-specific transfer to occur constantly in nature. Indeed, of the SaPI2 homologues with identical Stl repressors the ScCIM23864:W1 (*S. caprae*) and SICIFDAARGOS_141 (*S. lugdunensis*) elements were actually found to be identical to each other, with just 3 mismatches over 13,847 nt³¹⁰. All this implies that the targeting of conserved phage mechanisms has created versatile elements that can successfully spread inter- and intra-specifically in nature.

Originally the SaPIs may have targeted a single phage protein for their derepression. In response, to avoid detrimental SaPI induction, the phage could have substituted the encoding gene to a non-related, but functionally similar protein. This in turn could have triggered a second stage of SaPI evolution, leading to the divergence of the SaPI Stl to target this new protein, enabling it to interact with structurally unrelated phage proteins. A more simple strategy might have been to target multiple phage proteins performing different functions for the phage. However, given the mosaicism exhibited by phages encoding multiple versions of functionally similar, structurally diverse proteins, this would select for insensitive phages with a combination of non-inducing proteins. By targeting multiple versions of proteins involved in the same conserved process the SaPIs make it difficult for the phages to overcome SaPI parasitism and increase their probability of inter- and intra-specific transfer.

5.4 Conclusions and future work

The possibility of inter- and intra-specific transfer of SaPIs, and more broadly PICIs, occurring frequently in nature has previously been implied by the ease at which this occurs under laboratory conditions. This study suggests that this is likely the case, establishing that the SaPIbov1 Stl repressor has homologues in

Chapter 5. Establishing the inter-specific impact from SaPIbov1 targeting dimeric and trimeric Duts multiple species that are capable of interacting with SaPIbov1-inducing Duts. The apparent strategy of the Stl repressors targeting structurally unrelated proteins performing the same function for the phage is unique in nature and appears to be highly effective. This highlights the SaPIs as intriguing, highly efficient subcellular parasites.

Preliminary results in this Chapter hint at the potential existence of another PICI-inducing phage protein family, the MazG family of NTP pyrophosphohydrolases. MazG family proteins were identified in prophages for *S. sciuri*, the species encoding the SaPIC0673 homologue of the SaPIbov1 Stl that did not interact with the Duts. It would be interesting to check whether these MazG proteins were capable of inducing the SaPIC0673 and, if so, whether this represents another conserved phage process targeted by the PICIs.

It would be helpful to test whether the SsiCIFDAARGOS_124 (*S. simulans*) homologue can interact with the Duts or not. The fact that the SsiCIFDAARGOS_124 homologue groups with the SaPIC0673 in the phylogenetic tree in Fig 5.2 suggests it may not interact with the Duts. If this is the case, identifying recorded *S. simulans* prophages to establish whether they contain any Dut or MazG proteins could help elucidate a potential inducer of this homologue.

Preliminary data in this study has identified potential Dut inducers of the ShoCI794_SEPI and ShaCI51-48 homologues in prophages found in the same bacterial species as these putative islands. Establishing whether these Duts are capable of derepressing these islands would further confirm the hypothesis that the repressors are targeting the Dut process, but evolving to better interact with Duts encoded by phages in the vicinity of the islands.

Although we know that individual SaPI Stl repressors are capable of interacting with structurally dissimilar proteins, how they achieve this is not known. In this study, the assumption has been made that the Stl encodes multiple domains to achieve this interaction, but this hypothesis has not been tested. Identifying the presence of domains within the SaPIbov1 Stl and testing whether one is responsible for dimeric Dut interaction and another for trimeric Duts would clarify this. This will be addressed in Chapter 6.

Chapter 6 Characterisation of the SaPI_{bov1} Stl repressor

Results from this section are included in:

Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer (2017) **Bowring J***, Neamah MM*, Donderis J*, Mir-Sanchis I, Alite C, J Ciges-Tomas R, Maiques E, Mehmedov I, Marina A, Penadés JR. *eLIFE*. DOI: 10.7554/eLife.26487

Disclaimer on the work performed:

The structural modelling in this chapter was performed in conjunction with and under the guidance of Dr Alberto Marina and his laboratory group. The *in vitro* SaPI_{bov1} Stl truncation work referred to in the discussion (Section 6.3) was performed by the Marina laboratory group. The dimeric Dut:Stl_{SaPI_{bov1}} work referred to in the discussion (Section 6.3) was also performed by Dr Alberto Marina and his laboratory group.

6.1 Background

As key mobile genetic elements within the wider horizontal transfer map, understanding the transfer of the PICIs and bacteriophages is vital for establishing methods of virulence factor transfer in bacteria. Furthermore, given the inherent association between the PICIs and bacteriophages^{270,283,305}, understanding the interactions between these two MGEs is necessary to fully understand PICI spread. One key interaction between the PICIs and the bacteriophage is the protein:protein interaction that triggers induction of the PICI ERP cycle^{269,280}. This involves PICI-specific phage-encoded derepressor proteins binding to the PICI repressor (Stl)^{280,305,359}.

The Stl repressor of the well-characterised SaPIbov1 island has been shown to bind both the dimeric and trimeric Duts encoded by *S. aureus* phages for its induction (Chapter 3)^{305,359}. These two Dut types perform the same function for the phage, but are completely divergent in sequence and structure (Chapter 3 and 4)^{304,359}. Previous analysis has shown that there are no overlapping residues between the dimeric and trimeric Duts that the SaPIbov1 Stl could exploit to achieve interactions with both Duts using the same repressor domain (Chapter 4). As such, the simplest explanation for the SaPIbov1 Stl repressor interacting with two such different proteins is that the repressor encodes two domains, one for interacting with each Dut family. However, the mechanism by which the SaPIbov1 Stl repressor binds to these different proteins is currently unclear.

Indeed, none of the structures of the different SaPI Stl repressors have been solved and little has been established regarding the domains involved in Dut interaction. Studies have established the presence of a helix-turn-helix (HTH) domain in the SaPIbov1 Stl that is involved in DNA binding for island repression, and have identified the DNA binding sites in the integrated island to which the repressors bind^{269,280,399,404}. The HTH domain has been found to be fairly conserved throughout the different Stl repressors for multiple SaPIs, although the rest of these proteins are very divergent³⁹⁹. This seems to indicate that the DNA binding domain of the different SaPIs is conserved, whilst the regions interacting with various phage proteins have diverged. However, the questions of which domains in the SaPIbov1 Stl interact with the different Dut proteins and

how this Stl achieves both island repression and phage-inducer interaction are yet to be addressed.

The aim of this study was to further characterise the different domains of the SaPIbov1 Stl and to investigate the roles of these different Stl domains through point mutation. Using structural predictions of the SaPIbov1 Stl repressor, different areas of interest were established for DNA recognition, repressor dimerization, and Stl:Dut interaction. Within these areas of interest, residues that were predicted to have a key role for the protein were mutated to investigate the effect of this on island repression and induction.

This study characterises three distinct regions within the SaPIbov1 Stl, one forming the N-terminal HTH domain and the other two making up the C-terminal region, joined together by a linking region. A number of residues were identified as responsible for DNA binding for SaPIbov1 repression, all within the established HTH domain. The study also found significant changes in SaPI repression and Dut interaction when mutating residues predicted to enable Stl dimerization. Finally, the study confirmed the existence of multiple interacting domains in the SaPIbov1 Stl repressor.

6.2 Results

6.2.1 The predicted structure of the SaPIbov1 Stl repressor

The structure of the SaPIbov1 Stl repressor has not yet been solved, making it more difficult to elucidate the residues involved in binding the different phage Duts for island induction. Previous data for the SaPIbov1 Stl repressor indicates the presence of both monomers and dimers under native mass spectrometry experimental conditions³⁶⁸. Previous sequence analysis and *in silico* modelling has mainly focused on the HTH region responsible for DNA recognition³⁹⁹. Outwith this conserved domain little has been proposed for potential regions for interaction with phage proteins or for any oligomerization. Here both I-TASSER and Phyre2 servers were used to analyse the predicted structure of the SaPIbov1 Stl repressor with an aim to identifying areas of interest for phage-encoded Dut interaction, DNA recognition or oligomerization (Fig 6.1)^{369-371,388}.

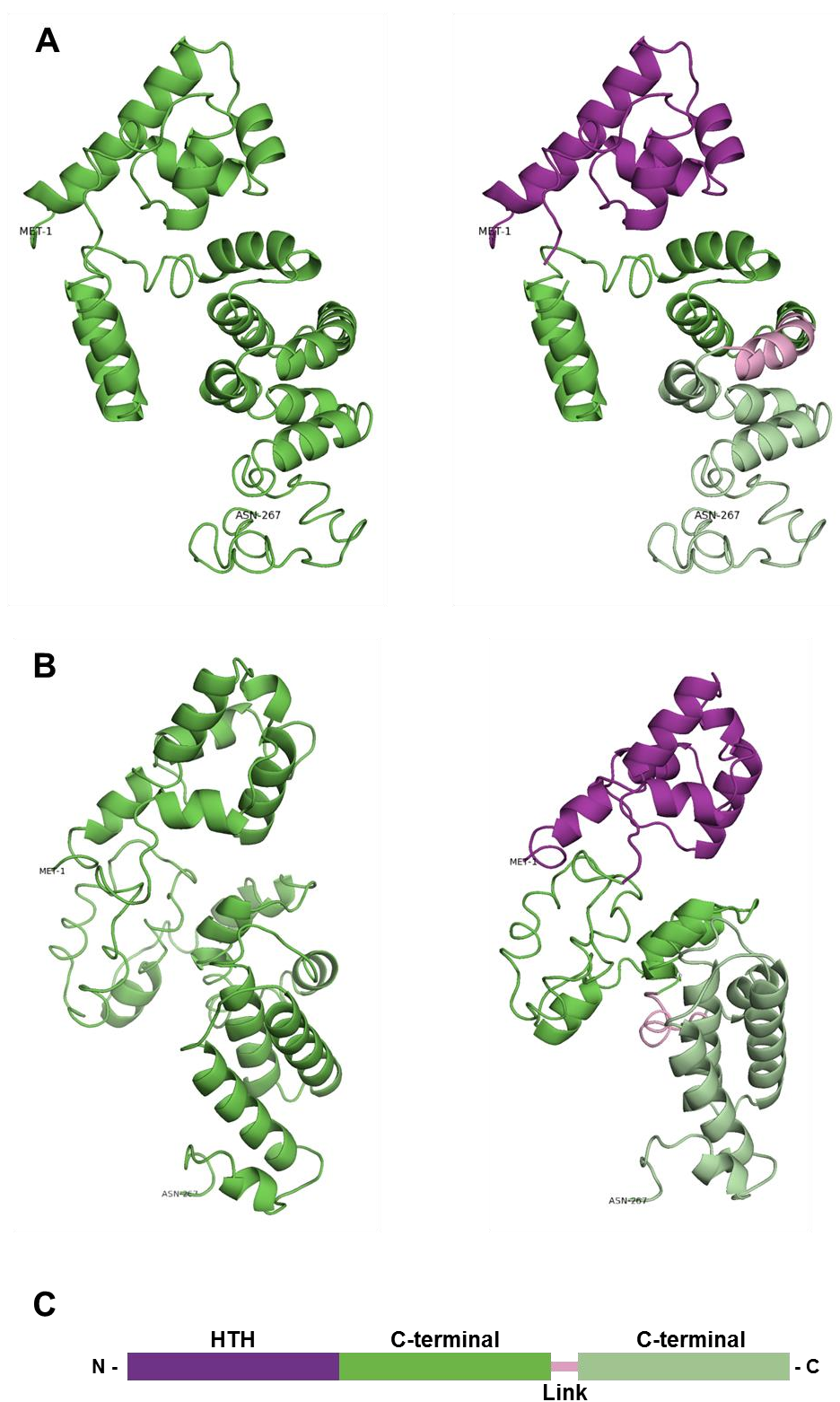


Figure 6.1 Predicted structures of the SaPIbov1 StI repressor protein constructed by the I-TASSER and Phyre2 servers. Cartoon representations of the monomer 3D structures of the SaPIbov1 StI repressor as generated by I-TASSER³⁶⁹⁻³⁷¹ (**A**) and Phyre2³⁸⁸ (**B**). The right-hand images are coloured by predicted domains as shown in **C**; the helix-turn-helix containing N-terminal in purple, C-terminal in dark and light green, and the C-terminal linking region in pink. I-TASSER and Phyre2 templates and confidence values can be found in Appendix 12.

Similar to previous analyses, both SaPIbov1 Stl predictions indicated a mainly α -helical protein containing an N-terminal HTH DNA binding domain (residues 1-80) and a C-terminal region of unknown function³⁹⁹. Both the I-TASSER and Phyre2 predictions of the SaPIbov1 Stl seem to indicate three domains in total, with the unknown C-terminal made up of two potential domains connected by a linking region (C-terminal 1, residues 81-166; linking region, residues 167-179; C-terminal 2, 180-267, Fig 6.1). The linking region was identified using the GeneSilico Metadisorder server to study the disorder tendency of the protein, which indicated residues ~167-179 as a disordered region of low complexity (Fig 6.2).

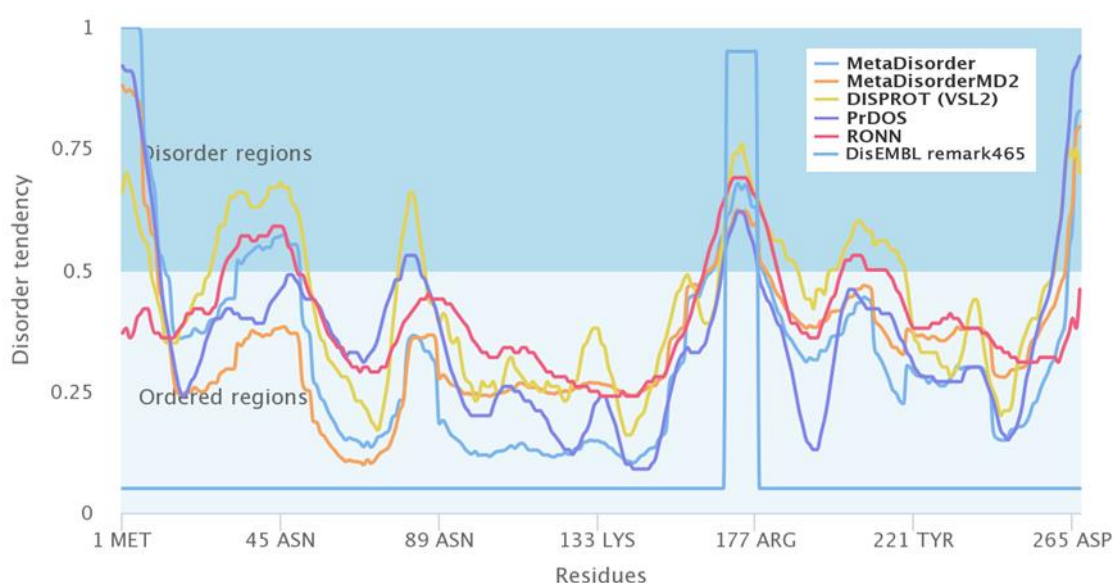


Figure 6.2 GeneSilico Metadisorder server protein sequence prediction of any disordered regions in the SaPIbov1 Stl repressor. Displaying the residue number along the x-axis and the disorder tendency (0-1) along the y-axis. Scores above 0.5 are considered disordered. Multiple methods for disorder prediction were used by the server and are indicated by the different coloured lines as described in the embedded key^{405–409}.

These predictions further clarify the potential domains within the SaPIbov1 Stl repressor, indicating that the large C-terminal region likely contains two domains in addition to the previously identified N-terminal HTH DNA binding domain. Using these predictions this study set out to establish specific residues of interest in order to investigate the DNA binding, oligomerization and Dut interaction of the SaPIbov1 Stl repressor.

6.2.2 Identifying the SaPIbov1 Stl residues responsible for DNA recognition for SaPI repression

The predicted structures in this study indicate a HTH DNA binding domain occupying the N-terminal between residues 1-80. Previous studies identified this N-terminal HTH region as vital for SaPIbov1 Stl repression of the island by DNA binding³⁹⁹. As such, the next step was to identify key residues within this domain that are responsible for the DNA binding function of the SaPIbov1 Stl. Previous work had identified a double mutation (Q40A,N41A) with greatly reduced affinity for the *stl* - *str* intergenic region of DNA, where the SaPIbov1 Stl normally binds³⁹⁹. However, it was unlikely that these were the only residues of importance for DNA binding and this study set out to identify other DNA binding residues.

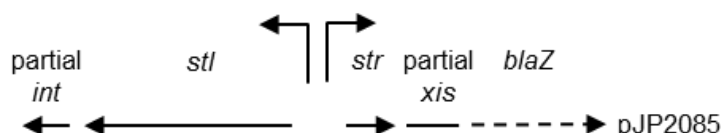


Figure 6.3 Schematic of the pJP2085 transcriptional fusion plasmid containing the regulatory module of the SaPIbov1, including part of the *integrase* gene, the *stl* repressor, the *str* anti-repressor, and a partial *xis* fused to a *blaZ* reporter gene. All the clones referred to in this chapter were constructed by cloning point mutations into the Stl of this fusion plasmid.

Modelling performed by the Marina group superimposed the SaPIbov1 Stl with the HTH domain from the controller protein of the Esp1396I restriction-modification system in *Enterobacter* sp. rfl1396, the structure of which is shown bound to DNA (PDB 3CLC)⁴¹⁰. From this, a series of point mutations were chosen that were predicted to prevent the DNA binding of the SaPIbov1 Stl and, as such, prevent SaPIbov1 repression. To test whether these mutants were capable of repressing SaPIbov1, the plasmid pJP2085 was utilised. This plasmid carries a β -lactamase reporter gene fused to *xis*, downstream of the encoded Stl-repressed *str* promoter and *str* (Fig 6.3). The different SaPIbov1 Stl point mutations were cloned into variations of this plasmid (Table 2.2). These plasmids were introduced into strains lysogenic for phages encoding either dimeric (ϕ NM1) or trimeric Duts (ϕ 11 and 80 α). This allowed analysis of the impact of phage induction on SaPIbov1 repression, but also of the impact of the point mutations. If the point mutations had the expected effect then the Stl would be unable to

repress the SaPIbov1 and the β -lactamase expression levels would be high even without phage induction.

Figure 6.4 shows the levels of β -lactamase expression with or without phage induction using MC ($2 \mu\text{g ml}^{-1}$) for the different point mutations targeting SaPIbov1 StI DNA binding. The β -lactamase expression levels can be interpreted as a surrogate for the expression of the StI-repressed *str* and *xis* genes, following derepression of StI from the intergenic region. Table 6.1 shows a list of all the point mutations, the effects seen on the β -lactamase levels compared to the wt and the significance and *p* values in comparison to the wt plasmid β -lactamase levels, with significant point mutants highlighted in yellow.

A number of point mutations led to significantly high levels of β -lactamase expression in absence of phage induction. These were point mutants R22A, Q29A, E47A, N48A, and N45A. In all these cases, the phage present in the strain did not affect the β -lactamase expression levels and the β -lactamase levels in the non-induced samples were significantly higher than with the wt (Fig 6.4 and Table 6.1). This indicates that the StI proteins with these point mutations were unable to bind the intergenic region of DNA and so could not repress the downstream genes, leading to unregulated β -lactamase expression. However, whether this is due to the residues mutated being vital for DNA binding or if it is because of protein misfolding/non-expression is not clear with this plasmid system. To confirm whether it is the former, the expression and aggregation of the point mutants needs to be checked.

Similarly, the N50A mutation seems to have leaky expression, with some β -lactamase expression in all the phage strains without induction and significantly higher levels in the induced sample of the ϕ NM1 lysogenic strain. This mimics the results seen for point mutant F38A, which also displays leaky β -lactamase expression in the non-induced samples, albeit at non-significant levels. This could indicate that these residues have some involvement in the DNA binding of the SaPIbov1 StI, but are not vital for this interaction.

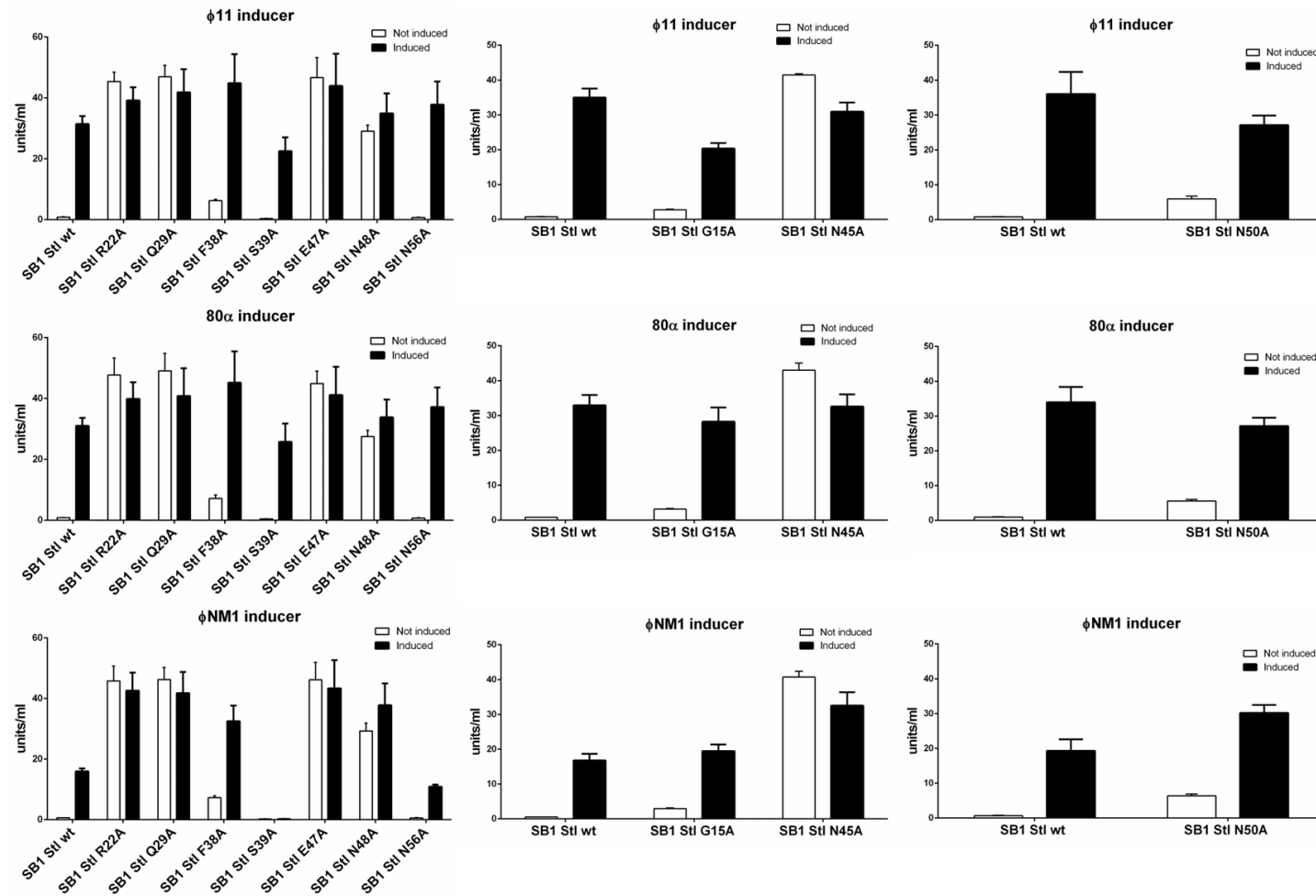


Figure 6.4 Identifying residues responsible for DNA recognition in the SaPIbov1 Stl repressor. Dut-encoding phage expression leading to the release of Stl repression of downstream genes. β -lactamase assays were performed on strains containing pJP2085 (see Fig 6.2) or equivalent Stl point mutant plasmids (Table 2.2) and positive for trimeric Dut encoding phages $\Phi 11$ or 80α or dimeric Dut encoding phage $\Phi NM1$. Samples were taken after 90 mins in the absence or following induction with MC. All data is the result of three independent experiments. Error bars represent standard error of the mean. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences between the wt and point mutants. Adjusted p values are shown in Table 6.1.

Table 6.1 List of point mutations shown in Fig 6.3 and predicted to affect SaPIbov1 StI DNA recognition, as cloned into the SaPIbov1 StI of transcriptional fusion vector pJP2085. The table indicates the effects seen and the *p* values in comparison to the wild-type vector with and without induction of different lysogenic phages encoding dimeric or trimeric Duts, as indicated. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences between the wt and point mutants. All *p* values have been adjusted for multiple comparisons. Significant mutations have been highlighted in yellow.

Phage (lysogenic)	StI point mutation (pCN41)	Effect	Not induced		Induced	
			Sig. diff. to wt	<i>p</i> value	Sig. diff. to wt	<i>p</i> value
Φ11 (Trimeric Dut)	R22A	No repression	Yes	< 0.0001 ^{****}	No	0.9335 ^{ns}
	Q29A	No repression	Yes	< 0.0001 ^{****}	No	0.7589 ^{ns}
	F38A	Leaky repression	No	0.9909 ^{ns}	No	0.4793 ^{ns}
	S39A	Reduced induction	No	> 0.9999 ^{ns}	No	0.8755 ^{ns}
	E47A	No repression	Yes	< 0.0001 ^{****}	No	0.5637 ^{ns}
	N48A	No repression	Yes	0.0045 ^{**}	No	0.9994 ^{ns}
	N56A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9753 ^{ns}
	G15A	Reduced induction/leaky repression	No	0.6489 ^{ns}	Yes	< 0.0001 ^{****}
	N45A	No repression	Yes	< 0.0001 ^{****}	No	0.1992 ^{ns}
	N50A	Leaky repression	No	0.5484 ^{ns}	No	0.2024 ^{ns}
80α (Trimeric Dut)	R22A	No repression	Yes	< 0.0001 ^{****}	No	0.8898 ^{ns}
	Q29A	No repression	Yes	< 0.0001 ^{****}	No	0.8298 ^{ns}
	F38A	Leaky repression	No	0.9811 ^{ns}	No	0.4470 ^{ns}
	S39A	Reduced induction	No	> 0.9999 ^{ns}	No	0.9935 ^{ns}
	E47A	No repression	Yes	< 0.0001 ^{****}	No	0.8068 ^{ns}
	N48A	No repression	Yes	0.0113 [*]	No	0.9999 ^{ns}
	N56A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9828 ^{ns}
	G15A	Leaky repression	No	0.7917 ^{ns}	No	0.4134 ^{ns}
	N45A	No repression	Yes	< 0.0001 ^{****}	No	0.9950 ^{ns}
	N50A	Leaky repression	No	0.4053 ^{ns}	No	0.1722 ^{ns}
ΦNM1 (Dimeric Dut)	R22A	No repression	Yes	< 0.0001 ^{****}	Yes	0.0009 ^{***}
	Q29A	No repression	Yes	< 0.0001 ^{****}	Yes	0.0014 ^{**}
	F38A	Leaky repression	No	0.9191 ^{ns}	No	0.0854 ^{ns}
	S39A	No induction	No	> 0.9999 ^{ns}	No	0.1175 ^{ns}
	E47A	No repression	Yes	< 0.0001 ^{****}	Yes	0.0006 ^{***}
	N48A	No repression	Yes	0.0003 ^{***}	Yes	0.0093 ^{**}
	N56A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9802 ^{ns}
	G15A	Leaky repression	No	0.6534 ^{ns}	No	0.5992 ^{ns}
	N45A	No repression	Yes	< 0.0001 ^{****}	Yes	0.0003 ^{***}
	N50A	Leaky repression	No	0.1593 ^{ns}	Yes	0.0104 [*]

There were some unpredicted effects from three point mutations. In the strain lysogenic for Φ11, which encodes a trimeric Dut, the G15A mutant displayed significantly lower β-lactamase levels than the wt following induction of the phage. This could indicate that the Φ11 Dut is less efficient at binding the

SaPIbov1 Stl containing this mutant or that the Stl is bound to the DNA more strongly than the wt. However, considering that this effect is confined to the $\phi 11$ strain and that there appears to be some leaky β -lactamase expression when the phage is not induced, this seems more likely to be due to the Dut:Stl interaction.

Additionally, the point mutation S39A appears to have reduced levels of β -lactamase expression in all three lysogenic strains, although these reductions are not significantly different to wt levels. The difference seen with the ϕ NM1 strain is striking, abolishing β -lactamase expression levels. Given that some reduction is seen with both the dimeric and trimeric encoding Duts, this could be due to the Stl binding to the DNA with greater affinity making it more difficult to derepress. However, the effects are more dramatic in the strain with the dimeric Dut encoding phage, suggesting some differences between the two Dut types in their ability to derepress the island.

The mutation N56A also seems to reduce the levels of β -lactamase expression in the dimeric Dut encoding phage strain, but not in the trimeric Dut encoding strains. This effect is less dramatic than with the S39A mutation. Without further investigation it is difficult to understand the cause of these differences. However, it is interesting that both these mutations seem to impact one Dut type more than the other. This supports the hypothesis that the Stl interacts with the dimeric and trimeric Dut differently.

Given that these point mutations were based on a superimposition of the Stl with another HTH-encoding transcriptional regulator interacting with DNA, it is unsurprising that not all the mutants had the expected effect. However, this study has identified five different residues that seem to be vital for the SaPIbov1 Stl DNA binding function. Further analysis is needed to confirm that the effects seen from these point mutations are due to an inability to bind the intergenic region of DNA rather than any protein misfolding. Furthermore, additional analysis into the mutations with unpredicted effects (G15A, S39A, and N56A) would help increase understanding of the cause of these results. This data gives some indication of the regions and potential residues of importance for SaPIbov1 island repression.

6.2.3 Establishing the SaPIbov1 Stl residues involved in repressor dimerization

The oligomerisation state of the SaPIbov1 Stl seems to impact on Stl function. Initial structural indications for the phage-encoded trimeric Dut:SaPIbov1 Stl interaction are that the Stl interacts with the trimeric Duts as a monomer³⁰⁴. On the other hand, the mechanism of transcriptional repressor proteins has traditionally been to bind two palindromic sequences as a dimer, correlating with the identification of two inverted repeat regions acting as SaPIbov1 Stl binding sites in the SaPI DNA sequence^{404,410}. One explanation for this monomer-dimer switch is that the inducing phage Duts disturb the Stl homodimer interface as well as overcoming the Stl:DNA interaction. Identifying the residues responsible for forming the SaPIbov1 Stl homodimer would help to understand more about this monomer-dimer switch.

As before, modelling performed by the Marina group superimposed the SaPIbov1 Stl with the controller protein of the Esp1396I restriction-modification system in *Enterobacter* sp. rfl1396 (PDB 3CLC)⁴¹⁰. One region of the SaPIbov1 Stl overlapped with part of the controller protein dimerization interface. From this, four point mutations were chosen that were predicted to affect the SaPIbov1 Stl dimerization: Y70A, R74A, D77A, and H73C. To test the effect of these mutants on SaPIbov1 repression and induction, derivatives of plasmid pJP2085 were utilised (Fig 6.3). The different SaPIbov1 Stl point mutations were cloned into variations of this plasmid (Table 2.2). These plasmids were again introduced into strains lysogenic for phages encoding either dimeric (ϕ NM1) or trimeric Duts (ϕ 11 and 80 α). This allowed analysis of the impact of phage induction on SaPIbov1 repression, but also of the impact of the point mutations on the Stl.

Given that the Stl is thought to repress SaPIbov1 as a dimer, if these mutations disrupted dimerization this could prevent the normal repression of the island. Similarly, if these mutants were still capable of repressing the island in some way as monomers, potentially they could better interact with the phage-encoded Duts. One of the mutants for this dimerization study (H73C) was designed to form a locked dimer due to a cysteine-cysteine di-sulphide bond. As such, it would be expected to have normal DNA binding, but reduced interaction with the phage Duts.

Figure 6.5 shows the levels of β -lactamase expression with or without phage induction using MC ($2 \mu\text{g ml}^{-1}$) for the different point mutations targeting SaPIbov1 Stl dimerization. The β -lactamase expression levels can be interpreted as a surrogate for the expression of the Stl-repressed *str* and *xis* genes, following derepression of Stl from the intergenic region. Table 6.2 shows a list of all the point mutations, the effects seen on the β -lactamase levels compared to the wt and the significance and *p* values in comparison to the wt plasmid β -lactamase levels, with significant point mutants highlighted in yellow.

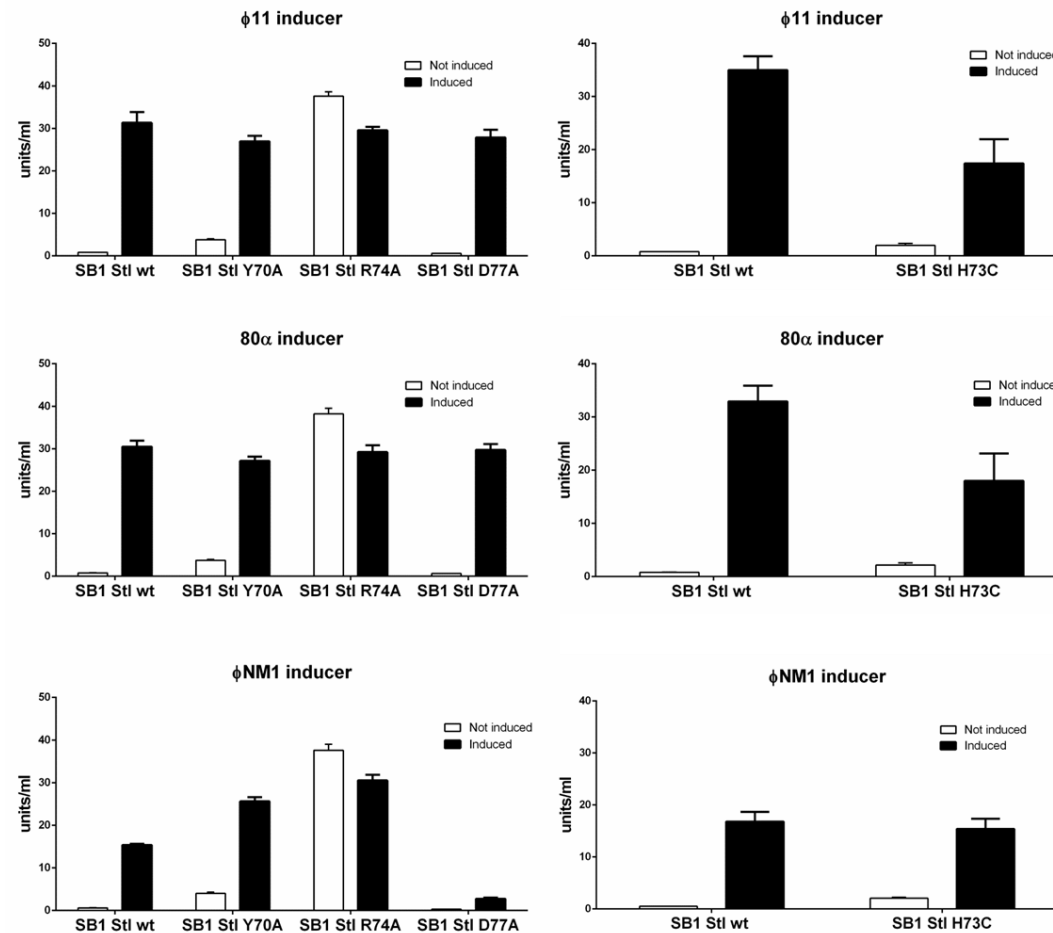


Figure 6.5 Identifying residues responsible for repressor dimerization in the SaPIbov1 StI repressor. Dut-encoding phage expression leading to the release of StI repression of downstream genes. β -lactamase assays were performed on strains containing pJP2085 (see Fig 6.2) or equivalent StI point mutant plasmids (Table 2.2) and positive for trimeric Dut encoding phages Φ 11 or 80α or dimeric Dut encoding phage Φ NM1. Samples were taken after 90 mins in the absence or following induction with MC. All data is the result of three independent experiments. Error bars represent standard error of the mean. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences between the wt and point mutants. Adjusted p values are shown in Table 6.2.

Table 6.2 List of point mutations shown in Fig 6.4 and predicted to affect SaPIbov1 Stl repressor dimerization, as cloned into the SaPIbov1 Stl of transcriptional fusion vector pJP2085. The table indicates the effects seen and the *p* values in comparison to the wild-type vector with and without induction of different lysogenic phages encoding dimeric or trimeric Duts, as indicated. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences between the wt and point mutants. All *p* values have been adjusted for multiple comparisons. Significant mutations have been highlighted in yellow.

Phage (lysogenic)	Stl point mutation (pCN41)	Effect	Not induced		Induced	
			Sig. diff. to wt	<i>p</i> value	Sig. diff. to wt	<i>p</i> value
Φ11 (Trimeric Dut)	Y70A	Leaky induction	No	0.3021 ^{ns}	No	0.0724 ^{ns}
	R74A	No repression	Yes	< 0.0001 ^{****}	No	0.7146 ^{ns}
	D77A	wt phenotype	No	0.9988 ^{ns}	No	0.1954 ^{ns}
	H73C	Reduced induction	No	0.9395 ^{ns}	Yes	0.0030 ^{**}
80α (Trimeric Dut)	Y70A	Leaky induction	No	0.1803 ^{ns}	No	0.1176 ^{ns}
	R74A	No repression	Yes	< 0.0001 ^{****}	No	0.8003 ^{ns}
	D77A	wt phenotype	No	0.9997 ^{ns}	No	0.9460 ^{ns}
	H73C	Reduced induction	No	0.9376 ^{ns}	Yes	0.0152 [*]
ΦNM1 (Dimeric Dut)	Y70A	Leaky induction	Yes	0.0168 [*]	Yes	< 0.0001 ^{****}
	R74A	No repression	Yes	< 0.0001 ^{****}	Yes	< 0.0001 ^{****}
	D77A	No induction	No	0.9886 ^{ns}	Yes	< 0.0001 ^{****}
	H73C	wt phenotype	No	0.6850 ^{ns}	No	0.7382 ^{ns}

Of the point mutations predicted to affect SaPIbov1 Stl dimerization, the R74A point mutation seems to have prevented the Stl from being able to bind the DNA region responsible for island repression. As such, the β -lactamase expression levels without any phage induction are significantly higher than for the wt plasmid in all strains. This could be because the Stl cannot form a dimer and so is unable to bind both palindromic sequences for island repression, but the possibility that the protein cannot fold correctly needs to be ruled out.

Similarly, the ‘locked-dimer’ H73C mutant shows reduced β -lactamase expression in the induced samples for the strains lysogenic for the trimeric Dut encoding phages. However, there is still some induction between the trimeric Duts and this mutant, causing moderate levels of β -lactamase expression. Furthermore, this effect is not seen at all in the strain with the dimeric Dut encoding phage. Indeed, although this mutant was designed to lock the Stl as a dimer, this is dependent on oxidant conditions, whilst the intracellular environment has more reductant conditions so potentially this mutant is still able to form a monomer.

The Y70A mutant appears to have some leaky repression, with β -lactamase expression seen in the non-induced samples. These expression levels are seen in each strain although they are only significant in the dimeric Dut encoding ϕ NM1 strain. However, this could indicate that the Y70A Stl mutant does not form as stable a dimer as the wt, leading to some derepression.

The D77A had some unpredicted effects, with the mutant showing a wt phenotype for the trimeric Dut-encoding phage strains, but having significantly lower levels of β -lactamase expression for the induced sample of the dimeric encoding ϕ NM1 strain. Potentially this residue could be involved in the SaPIbov1 Stl interaction with the dimeric Duts, but not with the trimeric Duts. This, therefore, supports the existence of multiple interaction domains in the SaPIbov1 Stl repressor.

Similar to the previous section (6.2.2), these point mutations were based on superimposition predictions. However, the majority of the point mutations seemed to work as expected, albeit to a lesser or greater extent and with some unexplained variations. Further analysis is needed to confirm that the effects seen in the Y70A and R74A point mutations are due to a lost or lesser ability to form the dimer needed for binding the intergenic region of DNA, rather than any protein misfolding. Furthermore, identifying the oligomerisation states of each mutant would help with the interpretation of these results. The unexpected result with the D77A mutant should be further investigated, both by checking the protein structure and oligomerisation state and by recombining this mutation into a SaPIbov1-positive strain to analyse the effect on dimeric Dut interaction. Overall, this data indicates that this region may be involved in SaPIbov1 Stl dimerization and identifies residues for further study.

6.2.4 Determining SaPIbov1 Stl residues that interact with the phage-encoded trimeric Duts

Having identified a number of residues implicated in both the DNA binding and dimerization of the SaPIbov1 Stl, the next step was to see if it was possible to identify residues that affected the Stl interaction with the trimeric Duts and not the dimeric Duts. Based on the hypothesis that the SaPIbov1 Stl encodes multiple domains, including one targeting the dimeric Duts and another targeting the

trimeric Duts, such residues should exist. Identifying which regions of the SaPIbov1 Stl target the different dimeric and trimeric phage Duts would help elucidate the mechanism by which this protein can target such structurally different proteins for its derepression.

Based on structural modelling done by the Marina group on the interaction of the trimeric Dut from phage $\phi 85$ and the SaPIbov1 Stl, a number of residues in the SaPIbov1 Stl were identified as potentially important for trimeric Dut:SaPIbov1 Stl interaction. These were as follows; the Y106 residue was predicted to interact with the trimeric Dut Motif I, the Y112 and Y113 residues that were predicted to interact with the trimeric Dut active centre, the Y116 residue was predicted to have contact with the β -hairpin of the trimeric Dut motif VI, the D117 residue was predicted to interact with the Motif IV, and the D155 was residue predicted to interact with the trimeric Dut motif VI.

To test whether these mutants were capable of interacting with the phage-encoded dimeric and trimeric Duts, the plasmid pJP2085 was utilised (Fig 6.3). A series of different single and double SaPIbov1 Stl point mutations were cloned into variations of this plasmid (Table 2.2). These plasmids were introduced into strains lysogenic for phages encoding either dimeric (ϕ NM1) or trimeric Duts (ϕ 11 and 80 α). This allowed analysis of the impact of phage induction on SaPIbov1 repression, but also of the impact of the point mutations. If the point mutations had the expected effect then the SaPIbov1 would be de-repressed as normal when the dimeric Dut encoding phage was induced, but not de-repressed in the presence of induced trimeric Dut encoding phages.

Figure 6.6 shows the levels of β -lactamase expression with or without phage induction using MC ($2 \mu\text{g ml}^{-1}$) for the different point mutations targeting SaPIbov1 Stl:trimeric Dut interaction. The β -lactamase expression levels can be interpreted as a surrogate for the expression of the Stl-repressed *str* and *xis* genes, following derepression of Stl from the intergenic region. Table 6.3 shows a list of all the point mutations, the effects seen on the β -lactamase levels compared to the wt and the significance and *p* values in comparison to the wt plasmid β -lactamase levels, with significant point mutants highlighted in yellow.

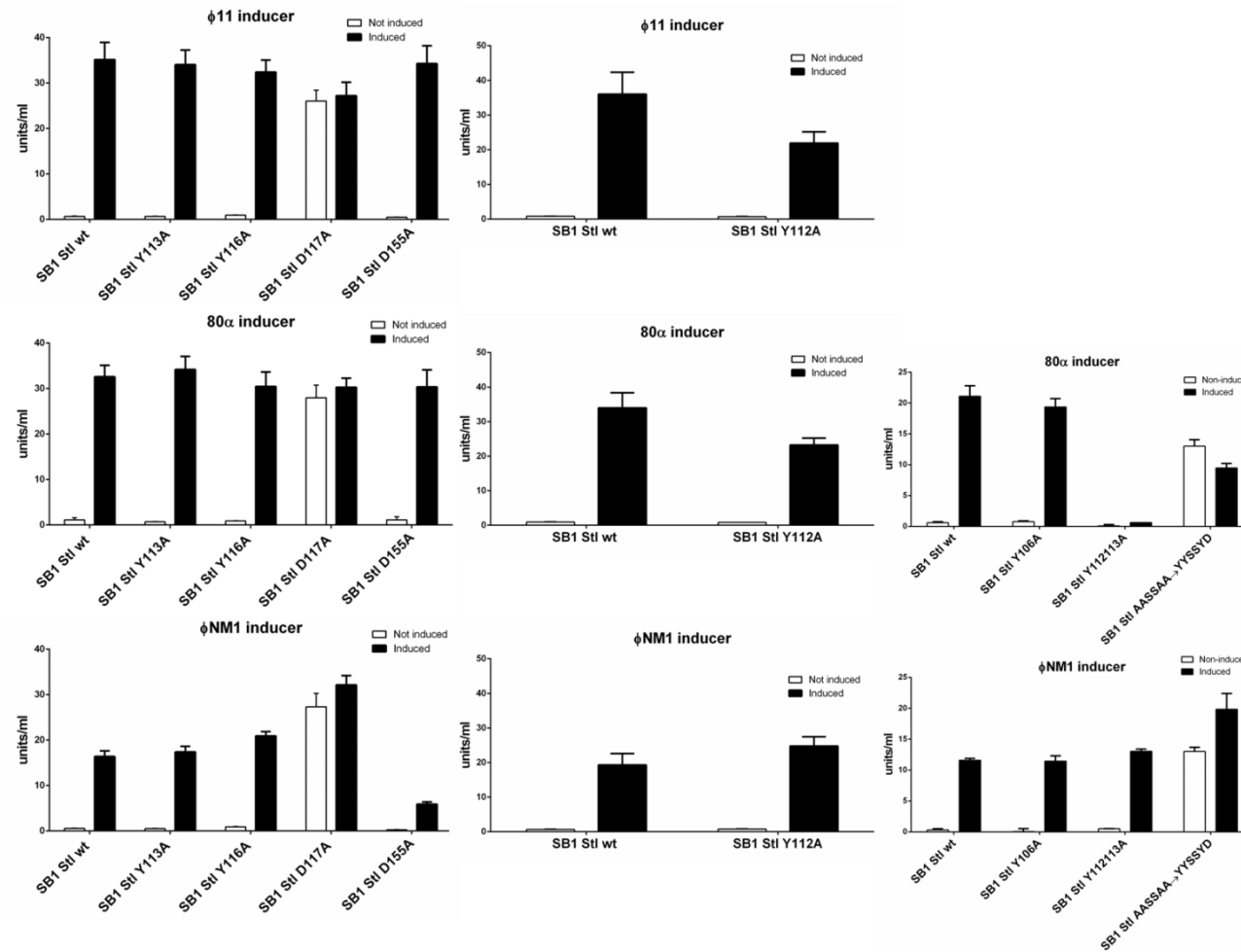


Figure 6.6 Identifying residues responsible for Dut interaction in the SaPIbov1 Stl repressor. Dut-encoding phage expression leading to the release of Stl repression of downstream genes. β -lactamase assays were performed on strains containing pJP2085 (see Fig 6.2) or equivalent Stl point mutant plasmids (Table 2.2) and positive for trimeric Dut encoding phages $\Phi 11$ or 80α or dimeric Dut encoding phage $\Phi NM1$. Samples were taken after 90 mins in the absence or following induction with MC. All data is the result of three independent experiments. Error bars represent standard error of the mean. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences between the wt and point mutants. Adjusted p values are shown in Table 6.3.

Table 6.3 List of point mutations shown in Fig 6.5 and predicted to affect Stl repressor interaction with the trimeric Duts, as cloned into the SaPIbov1 Stl of transcriptional fusion vector pJP2085. The table indicates the effects seen and the *p* values in comparison to the wild-type vector with and without induction of different lysogenic phages encoding dimeric or trimeric Duts, as indicated. A 2-way ANOVA with Sidak's multiple comparisons test was performed to compare mean differences between the wt and point mutants. All *p* values have been adjusted for multiple comparisons. Significant mutations have been highlighted in yellow.

Phage (lysogenic)	Stl point mutation (pCN41)	Effect	Not induced		Induced	
			Sig. diff. to wt	<i>p</i> value	Sig. diff. to wt	<i>p</i> value
Φ11 (Trimeric Dut)	Y113A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9961 ^{ns}
	Y116A	wt phenotype	No	> 0.9999 ^{ns}	No	0.8990 ^{ns}
	D117A	No repression	Yes	< 0.0001 ^{****}	No	0.1184 ^{ns}
	D155A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9984 ^{ns}
	Y112A	Reduced induction	No	0.9998 ^{ns}	Yes	0.0456 [*]
80α (Trimeric Dut)	Y113A	wt phenotype	No	0.9999 ^{ns}	No	0.9784 ^{ns}
	Y116A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9446 ^{ns}
	D117A	No repression	Yes	< 0.0001 ^{****}	No	0.9249 ^{ns}
	D155A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9330 ^{ns}
	Y112A	Reduced induction	No	0.9991 ^{ns}	Yes	0.0271 [*]
	Y106A	wt phenotype	No	0.9988 ^{ns}	No	0.4745 ^{ns}
	YY112113AA	No induction	No	0.9741 ^{ns}	Yes	< 0.0001 ^{****}
	112AASSAA→YYSSYD	No repression	Yes	< 0.0001 ^{****}	Yes	< 0.0001 ^{****}
ΦNM1 (Dimeric Dut)	Y113A	wt phenotype	No	> 0.9999 ^{ns}	No	0.9724 ^{ns}
	Y116A	wt phenotype	No	0.9996 ^{ns}	No	0.0761 ^{ns}
	D117A	No repression	Yes	< 0.0001 ^{****}	Yes	< 0.0001 ^{****}
	D155A	Reduced induction	No	0.9998 ^{ns}	Yes	< 0.0001 ^{****}
	Y112A	wt phenotype	No	0.9992 ^{ns}	No	0.2037 ^{ns}
	Y106A	wt phenotype	No	0.9973 ^{ns}	No	0.9994 ^{ns}
	YY112113AA	wt phenotype	No	0.9988 ^{ns}	No	0.7028 ^{ns}
	112AASSAA→YYSSYD	No repression	Yes	< 0.0001 ^{****}	Yes	< 0.0001 ^{****}

Some of the point mutations failed to have the expected effect, with the Y106A, Y113A, and Y116A mutations all having a wt phenotype. However, the Y112A point mutation showed a significant decrease in β-lactamase levels in strains containing a trimeric Dut encoding phage, but a wt phenotype for the dimeric encoding phage strain. This effect correlates with the prediction of this residue being involved in trimeric Dut interaction and with the expectation that the SaPIbov1 Stl has different residues for interacting with dimeric or trimeric Duts.

In tandem with the Y112A single point mutation, this study looked at the effect of a double mutation including residues 112 and 113 (Y112Y113→A112A113). This double mutant had a significant effect on the β -lactamase levels for the induced 80 α trimeric Dut-encoding lysogenic strain. This mutation shows very low levels of β -lactamase expression for the strain with the trimeric Dut encoding phage, but a wt phenotype for the dimeric encoding phage strain. This supports the previous data that the Y112 residue is involved in interaction with the phage-encoded trimeric Duts, but not the dimeric Duts, and introduces Y113 as another important residue. Interestingly, the Y113A single point mutation showed a wt phenotype and the drop in β -lactamase levels was less dramatic for the Y112A single point mutation, indicating that potentially these residues complement each other; to completely abolish trimeric Dut:Stl_{SaPI_{bov1}} interaction both must be mutated. Furthermore, this data supports the idea that the SaPI_{bov1} Stl has one set of residues for trimeric Dut interaction, and another for dimeric Dut interaction.

The D155A mutation had an unexpected effect, with this mutant showing a significantly decreased level of β -lactamase expression for the induced sample of the strain lysogenic for the dimeric Dut encoding ϕ NM1. This effect is not seen in the strains lysogenic for trimeric Dut encoding phages, which display a wt phenotype. Although this was not the predicted effect of the mutation, it is interesting that this residue again seems to be involved in the interaction with one type of Dut, but not the other. This further supports the idea that the SaPI_{bov1} Stl has different residues for interacting with the dimeric and trimeric Duts.

The D117A mutation had the unexpected effect of preventing Stl repression of the SaPI_{bov1} island, leading to constantly high levels of β -lactamase expression in all strains, independent of phage induction. This effect was also seen in a larger mutation including the D117 residue (112YYSSYD→112AASSAA); the implication being that mutating the D117 residue has an effect on either the DNA-binding of the Stl, affects dimerization, or prevents the correct folding. That this residue is outside of the HTH region thought to be involved in DNA-binding suggests that the D117 residue is in some way important for the adoption of the correct conformation of the Stl or for dimerization. However, further

investigation is needed to understand why these D117 mutations have this effect.

This study establishes residues Y112 and Y113 as highly important for the SaPIbov1 Stl interaction with the phage-encoded trimeric Duts, but not for the interaction with the dimeric Duts. Further mutations changing the charge of these residues could increase the effects seen in the single point mutations, a tactic which could also be applied to the Y106 and Y116 residues. There is the possibility that the mutation to alanine is too conservative to overcome any interaction compensation from other residues and some of these residues could still be involved in the interaction with the trimeric Duts. The unexpected result with the D155A mutant should be further investigated; recombining this mutation into a SaPIbov1-positive strain would allow further analysis of the effect on dimeric Dut interaction. The protein expression and structure of the D117 mutant should also be investigated to establish why this mutant could not repress SaPIbov1. Overall, the data here supports the hypothesis that the SaPIbov1 Stl encodes different regions to interact with the different Dut types; one set of residues is important for the trimeric interaction, but not the dimeric, and vice versa.

6.3 Discussion

Understanding the function of the different SaPIbov1 Stl domains is key to elucidating the mechanism by which this protein performs three different interactions: DNA binding, dimeric Dut interaction, and trimeric Dut interaction. It has been established here and in the literature that the SaPIbov1 Stl contains a N-terminal HTH domain for DNA binding, but otherwise these interactions remain uncharacterised^{310,399}. This chapter aimed to characterise the role of the C-terminal region and establish what residues are important for DNA binding, Stl dimerization, and Dut interaction. This would help establish the mechanism by which the SaPIbov1 Stl binds the DNA region and the role of dimerization with regard to repression/induction. Furthermore, this would confirm whether the SaPIbov1 Stl interacts with the phage encoded dimeric Duts using one set of residues, whilst using another set to interact with the phage encoded trimeric Duts.

This study used structural predictions of the SaPIbov1 Stl to identify three different domains within the protein; one in the N-terminal encoding a HTH and two in the C-terminal linked by a disordered region. Following on from this, structural modelling work done by the Marina laboratory group identified a number of residues of interest, which were tested *in vitro* here. Of these, the β -lactamase expression tests identified a number of residues that followed the predictions. The R22, Q29, E47, N48, and N45 residues all seem to be important for DNA binding. The Y70, R74, and H73 residues are implicated in SaPIbov1 Stl dimerization to lesser or greater extents, although these results are less easily interpreted. Finally, the Y112 and Y113 residues appear vital for SaPIbov1 Stl interaction with the trimeric Duts, but not the dimeric Duts.

The identification of residues R22, Q29, E47, N48, and N45 as important for SaPIbov1 Stl DNA binding correlate with previous work that identified the N-terminal HTH and more specifically residues Q40 and N41 as important for DNA binding³⁹⁹. Furthermore, this matches other known transcriptional regulators that encode a HTH domain and form dimers to bind two palindromic sequences to repress transcription⁴¹⁰.

This leads on to the role of Stl dimerization in DNA binding and the possibility of a dimer-monomer switch for release of DNA binding and interaction with the Dut proteins. It is expected that the SaPIbov1 Stl undergoes DNA binding as a dimer, allowing two HTH regions to interact with the two palindromic sequences that have been identified at the *stl* - *str* intergenic region⁴⁰⁴. However, the phage-encoded trimeric Duts are thought to interact with the SaPIbov1 Stl by binding three Stl monomers³⁰⁴. Furthermore, recent preliminary structural work by members of the Marina laboratory has confirmed that not only do the trimeric Duts interact with monomeric Stl, but the dimeric Duts also bind a monomeric Stl. These initial results indicate that the SaPIbov1 Stl dimer-monomer switch has an important functional role for switching between DNA binding and Dut interaction. As such, understanding what residues are involved in this dimerization could help explain this process. Here, three residues are identified as potential points of dimerization: Y70, R74, and H73.

The preliminary work done solving the structure of the dimeric ϕ O11 Dut bound to the C-terminal end of the SaPIbov1 Stl indicates that the Stl mimics a Dut

monomer in order to interact with a Dut monomer in the form of a heterodimer. This correlates with the work from Chapter 4, which implicated the dimeric Dut dimerization interface as important for Stl binding. This also matches published results indicating that a SaPIbov1 Stl N-terminal truncation mutant (residues 1-86; Stl^{ΔHTH}) could interact with the dimeric Duts but not the trimeric Duts, whilst an extreme C-terminal truncation mutant (residues 176-267; Stl^{ΔCter}) could interact with the trimeric Duts and not the dimeric Duts³¹⁰. Attempts to replicate this experiment using the plasmid system described in this chapter were unsuccessful, as the truncated SaPIbov1 Stl was no longer capable of repressing the island.

These results support the hypothesis that the SaPIbov1 Stl has different residues and domains for interacting with the dimeric Duts and the trimeric Duts. However, there seem to be some discrepancies between the residues identified here as important for Dut interaction and those implicated in the truncation and preliminary data just described. The N-terminal has been implicated in the interaction with the trimeric Duts, but here we establish that residues Y112 and Y113 seem vital for this interaction. Furthermore, the HTH makes up the majority of the N-terminal and has been established as the region responsible for DNA binding. Similarly, preliminary results indicate the most C-terminal region is the main region of interaction with the dimeric Duts, but here we find some indication that the D77 and D155 may play a role in this interaction. It seems there is some nuance to the SaPIbov1 Stl mechanism of interaction between the dimeric and trimeric Duts that is not yet clear. Confirming the results found here by recombining the mutations into the SaPIbov1 itself and further structural work solving the interactions between the full SaPIbov1 Stl protein and the different Duts would help fully establish the mechanisms of Stl:Dut interaction.

6.4 Conclusions and future work

Establishing the mechanism by which the SaPIbov1 Stl is able to repress SaPIbov1 through DNA binding, whilst also interacting with the structurally divergent dimeric and trimeric Duts is vital to understanding the efficient transfer of SaPIbov1. This study identifies a number of residues of interest to be taken forward for further study: DNA binding residues R22, Q29, E47, N48, and N45, dimerization residues Y70, R74, and H73, and trimeric Dut interaction residues

Y112 and Y113. For all of these mutations the expression and folding of the SaPIbov1 Stl needs to be confirmed, to establish that any effects seen were not due to the protein not expressing or misfolding. Furthermore, by recombining these mutations into a SaPIbov1-positive strain it would be possible to check replication and transfer of these SaPIbov1 mutants. This should also be done for the mutants that had an unexpected effect; particularly G15, D77, D155, and D117. Finally, by continuing to try and solve the structure of the SaPIbov1 Stl and eventually to gain the structure of the full SaPIbov1 Stl interacting with the dimeric or trimeric Duts, the residues key to these different interactions will be far easier to determine.

Chapter 7 Concluding remarks

7.1 The dimeric Duts as a subset of SaPIbov1 inducers

Bacteria, both commensal and pathogenic, are highly successful at evolving under selective pressures to be maintained in changing environments and to colonise new niches. One system that allows rapid evolution in response to external pressures is HGT⁴¹¹. MGEs make up a vital aspect of this wider HGT map, playing a key role in the transfer of important resistance and virulence genes⁵⁵. This is particularly true of *S. aureus*, a species of bacteria in which ~20-22% of the genome is made up of MGEs and non-mobile GIs^{58,300,390,412,413}. Through these mobile elements the *S. aureus* has access to an interchangeable ‘genome library’, allowing acquisition of resistance and virulence genes that enable the bacteria to survive and flourish in different niches. As *S. aureus* evolves quickly using MGEs to adapt to changing ecological pressures, this differentiation of the genome creates diversity within clonal lineages as the bacteria facing different pressures diverge.

One important subset of the *S. aureus* MGEs is made up by the SaPIs, which are the best characterised of a wider MGE group known as the PICIs. The SaPIs are phage satellites that exploit the life cycle of the phages for their own replication and transfer^{262,270,278}. They encode clinically relevant virulence factors, including enterotoxins B and C and toxic shock syndrome toxin-1, as well as being implicated in antibiotic resistance transfer^{29,57,297,414}. The SaPIs reside stably in the *S. aureus* chromosome under control of the SaPI master repressor, Stl. Stl targets phage-encoded anti-repressor proteins for its own island derepression, further linking the SaPIs to the helper phage life cycle. Different SaPIs encode different Stl repressors, which are divergent in sequence to one another, and so each Stl targets a different phage protein for derepression. In this way, different islands are mobilised by different helper phages, depending on the SaPI Stl and the derepressor proteins encoded by that phage²⁸³.

The best characterised of the known SaPIs is the SaPIbov1 island, which carries the virulence genes for toxic shock syndrome toxin-1, enterotoxin C, and enterotoxin L²⁹⁷. Numerous investigations into the induction of SaPIbov1 by helper phages $\phi 11$ and 80 α have established that the phage proteins that bind the SaPIbov1 Stl for island derepression are the trimeric Duts encoded by these phages^{304,305,368}. Indeed, the most recent publication proposed the mechanism

for this interaction³⁰⁴. In this mechanism the trimeric Dut motifs IV and VI act as binding sites for the SaPIbov1 Stl. Stl binding is followed by a conformational change bringing the motif V from the neighbouring protomer over the bound Stl monomer to stabilise the interaction. Furthermore, these studies implicate the dUTP as a secondary signalling molecule that blocks the trimeric Duts from binding Stl to prevent the SaPI-inducing function^{304,368}. Finally, the studies also identified an unconserved central region, making up motif VI, that seems to impact the affinity of the different phage-encoded trimeric Duts for the SaPIbov1 Stl³⁰⁴.

Previous studies had identified four prophages (ϕ NM1-4) in the *S. aureus* Newman strain, a clinical isolate⁴¹⁵. The ϕ NM1 and ϕ NM2 phages were found to have similar structural proteins to the SaPIbov1-inducing 80 α , although the Dut encoded by these two phages was highly divergent from the 80 α trimeric Dut³¹⁷. Despite the differences seen in the Dut encoded by the ϕ NM1 and ϕ NM2 phages to that of 80 α , both these phages were able to induce SaPIbov1³¹⁷. Indeed, it was confirmed that the type 2 dimeric Dut encoded by ϕ NM1 and ϕ NM2 was sufficient to induce the SaPIbov1, raising the question of whether the staphylococcal phage-encoded dimeric Duts formed a second subset of SaPIbov1 inducers³⁵⁹.

Chapter 3 of this study describes the identification of 59 different phage-encoded dimeric Duts and their subsequent grouping into 9 families. Representative phage-encoded dimeric Duts from each family were analysed and predicted to exhibit a reduced protein in comparison with the solved structures of dimeric Duts from other organisms³⁵⁵⁻³⁵⁸. This was later confirmed once the structures of two of these phage-encoded dimeric Duts (ϕ O11 and ϕ DI) had been solved, as described in Chapter 4. These sequence and structural analyses confirmed that the phage-encoded dimeric and trimeric Duts exhibit totally different structures. However, in a surprising parallelism with the phage-encoded trimeric Duts, the protein sequences of the different dimeric Dut families exhibited conserved N- and C-terminals with sequence divergence mainly isolated to an unconserved central region. In the phage-encoded trimeric Duts this region of divergence seems to impact on the efficiency of the Dut:Stl interaction. The expectation based on this was that some of the phage-encoded dimeric Dut would induce the island more effectively than others.

This expectation was partly confirmed when the SaPIbov1-inducing capabilities of 6 different dimeric Dut protein representatives from different families were tested, although to an unpredicted extent. Of the dimeric Dut proteins tested only four were capable of SaPIbov1-induction, whilst the $\phi 55$ and ϕDII Duts were not. It seems that, like the trimeric Duts, the differences in this unconserved motif VI region contribute to differences in SaPIbov1 induction. However, whilst the differences in the $\phi 11$ and 80α trimeric Duts affect how well these Duts induce the island, with the phage-encoded dimeric Duts this seems to impact whether the Dut can induce the island or not. Perhaps testing more phage-encoded trimeric Duts would identify one that has evolved to escape SaPIbov1 induction, or maybe the dimeric Dut mechanism of interaction with Stl is more easily escaped. Either way, it is striking that two structurally different protein families both display a similar region of divergence that impacts their ability to induce SaPIbov1.

This data establishes the phage-encoded dimeric Duts as another subset of SaPIbov1 inducers, in addition to the previously characterised trimeric Duts. However, the dimeric Duts appear to divide into two groups: the non-inducers and the inducers. The non-inducers appear to have evolved to overcome SaPIbov1 parasitism by avoiding interaction with the SaPIbov1 Stl. Given the high cost for the phage following SaPI induction, it would be expected that a higher proportion of phage-encoded Duts would display this non-inducing phenotype. However, the fact that they do not supports the idea that the SaPIbov1 Stl is exploiting an unknown Dut function that confers a selective advantage for the phage. By mimicking a proposed Dut signalling partner the Stl could compete with this partner to interact with the Dut for island induction, utilising this competing affinity to prevent the Dut escaping Dut:Stl interaction due to the need to maintain the Dut:partner interaction. If this is the case, it could be difficult to escape island induction whilst maintaining this advantageous function. This possibility is supported by the fact that, although in laboratory conditions there seems to be no detrimental effect on the phage following Dut deletion, in nature Dut mutants are lost in favour of phages encoding a Dut protein²⁸¹.

Recently there have been a number of papers associating secondary signalling functions to Duts from different organisms. For example, different

immunoregulatory functions have been found for a number of Duts encoded by different members of the *Herpesviridae* family^{360,416,417}. This includes multiple studies on the dysregulation of the immune system by EBV, in one case leading to an exacerbated sickness response in mice^{330,418-421}. Furthermore, susceptibility to psoriasis has been associated with polymorphisms in the Dut encoded by human endogenous retrovirus K^{333,422,423}. These signalling functions are not confined to viruses, with the Dut encoded by rats identified as interacting with a transcriptional regulator, the peroxisome proliferator-activated receptors (PPARs)³³⁴. These studies indicate the potential of Dut proteins to exhibit moonlighting functions.

This supports the idea of competing priorities in which a potential signalling mechanism of the phage-encoded Duts is maintained, preventing the phage from completely escaping SaPI induction. However, if the Dut-associated signalling advantage is lost in the non-inducers, how then are the non-inducing dimeric Dut encoded phages maintained in nature? It could be that they occupy a niche in which the selective advantage is not as important as avoiding island induction, for example in a bacterial population encoding high numbers of SaPIs.

7.2 SaPIs exploit conserved phage mechanisms to achieve efficient intra- and interspecific transfer

Answering the question of how the phages and SaPIs interact and are maintained in nature remains key to understanding the full impact of these MGEs outside of the laboratory. The importance of SaPIs in conferring pathogenicity to *S. aureus* strains is epitomised by the example of SaPI transfer of toxic shock syndrome toxin 1²⁹. Transfer of this gene is capable of transforming a relatively benign strain into a pathogen capable of causing toxic shock syndrome²⁹. SaPIs also encode other virulence factors including a number of enterotoxin superantigens and resistance genes^{262,297}. The potential of SaPIs, and more widely PICs, to transfer between species in nature needs to be understood to establish the extent of virulence transfer via these MGEs.

Inter- and intra-specific transfer of SaPIs and even intergeneric transfer of SaPIs has been easily achieved under laboratory conditions. Studies have demonstrated the ability of both *pac* and *cos* phages to transfer different SaPIs

to a variety of staphylococcal species and even different genera of bacteria. Furthermore, the existence of PICI elements in a wide range of gram-positive bacteria has become increasingly apparent. Indeed the *Enterococcus faecalis* phage-inducible island, EfCIV583, has been confirmed as a functional PICI, innately linked to the helper phage (p1) life cycle in a similar way to that characterized for the SaPIs^{275,320}. The ease at which the SaPIs have been transferred to new species and genera in the laboratory, and the ever increasing range of bacteria identified as encoding PICI-like elements, has led to an assumption that inter- and intraspecific transfer is common in nature.

In Chapter 5 the reality of such transfer in nature is addressed, the results of which support the commonly held conception that SaPIs and PICIs undergo high rates of transfer within and between bacterial species. This study identified a number of SaPI_{bov1} Stl repressor homologues in different staphylococcal species. From these, the least homologous were analysed for the ability to interact with dimeric and trimeric phage-encoded Duts, including the trimeric Dut encoded by a *S. epidermidis* phage, ϕ IPLA6. In support of the SaPIs being able to transfer to other staphylococcal species, some of the SaPI_{bov1} Stl homologues were derepressed by the different phage-encoded Duts. This indicates that these different SaPI_{bov1} Stl homologues likely came from a common ancestor and have spread through a number of staphylococcal species, diverging in sequence when encoded by these new bacterial hosts. Furthermore, the ability of the *S. epidermidis* ϕ IPLA6 encoded Dut to induce the SaPI_{bov1} Stl, and homologues, indicates that the conserved Duts from other species of phage are sufficient to induce the related Stl repressors. This supports the idea that, following SaPI or PICI interspecific transfer, the resident phages encoding conserved proteins would be able to maintain and transfer these elements within the new host.

A number of findings further support the idea that interspecific transfer of PICIs is viable in nature. One of these is the increasing awareness of silent phage transfer as a mechanism for SaPI/PICI transfer to bacterial species that the phage cannot lyse^{217,262,288}. Previously, the assumption was that a phages host range was limited to those bacteria that facilitated the phage lytic or lysogenic life cycle. However, it has become apparent that SaPIs and virulence factors can be transferred by phages to bacterial strains that the phages themselves cannot

infect. This changes the long held belief that because phages have a limited host range this limits *S. aureus* phage-mediated MGE transfer. As such, a wider range of SaPI and PICI interspecific transfer could be facilitated in nature by phages with an otherwise narrow bacterial host range.

Another factor found to impact on the ability of PICIs to transfer between different species and more distant genera is the bacterial wall teichoic acid (WTA)⁴¹². A study into the WTA of different *S. aureus* strains found that atypical WTA structures in certain naturally occurring *S. aureus* strains facilitated the HGT of SaPIs to bacteria in phylogenetically distant genera⁴¹². The results indicated that the WTA found at the surface of Gram-positive bacteria vary in structure, thus forming a ‘glycocode’. This ‘glycocode’ allowed or negated binding by different phage receptors and this facilitated the transfer of MGEs, such as the SaPIs, to distantly related species or genera encoding the appropriate WTA structure⁴¹². SaPI virions encode the same receptor binding proteins as the helper phages that induce them. This indicates that PICIs would be able to achieve interspecific and even intergenic transfer so long as the bacterial WTA of the new host facilitates binding. Similarly, although only five SaPI *attC* sites have been identified in the *S. aureus* chromosome, the sequence specificity of the SaPI integrases is less stringent than for the phages²⁸⁸. This allows SaPIs, such as SaPI1, to integrate at alternative *attC* sites in other bacterial species²⁸⁸.

The results of this study, together with previous investigations into interspecific transfer, indicate that interspecific transfer is possible and occurs in nature. Indeed, the presence of SaPIbov1 Stl homologues in various staphylococcal species and the identification of PICI-like elements in multiple Gram-positive bacteria imply interspecific transfer is common in nature.

7.3 The SaPIbov1 Stl: a multi-domain protein targeting the conserved phage dUTPase mechanism for derepression

As the potential for interspecific transfer of SaPIs and PICIs in nature becomes more apparent, understanding how these islands are induced by different phage proteins becomes increasingly important. The initial interaction between phages

and SaPIs consists of a protein:protein interaction between specific phage proteins and the SaPI master repressor, the Stl. Understanding the role of the Stl repressors and the mechanisms by which they achieve phage protein targeting for island induction will help elucidate the phage-mediated activation and spread of PICIs in nature.

The Stl master repressor of the SaPIs plays a multifunctional role, binding the integrated SaPI DNA to repress the island and targeting phage proteins for island induction. Initially, the Stl repressors of the different islands were thought to target specific phage proteins for island induction, with different Stl proteins pairing with different phage proteins. However, investigations into the Stl repressor of the well-characterised SaPI_{bov1} island indicated that the phage protein-targeting of the Stl repressors is more sophisticated than this.

As previously mentioned, Chapter 3 of this study confirms the phage-encoded dimeric Duts as a subset of SaPI_{bov1} inducers. That SaPI_{bov1} is able to target both the dimeric and trimeric phage-encoded Duts when these proteins are so different in both sequence and structure implied the SaPI_{bov1} Stl had evolved to target a conserved phage mechanism over a specific protein.

Following this, Chapter 4 confirmed the structural differences between the phage-encoded dimeric and trimeric Duts, but identified a number of surprising similarities. Both the dimeric and trimeric phage-encoded Duts seem to undergo a conformational change on binding dUTP that blocks the binding of the SaPI_{bov1} Stl. Furthermore, they both exhibit a central unconserved region whilst displaying conserved N- and C-terminals. However, this Chapter also established that there were no overlapping residues between the dimeric and trimeric Duts and the conformational changes induced by dUTP binding were different between the two Dut types. This indicated the mechanisms by which the SaPI_{bov1} Stl bound to each of these proteins were likely to be different, and hinted at the possibility of separate regions or domains in the Stl to achieve these different interactions.

This idea was supported by the data from Chapter 6, which predicts that the SaPI_{bov1} Stl is made up of three domains: one HTH domain in the N-terminal for DNA binding and island repression, and two in the C-terminal that seem to be

involved in Dut interaction. This correlates with previous studies that predicted the DNA binding function of the HTH³⁹⁹. This HTH region seems to be conserved throughout the different SaPI Stls, which otherwise exhibit a high level of sequence divergence³⁹⁹. The implication of this conserved HTH domain is that the different Stl repressors perform the island repressing DNA binding function via the same mechanism. This is supported by the presence of conserved motif regions within the *stl-str* intergenic DNA of different SaPIs, which correlate with the motif previously identified for SaPIbov1 (Appendix 10)⁴⁰⁴. The C-terminal domains of the Stls are then free to evolve to target the phage processes specific for the induction of the different SaPIs. That the SaPIs have maintained a conserved DNA binding domain makes sense given that uncontrolled SaPI excision and replication confers a high cost to the bacteria^{278,281}.

The data from Chapter 6 indicates that residues within the central domain of the SaPIbov1 Stl may be involved in the interaction of the Stl with the phage-encoded trimeric Duts. This seemingly contradicts other experiments that associated the N-terminal with trimeric Dut interaction³¹⁰, although it is entirely possible both these regions contribute to Stl binding of the trimeric Duts. The proposed mechanism of interaction between the SaPIbov1 Stl and the trimeric Duts is that three Stl monomers can bind to each of the trimer protomers at motifs IV and VI³⁰⁴. At this point the flexible motif V of the neighbouring protomer comes round to cover the Stl binding site and stabilises the interaction^{304,311}. Further research into which SaPIbov1 Stl and trimeric motif IV, V and VI residues are involved would help fully elucidate this mechanism of interaction.

The predicted dimeric Dut:SaPIbov1 Stl interaction is quite different to the Stl interaction with the phage-encoded trimeric Duts. Data from Chapter 4 proposed that the dimerization interface of the dimeric Duts plays a vital role in the interaction with SaPIbov1 Stl. Preliminary structural data from the Marina laboratory group supports this, proposing a mechanism of interaction in which a Stl monomer binds the dimerization interface of a dimeric Dut monomer by mimicking the Dut neighbouring protomer. This involves the C-terminal most domain of the SaPIbov1 Stl and correlates with data that also implicated the Stl C-terminal most region in dimeric Dut interaction³¹⁰. However, data from this study implicated residues positioned towards the middle of the Stl protein in this

interaction. How the complete SaPIbov1 Stl binds to the phage-encoded dimeric Duts is as yet unknown and further research into the Dut and Stl residues involved in this interaction would help complete this mechanism.

That the SaPIbov1 Stl is able to target the structurally unrelated dimeric and trimeric Dut proteins, seemingly through quite different mechanisms, indicates the SaPIs have evolved to exploit conserved phage processes. By targeting unrelated proteins that perform the same process for the phage, this represents an interesting evolutionary adaptation. Through targeting conserved processes it seems the SaPIs have increased their opportunity for phage induction and optimised their chance of interspecific transfer; most phages encode either a dimeric or trimeric Dut so the SaPIs have a greater likelihood of being induced or maintained in a new host species. As such, although SaPI targeting of a process rather than of multiple proteins involved in different processes would have been a more difficult strategy to achieve, it is a more effective approach. Indeed, studies have shown that all *S. aureus* phages encode one of these two Dut types²⁸¹. Equally, this strategy makes it difficult for the phages to escape induction of the island as phages with a *dut* gene are preferentially maintained in nature²⁸¹. This goes some way to explain the divergent regions within both the dimeric and trimeric phage-encoded Duts; the phage is attempting to evade island induction whilst maintaining any functional advantage the Dut confers.

It seems that Stl repressor targeting of conserved phage proteins is not limited to SaPIbov1. This method of targeting a conserved phage processes has been extended to include SaPI2, which is induced by a number of different phage-encoded recombinase families³¹⁰. This indicates PICI-phage interactions have evolved in a more complex way than previously anticipated, and could explain the wide-ranging existence of PICIs in many Gram-positive bacteria.

Overall, this study indicates that both the dimeric and trimeric phage-encoded Duts are moonlighting phage proteins capable of inducing SaPIbov1. It also establishes the PICIs as sophisticated and effective parasites of the helper phages, with novel methods of targeting a whole host of phage proteins that perform the same phage function. It demonstrates how the SaPIbov1 Stl and other SaPI repressors have achieved an optimal balance of repression; only induced in the presence of helper phages, but rarely evaded. The study also

confirms the presence of homologous repressors in different staphylococcal species, indicative of interspecific transfer occurring in nature. All of this goes some way to explaining the widespread occurrence of PICIs throughout Gram-positive bacterial species.

7.4 Future work

The data generated in this study establishes the phage-encoded dimeric Duts as a subset of SaPIbov1 inducers. It goes some way to characterising the interaction of the phage-encoded dimeric Duts with the SaPIbov1 Stl repressor, identifying the dimerization interface as a region of interest. It establishes the existence of a number of homologues of the SaPIbov1 Stl in different staphylococcal species, two of which were capable of interacting with both dimeric and trimeric Duts from staphylococcal phages. Finally, it predicts the division of the SaPIbov1 Stl into three domains, the DNA binding HTH N-terminal domain, the central C-terminal domain, and the far C-terminal domain. Data from this study supports that the central domain may be involved in the binding of the phage-encoded trimeric Duts, whilst the far C-terminal domain may bind the dimeric Duts. However, the exact mechanism by which the SaPIbov1 Stl achieves an interaction with phage-encoded dimeric or trimeric Duts is as yet unknown. Furthermore, there was some *in silico* suggestion of potential PICI inducers in Chapter 5, but this was not explored *in vitro*. There are a number of questions that the study leaves open, additional work to help address this should include:

- Chapter 3 identified the SaPIbov1- and SaPIbov5-inducing capabilities of 6 of the 9 identified families of phage-encoded dimeric Duts. However, it did not address the induction ability of representatives from the other 3 families. Given the phylogenetic tree in this chapter indicated the potential grouping of non-inducers vs inducers, identifying the SaPIbov1/5 inducing ability of representatives HKU10-03, ΦStauST398-3, and ΦMR25 would confirm or refute this. This could be achieved by cloning these Duts into the overexpression vector pCN51 and introducing the plasmid into a SaPIbov1/5-positive strain. Southern and western blots could then be performed. Equally, the described β-lactamase experiments could be performed and a two-hybrid assay could be used to assess interaction between the SaPIbov1 Stl and the Duts. If these

experiments confirmed the grouping, establishing any conserved sequence or structural differences between the inducers and non-inducers could help direct future *in vitro* investigations into induction capability. In turn, understanding why one dimeric Dut can induce SaPIbov1/5 whilst another cannot would help elucidate the residues of importance in the dimeric Duts with regard to SaPIbov1 StI binding.

- Chapter 4 addressed the role of the dimerization interface of the phage-encoded dimeric Duts in SaPIbov1 induction. Changing this alone was not sufficient to confer induction to a non-inducing dimeric Dut, although a single point mutation in this region could abolish induction of the ‘inducing’ Duts. This Chapter did not address the potential role of the ‘latch’ in facilitating induction. This region is predicted to stabilise the Dut dimer. It would be beneficial to look at the differences between the ‘latch’ of a non-inducing dimeric Dut compared with that of an inducing dimeric Dut. Any obvious differences could be cloned into the existing dimerization interface exchange clones and the effect of this on SaPIbov1 induction assessed. To fully understand the method of interaction between the dimeric Duts and the SaPIbov1 StI, the most direct method would be to solve the structure of the two proteins interacting through crystallography. Any insight gained from this could be tested in the phage through more guided point mutations and chimeras.
- A number of speculations were made in Chapter 5 about the presence of potential PICI-inducers in prophage within other bacterial species (the Dut proteins of ϕ StB12 and ϕ Shaem1, *S. hominis* and *S. haemolyticus* respectively). These correlated with the identification of SaPIbov1 StI homologues in the same strain (ShoCI794_SEPI and ShaCI51-48). It would be a nice experiment to test the combinations of StI homologues with the identified potential Dut inducers from these prophages. This could be done using the 2 plasmid system described earlier analysing the β -lactamase expression levels as an indicator of downstream translation. It could also be done using a 2-hybrid plasmid overexpression system.

- Also in Chapter 5, the phylogenetic tree analysing the different SaPIbov1 Stl homologues indicated possible groupings between those homologues that could be induced by Duts and those that could not. It would be interesting to test the SsiCIFDAARGOS_124 (*S. simulans*) homologue to see if those groupings continued to be a correct prediction of Dut induction.
- Chapter 5 also identified the potential of the SaPIC0673 homologue (*S. sciuri*) to target the MazG family of proteins. It would be helpful to confirm this. This could be achieved using the β -lactamase assay or by a 2-hybrid assay. If the SsiCIFDAARGOS_124 (*S. simulans*) homologue was not induced by the Duts, identifying whether the MazG proteins could induce this homologue would be useful too.
- Chapter 6 identified a number of residues of interest within the SaPIbov1 Stl: DNA binding residues R22, Q29, E47, N48, and N45, dimerization residues Y70, R74, and H73, and trimeric Dut interaction residues Y112 and Y113. It also identified some residues that had unexpected effects: G15, S39, N56, D77, D155, and D117. All these should be taken forward for further analysis. To do this the mutations could be recombined into SaPIbov1 Stl, where the effect on SaPI induction, replication and transfer could be analysed using Southern blots and transduction/titre assays. It would also be useful to assess the structural properties of these SaPIbov1 Stl mutants, which could be achieved through expression tests and purification. The affinity of these purified mutants for the different Duts could be assessed by analysing their ability to inhibit the dUTPase activity of the Duts using a malachite green assay. Furthermore, the interaction of these mutants with the different Duts could be investigated using a Native gel mobility shift assay.
- Preliminary data from the laboratory indicates the structural interactions between the SaPIbov1 Stl and the phage-encoded dimeric and trimeric Duts may soon be solved. This would allow guided point mutation of the dimeric and trimeric Duts as well as the SaPIbov1 Stl. By recombining effective mutations into SaPIbov1 or inducing phages, the

two mechanisms the SaPI_{bov1} StI uses to interact with these different proteins could be fully established.

Appendices

Appendix 1 List of the oligonucleotides and Southern blot probes used in this study.

Chapter 3. The dimeric Duts are a subset of phage-encoded SaPI inducers		
Plasmid	Oligonucleotides	Sequence
pJP674	SaPI _{bov1} -149cB	CGCGGATCCGATCAGTACCTAAATATGCG
	NY-24mK	CGGGGTACCCACTCGGTTATAACCTT
pJP1927	dutNM1-1mS	ACGCGTTCGACATTATGACGGGTCAAGTTGTCTATAA
pJP1928		ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC
pJP2040		ACGATGGCGATTATAAAGATC
pJP2041	dutNM1-2m	CACGATGGCGATTATAAAGATCACGATATCGATTAT
		AAAGATGATGATGATAAAATGACTAACACATTAACA
		ATTGATCAG
	dut-DI-2cB	CGCGGATCCTTACACGTATCCTTTTCCTGC
		(These plasmids were constructed using the same primers with different templates.)
pJP2042	dutNM1-1mS	ACGCGTTCGACATTATGACGGGTCAAGTTGTCTATAA
		ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC
		ACGATGGCGATTATAAAGATC
	dutNM1-2m	CACGATGGCGATTATAAAGATCACGATATCGATTAT
		AAAGATGATGATGATAAAATGACTAACACATTAACA
		ATTGATCAG
	dutO46-4cB	CGCGGATCCTTACACGTATCCTTTTCCTG
pJP2043	dutNM1-1mS	ACGCGTTCGACATTATGACGGGTCAAGTTGTCTATAA
		ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC
		ACGATGGCGATTATAAAGATC

	phi55-dut-22m	CACGATGGCGATTATAAAGATCACGATATCGATTAT AAAGATGATGATGATAAAATGACTAACACATTAACA ATTGATC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2055 pJP1932	(Stl _{SaPIbov1}) SaPIbov1-203mB	<u>CGCGGATCCC</u> ATGGAAGGAGCTGGTCAAATGGC
	SaPIbov1-223cE	<u>CCGGAATT</u> CGATTAATTAGTGTCTTTTCAAG
	(<i>E. coli</i> RBS-Dut _{ΦNM1/ΦO11}) dutNM1-35mS	<u>ACGCGTCG</u> ACTTCACACAGGAAACAGACCATGACT AACACATTAACAATTG
	phiO11_Dut_1cP	<u>CTGCAG</u> TTACACGTATCCTTTTCCTGCG (These plasmids were constructed using the same primers with different templates.)
pJP2056	Using pJP2055 as template.	pJP2056 was created by using Klenow to transfer the His-tag in pJP2055 to a different reading frame from the Stl _{SaPIbov1} .
pJP1933	Using pJP1932 as template.	pJP1933 was created by using Klenow to transfer the His-tag in pJP1932 to a different reading frame from the Stl _{SaPIbov1} .

Sequences recognized by the restriction enzymes used in cloning are underlined.

Chapter 4. Characterisation of the dimeric Duts interaction with SaPIbov1

Plasmid	Oligonucleotides	Sequence
pJP674	SaPIbov1-149cB	<u>CGCGGATCC</u> GATCAGTACCTAAATATGCG
	NY-24mK	<u>CGGGGTACCC</u> ACTCGGTTATAACCTT
pJP2040	dutNM1-1mS	<u>ACGCGTCG</u> ACATTATGACGGGTCAAGTTGTCTATAA ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC ACGATGGCGATTATAAAGATC

	dutNM1-2m	CACGATGGCGATTATAAAGATCACGATATCGATTAT AAAGATGATGATGATAAAATGACTAACACATTAACA ATTGATCAG
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2043	dutNM1-1mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTCTATAA ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC ACGATGGCGATTATAAAGATC
	phi55-dut-22m	CACGATGGCGATTATAAAGATCACGATATCGATTAT AAAGATGATGATGATAAAATGACTAACACATTAACA ATTGATC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2044	dutNM1-1mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTCTATAA ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC ACGATGGCGATTATAAAGATC
	DI-Dut-4c	GATTTGCAATACTCAGTCCAAAT <u>AATAAGATAT</u> CCG CCAATTCATCTAACTG
	DI-Dut-3m	CAGTTAGATGAATTGGCGGATATCTTATTATTGGA CTGAGTATTGCAAATC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2046	DI-dut-29mB	<u>CGCGGATCC</u> ATGACTAACACATTAACAACCTGATCAG
	dutNM1-19cS	<u>ACGCGTCGAC</u> TTACACGTATCCTTTTCCTGC
pJP2048	dutNM1-18mB	<u>CGCGGATCC</u> ATGACTAACACATTAACAATTGATC
pJP2049	dutNM1-19cS	<u>ACGCGTCGAC</u> TTACACGTATCCTTTTCCTGC (These plasmids were constructed using the same

		primers with different templates.)
pJP2057	dutNM1-100mS DI-dut-13m DI-dut-14c dut-DI-2cB	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC GGAAGAACTACTGAGGTTATGGTTGATTTGATTGA AAATGAATACTTTTTAGATTTCCCTTAC ACCTCAGTAGTTTCTTCCAAATCTTCTTCATCAACAA TTTGATTTGCAATACTCAGTC <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2058	dutNM1-100mS DI-dut-10c DI-dut-15m dut-DI-2cB	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC AGTGTATGCATTACATGAACAAAATAAACTGAATGT AGTTTAGGTAAAGTAACTTCTTCGTCAAACGATTCA AAAA GATTGAAAATGAAGTTACTTTACCTAACTACATTCA GTTTATTTTGTTTCATGTAATGCATACACTAACAGAAG ATGATTTGACATCGATTAG <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2059	dutNM1-100mS DI-dut-17m DI-dut-18c DI-dut-19c dut-DI-2cB	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC GACTGAGTATTGCAAATCAAATTGTTGATGAAGAAG ATTTG TTGATTTGCAATACTCAGTC CGTGGTTCCTTTTCATTTTCTTTTGTATGCGTCAAT GAGTTGGTCGATTGTATAGTAAGTATTGGCGTAC <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2060	dutNM1-100mS DI-dut-7m	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC GAGGTTATGGTTGATTTGATTGAAAATGAAGTTACT TTACCTAACTACATTCAGACATGATTTATGACATG ATCAAATCAACCATAACCTCAGTAGTTTCTTCCAAAT

	DI-dut-8c	CTTCTTCATCAACAATTTGATTTGCAATACTCAGTC
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2061	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DI-dut-10c	AGTGTATGCATTACATGAACAAAATAAACTGAATGT AGTTTAGGTAAAGTAACTTCTTCGTCAAACGATTCA AAAA
	DI-dut-9m	GTTTCATGTAATGCATACACTAACAGAACAAATTTGTA AAAGGTATTGATAATAGCATTATTAGAAGATTAGTAA TAG
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2062	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DI-dut-11m	TTGTACAAGTTTTAATAATGCCGTTTTTTGTACGCCAA TACTTACTATACAATCGACCAACTCATTGACGC
	DI-dut-12c	CATTATTA AAACTTGTACAATGCTATTATCAATACCT TTTACAAATTGTTTCGTAAAACTCACTCATCATGTC
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2063	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DI-dut-7m	GAGGTTATGGTTGATTTGATTGAAAATGAAGTTACT TTACCTAAACTACATTCAGACATGATTTATGACATG
	phi55-dut-8c	AAACAAATCACGGTCATACTCTTCAAATCCTTGAGT CAAAGTTAATTGCAAAC
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2064	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-9m	AGAGTACTTTTTAGATTTCCCTTACTTAAGAAATCAA

		GACATGATTTATGACATGATGAGTGAGTTTTACGAT CAATTTGTAAAAGGTATTGA
	phi55-dut-10c	GGAAATCTAAAAAGTACTCTTCGTCAAACGATTCAA AAAACAAATCACGGTCATACTCTTCAAATCCTTGAG TCAAAGTTAATTGCAAAC
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2065	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-11m	CAGTTTGCAATTAACCTTTGACTCAAGGATTTGAAGA GTATGAC
	phi55-dut-12c	AGTCAAAGTTAATTGCAAACCTG
	phi55-dut-13c	GATTCCTTTTCATTTTCTTTTGTATGCGTCAATGAG TTGGTCGATAGAATATAATTGCTCTGCTATTTTAAAG
	phi55-dut-14cB	CGCGGATCCTTACACGTATCCTTTTCCTGCGTCTGC TGTTCCATCTTGTCTTTCATGATTCCTTTTCATTTTC TTTTTG
pJP2066	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-1m	TGAATCGTTTGACGAAGAGTACTTTTTAGATTTCCC TTACTTAAGAAATCAAGTTTATTTTGTTTCATGTAAT
	phi55-dut-2c	ACTCTTCGTCAAACGATTCAAAAAACAAATCACGGT CATACTCTTCAAATCCTTGAGTCAAAGTTAATTGCAA AC
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2067	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-3m	AAGACATGATTTATGACATGATGAGTGAGTTTTACG ATGATGATTTGACATCGGTACAAGTTTTAATAATGC

		CGT
	phi55-dut-4c	TCATGTCATAAATCATGTCTTGATTTCTTAAGTAAGG GAAATCTAAAAAGTACTCATTTTCAATCAAATCAACC A
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2068	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-5m	AAGATTAGTAATAGTCTTTAAAATAGCAGAGCAATTA TATTCTATCGACCAACTCATTGA
	phi55-dut-6c	TTAAAGACTATTACTAATCTTCTAATCGATGTCAAAT CATCATCTGTTAGTGTATGCATTACATG
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2069	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DI-dut-2c	TTCTATCGTCGAACTCCTTCATTATCTGCAGTAATTC TTGTAACTGATCA
	DI-dut-1m	TGATCAGTTACAAGAATTACTGCAGATAATGAAGGA GTTCGACGATAGAA
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2070	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DI-dut-5m	CAACTCATTGACGCATACTGAAAGAAAATGAAAAGG AAC
	DI-dut-6c	GTTCCTTTTCATTTTCTTTCAGTATGCGTCAATGAGT TG
	dut-DI-2cB	<u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2071	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DI-dut-7m	GAGGTTATGGTTGATTTGATTGAAAATGAAGTTACT TTACCTAAACTACATTCAGACATGATTTATGACATG

	DI-dut-31m phi55-dut-12c dut-DI-2cB	CAAAAGGAGTTCGACGATAGAATACCAACTAGAAAT TTAAATGAC AGTCAAAGTTAATTGCAAACCTG <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2072	dutNM1-100mS phi55-dut-15m DI-dut-30c DI-dut-17m DI-dut-18c dut-DI-2cB	<u>ACGCGTCGACATTATGACGGGTCAAGTTGTC</u> CAAAAGGAGTTCGACGATAGAATACCGACGCTGAA CTTAGGAGATAGC TATTCTATCGTCGAACTCCTTTTG GACTGAGTATTGCAAATCAAATTGTTGATGAAGAAG ATTTG TTGATTTGCAATACTCAGTC <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u>
pJP2073 pJP2076	dutNM1-100mS DI-dut-38m DI-dut-37c dut-DI-2cB	<u>ACGCGTCGACATTATGACGGGTCAAGTTGTC</u> GTAGTCGAATTCTGTGAATGGTTTAATACATTGG CCAATGTATTAAACCATTACAGAATTCGACTAC <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u> (These plasmids were constructed using the same primers with different templates.)
pJP2074 pJP2077	dutNM1-100mS DI-dut-40m DI-dut-39c dut-DI-2cB	<u>ACGCGTCGACATTATGACGGGTCAAGTTGTC</u> GTAGTCGAATTCGATGAATGGTTTAATACATTGG CCAATGTATTAAACCATTACATCGAATTCGACTAC <u>CGCGGATCCTTACACGTATCCTTTTCCTGC</u> (These plasmids were constructed using the same primers with different templates.)

pJP2075	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
pJP2078	DI-dut-42m	GTAGTCGAATTCGCTGAATGGTTTAATACATTGG
	DI-dut-41c	CCAATGTATTAAACCATTTCAGCGAATTCGACTAC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
		(These plasmids were constructed using the same primers with different templates.)
pJP2079	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2080	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	DII-dut-2m	GTAGAAGAATTTTGCGAATGGTATAACAC
	DII-dut-1c	GTGTTATACCATTTCGCAAATTCTTCTAC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2081	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-29m	GTATGATTATTGAATTTGTGAGTGGATTAACACACT TG
	phi55-dut-28c	CAAGTGTGTTAATCCACTCACAAAATTCAATAATCAT AC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2082	dutNM1-100mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTC
	phi55-dut-31m	GTATGATTATTGAATTTGCAGAGTGGATTAACACAC TTG
	phi55-dut-30c	CAAGTGTGTTAATCCACTCTGCAAATTCAATAATCAT AC
	dut-DI-2cB	<u>CGCGGATCC</u> TTACACGTATCCTTTTCCTGC
pJP2083	Achieved using gene synthesis as described in Section 2.2.3	

Appendices

pJP2084	Achieved using gene synthesis as described in Section 2.2.3	
Sequences recognized by the restriction enzymes used in cloning are underlined.		
Chapter 5. Establishing the inter-specific impact from SaPI _{bov1} targeting dimeric and trimeric Duts		
Plasmid	Oligonucleotides	Sequence
pJP674	SaPI _{bov1} -149cB	CGC <u>GGATCC</u> GATCAGTACCTAAATATGCG
	NY-24mK	CGG <u>GGTACCC</u> ACTCGGTTATAACCTT
pJP1927	dutNM1-1mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTCTATAA
pJP1928		ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC
		ACGATGGCGATTATAAAGATC
	dutNM1-2m	CACGATGGCGATTATAAAGATCACGATATCGATTAT
		AAAGATGATGATGATAAAATGACTAACACATTAACA
		ATTGATCAG
	dut-DI-2cB	<u>CGCGGATCCTT</u> ACACGTATCCTTTTCCTGC
		(These plasmids were constructed using the same primers with different templates.)
pJP1929	dutNM1-1mS	<u>ACGCGTCGAC</u> ATTATGACGGGTCAAGTTGTCTATAA
		ATATGAGGAGGCACAGGAAAATGGATTATAAAGATC
		ACGATGGCGATTATAAAGATC
	IPLA5_0063_2m	CGATGGCGATTATAAAGATCACGATATCGATTATAA
		AGATGATGATGATAAAATGAGTAAGAAATTAGAAAT
		TAAATTG
	IPLA5_0063_3cB	<u>CGCGGATCCTT</u> AGAATCCTGTTGATCCAAACCC
pJP1930	orf32_80alpha-16mS	<u>ACGCGTCGAC</u> ATTATGGCAGGTCAAGTTGTC
	Orf25-phi11-39cB	<u>CGCGGATCCCTTT</u> ACACTCCGCTACTTCCG
pJP1924	SaPI-StI-1mB	<u>CGCGGATCCC</u> ACCTTATTTAAGATTTAAATTATTTCT
		AAACTC
	SaPI-StI-2cK	<u>CGGGGTACCCTTTT</u> GCTAATTCCTCTACTCGTTTAT

		C
pJP1925	Sepid-Stl-1mB	<u>CGCGGATCC</u> GGAGAAGTTATTTTGAATTTTGGATT GTTTC
	Sepid-Stl-2cK	<u>CGGGGTACC</u> CTTTTGCTGGGCGATTGCTCTGCAA GTTTC
pJP1926	Shaem-Stl-1mB	<u>CGCGGATCC</u> GCGAGAAAAGTCATTTACAATTTTAA ATTG
	Shaem-Stl-2cK	<u>CGGGGTACC</u> CGCTTCTGTTGTTGTTCTATCTTTTGC TGTG

Sequences recognized by the restriction enzymes used in cloning are underlined.

Chapter 6. Characterisation of the SaPIbov1 Stl repressor

Plasmid	Oligonucleotides	Sequence
pJP2085	SaPIbov1-149cB	CGC <u>GATCC</u> GATCAGTACCTAAATATGCG
	NY-24mK	CGGGT <u>ACC</u> CACTCGGTTATAACCTT
pJP2086*	SaPIbov1-orf20-34c	GTTAATTTTCATGTATTTTGCAAGAGTTTAAATAATTG
	SaPIbov1-orf20-35m	CAATTATTA ^{AA} ACTCTTGCAAATACATGAAATTAAC
pJP2087*	SaPIbov1-orf20-36c	CACTCAATTTGCTTGCAGTTAATTTTCATG
	SaPIbov1-orf20-37m	CATGAAATTA ^{AA} CTGCAAGCAAATTGAGTG
pJP2088*	SaPIbov1-orf20-26c	GGTGTTTTGACTAGCTCCTGTCCTTTCAC
	SaPIbov1-orf20-27m	GTGAAAGGACAGGAGCTAGTCAA ^{AA} ACACC
pJP2089*	SaPIbov1-orf20-28c	GAAATGGTGTTTTGAGCAAATCCTGTCCTTTCAC
	SaPIbov1-orf20-29m	GTGAAAGGACAGGATTTGCTCAA ^{AA} ACACCATTTTC
pJP2090*	SaPIbov1-orf20-40c	GTTTCTGTTGCCGTTTCGCGTGATTGAAATGG
	SaPIbov1-orf20-41m	CCATTTCAAATCACGCGAACGGCAACAGAAAC
pJP2091*	SaPIbov1-orf20-30c	GTTTCTGTTGCCGGCCTCGTGATTG
	SaPIbov1-orf20-31m	CAAATCACGAGGCCGGCAACAGAAAC
pJP2092*	SaPIbov1-orf20-38c	GTAGATTTCAATTTCACTACTCCAATGTTTCTG
	SaPIbov1-orf20-39m	CAGAAACATTGGAGTAGCTGAAATTGAAATCTAC
pJP2093*	SaPIbov1-orf20-58c	GTTTTAATAATTGTGGCGTAATGTGTTGG
	SaPIbov1-orf20-59m	CCAACACATTACGCCACAATTATTA ^{AA} AC
pJP2094*	SaPIbov1-orf-20-60c	GTTGCCGTTCTCGTGAGCTGAAATGGTGTTTTG

	SaPIbov1-orf-20-61m	CAAAACACCATTTCAGCTCACGAGAACGGCAAC
pJP2095*	SaPIbov1-orf20-32c	CTCCAATGTTTCTGGCGCCGTTCTCGTG
	SaPIbov1-orf20-33m	CACGAGAACGGCGCCAGAAACATTGGAG
pJP2096*	SaPIbov1-orf20-46c	CTGTGTAGAATAGCGCTGGGTATACC
	SaPIbov1-orf20-47m	GGTATACCCAGCGCTATTCTACACAG
pJP2097*	SaPIbov1-orf20-42c	CTTTAACTCATCTGATATCGCGTGTAGAATATAGC TGGG
	SaPIbov1-orf20-43m	CCCAGCTATATTCTACACGCGATATCAGATGAGTTT AAAG
pJP2098*	SaPIbov1-orf20-44c	CCTTTTTCTTTAACTCAGCTGATATTCTGTGTAG
	SaPIbov1-orf20-45m	CTACACAGAATATCAGCTGAGTTTAAAGAAAAAGG
pJP2099*	SaPIbov1-orf20-62c	CTCATCTGATATTCTGCATAGAATATAGCTGGG
	SaPIbov1-orf20-63m	CCCAGCTATATTCTATGCAGAATATCAGATGAG
pJP2100*	SaPIbov1-orf20-52c	CATATAAATCGTATGATGAAGCATAAATGTCGCCGT C
	SaPIbov1-orf20-53m	GACGGCGACATTTATGCTTCATCATACGATTTATAT G
pJP2101*	SaPIbov1-orf20-54c	GTTTCATCATATAAATCGGCTGATGAATAATAAATG
	SaPIbov1-orf20-55m	CATTTATTATTCATCAGCCGATTTATATGATGAAAC
pJP2102*	SaPIbov1-orf20-56c	GATTGTTTCATCATATAAAGCGTATGATGAATAATAA ATG
	SaPIbov1-orf20-57m	CATTTATTATTCATCATACGCTTTATATGATGAAACA ATC
pJP2103*	SaPIbov1-orf-48c	GATTTTTCTGTAGCAGTGGAAAGG
	SaPIbov1-orf-49m	CCTTTCCACTGCTACAGAAAAATC
pJP2104*	SaPIbov1-orf20-50c	CGTATGATGAATAAGCAATGTCGCCGTC
	SaPIbov1-orf20-51m	GACGGCGACATTGCTTATTCATCATACG
pJP2105*	SaPIbov1-orf20-16c	ATAAATGTCGCCGTCATTTGCATAGGCTTTATTAAC ATAA
	SaPIbov1-orf20-15m	TTATGTTAATAAAGCCTATGCAAATGACGGCGACAT TTAT
pJP2106*	SaPIbov1-orf20-18c	TATAAATCGTATGATGATGCTGCAATGTCGCCGTCA

	SaPIbov1-orf20-17m	TTAT ATAATGACGGCGACATTGCAGCATCATCATACGATT TATA
pJP2107*	SaPIbov1-orf20-22c	CAATTTGATTGTTTCATCATATAATGCTGCTGATGAT GCTGCAATGTCGCCGTCATTATAATAGGC
	SaPIbov1-orf20-21m	GCCTATTATAATGACGGCGACATTGCAGCATCATCA GCAGCATTATATGATGAAACAATCAAATTG

Sequences recognized by the restriction enzymes used in cloning are underlined.

*All asterisked plasmids were constructed using the same external primers as pJP2085, but internal primers were used to construct the point mutations and these are listed separately for each mutant plasmid.

Southern Blot probe oligonucleotides		
Probe	Oligonucleotides	Sequence
SaPIbov1/ SaPIbov5 <i>Int</i>	SaPIbov1-112mE	CCGGAATTCAATTGCTGAGGCAAACTTC
	SaPIbov1-113cB	CGCGGATCCTAATTCTCCACGTCTAAAGC
Sequences recognized by the restriction enzymes used in cloning are underlined.		

Plasmid sequencing oligonucleotides		
Probe	Oligonucleotides	Sequence
pCN41	pCN41-3m	CCGCCTTTGAGTGAGCT
	pCN41-2c	CATGCACTTAAACTAAAGC
pCN51	pCN51-1m	TAGGTGATGAACATATCAGGC
	pCN51-4c	CAGTATTTATTATGCATTAGAATA
pMAD	pMAD-2m	AAGCGAGAAGAATCATAATGGG
	pMAD-4c	CTAGCTAATGTTACGTTACAC
pET-28a	pET-28a-1m	ATGCGTCCGGCGTAGA
	pET-28a-2c	GCTAGTTATTGCTCAGCGG
pPROEX- HTa	pPROEX-13m	AGCGGATAACAATTCACACAGG
	pPROEX-14c	TTTTATCAGACCGCTTCTGCG
Sequences recognized by the restriction enzymes used in cloning are underlined.		

Appendix 1 List of the oligonucleotides and Southern blot probes used in this study.

Appendix 2 Templates and confidence values of phage encoded dimeric Dut models generated by I-TASSER^a for Fig 3.12

Server	Dimeric Dut	Top threading templates	Closest structural analog	Top final model		
				C-score ^b	TM-score ^c	RMSD (Å) ^d
I-Tasser	ΦNM1Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.33	0.76±0.10	4.4±2.9
	ΦDI Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.12	0.73±0.11	4.8±3.1
	ΦO11Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.30	0.75±0.10	4.4±2.9
	ΦDII Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.28	0.75±0.10	4.6±3.0
	ΦO46 Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.33	0.76±0.10	4.3±2.9
	Φ55 Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.49	0.78±0.10	4.1±2.8
	ΦMR25 Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.32	0.76±0.10	4.4±2.9
	ΦStauST3983 Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.42	0.77±0.10	4.1±2.8
	HKU1003 Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.32	0.76±0.10	4.3±2.9

Appendix 2 I-TASSER templates and values for the generated phage-encoded dimeric Dut models. ^aSee reference for details³⁷⁰. ^bC-score: confidence score estimating the quality of predicted models, typically ranging between -5 and 2, a higher value signifies a model with higher confidence. ^cTM-score; a scale for measuring the structural similarity between two structures. TM-score >0.5 indicates a model of correct topology, TM-score <0.17 means a random similarity. ^dRMSD; average distance of all residue pairs in two structures.

Appendix 3 The 59 phage encoded dimeric Duts used to construct Fig 3.7

A

NCBI protein accession number (phage name)	Conserved protein domain (Conserved Domain Database, CDD – accession number)	NCBI organism (assigned)
YP_873980 (ΦNM1 Dut)*	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage ΦNM1
AFO71480	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage SA13
YP_008239533	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage JS01
EFB46834	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> C427
EFB97382	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> A9765
EHM57234	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> 21209
YP_240523 (Φ55 Dut)*	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> phage 55
ACY11889	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> ED98
ADI96843	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> ED133
EFG58024	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> EMRSA16
CAG40530	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> MRSA252
CAI81421	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> RF122
EWO63645	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> M1273
YP_240297	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus</i> phage 96
YP_009204059	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus</i> phage B166
EZU98143	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> 11S00627
EUX77280	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> M0357
Not assigned (ΦDI Dut)*	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus</i> phage ΦDI
CDR62164	cl16941: NTP-PPase Superfamily cl05653: DUF1024 Superfamily	<i>Staphylococcus schweitzeri</i>
WP_049431351	cl16941: NTP-PPase Superfamily	<i>Staphylococcus hominis</i>
WP_072098691	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus capitis</i>
SAO04495 (ΦDII Dut)*	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i>
EZY57710	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> R0294
EZY59344	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> R0353
YP_240735	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage 88

YP_240809	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage 92
EZY70921	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> R0545
YP_240887	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage X2
YP_002332391	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage Ipla35
YP_001949832 (ΦMR25 Dut)*	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus</i> phage ΦMR25
EWR22849	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> M1120
EVP78464	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> M0763
EUR36964	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> GGMC6008
EVJ54045	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> WAMC6080
EWV51383	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> F63687
EYP21990	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> DAR1158
EVM02059	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> M0897
EGA99443 (ΦO46 Dut)*	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> O46
CXE15073	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i>
EZY69567	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> R0611
CXK87292	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i>
EGA96175 (ΦO11 Dut)	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> O11
YP_239632	cl16941: NTP-PPase Superfamily	<i>Staphylococcus</i> phage 69
SGS40384	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i>
SAO24456	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i>
ENL95507	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> M1034
EUR29836	cl16941: NTP-PPase Superfamily	<i>Staphylococcus aureus</i> SJOS6061
YP_008059848 (StauST398-3 Dut)*	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus</i> phage StauST398-3
WP_057504496	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i>
WP_031880022	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i>
AMO52038	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> Tager 104
EHO90986	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> 21262
EEV67636	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> A9719
EYN68054	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> DAR3762
EMY95987	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i> M0075
ADV05101 (HKU10-03 Dut)*	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus pseudintermedius</i> HKU10-03
ANQ81944	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus pseudintermedius</i>

SGS40384	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus aureus</i>
OAW38013	cd11527: NTP-PPase_dUTPase	<i>Staphylococcus epidermidis</i>
Trimeric dUTPases (outliers)		
NCBI protein accession number (phage name)	Conserved protein domain (Conserved Domain Database, CDD – accession number)	NCBI organism (assigned)
WP_011447036 (Φ11)	cl00493: trimeric_dUTPase Superfamily	<i>Staphylococcus aureus</i>
YP_001285346 (80α)	cl00493: trimeric_dUTPase Superfamily	<i>Staphylococcus aureus</i> phage 80alpha
AAL82303 (Φ12/Φ85)	cl00493: trimeric_dUTPase Superfamily	<i>Staphylococcus aureus</i> phage Φ12
NP_075491 (ΦSLT)	cl00493: trimeric_dUTPase Superfamily	<i>Staphylococcus</i> phage SLT

*Representatives of the 9 different families identified by the construction of a neighbour joining phylogenetic tree based on the alignment shown in Appendix 2B.

B

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

	10	20	30	40	50
YP_873980	-----	-----	-----	MT	NLT
AFO71480	-----	-----	-----	MT	NLT
YP_008239533	-----	-----	-----	MT	NLT
EFB46834	-----	-----	-----	MT	NLT
EFB97382	-----	-----	-----	MT	NLT
EHM57234	-----	-----	-----	MT	NLT
YP_001949832	-----	-----	-----	MT	NLT
EVP78464	-----	-----	-----	MT	NLT
EVM02059	-----	-----	-----	MT	NLT
EYP21990	-----	-----	-----	MT	NLT
EW22849	-----	-----	-----	MT	NLT
EUR36964	-----	-----	-----	MT	NLT
EVJ54045	-----	-----	-----	MT	NLT
EWV51383	-----	-----	-----	MT	NLT
phiDI_Dut	-----	-----	-----	MT	NLT
WP_049431351	-----	-----	-----	MT	NLT
WP_072098691	-----	-----	-----	MT	NLT
EGA99443	-----	-----	-----	MT	NLT
CXE15073	-----	-----	-----	M	NNLT
CXK87292	-----	-----	-----	M	NNLT
EZY69567	-----	-----	-----	M	NNLT
CDR62164	MNNREQIEQS	VISASAYNGN	DTEGLLKEIE	DVYKKAQAFD	EINEDIINT
YP_008059848	-----	-----	-----	M	NNLT
WP_031880022	-----	-----	-----	M	NNLT
AMO52038	-----	-----	-----	MN	NLT
EHO90986	-----	-----	-----	MT	NLT
WP_057504496	-----	-----	-----	MN	NLT
EEV67636	-----	-----	-----	MRRIRKMT	NLT
EMY95987	-----	-----	-----	MT	NLT
EYN68054	-----	-----	-----	T	NLT
ADV05101	-----	-----	-----	MT	NLT
OAW38013	-----	-----	-----	MK	NKLT
ANQ81944	-----	-----	-----	MT	NLT
EGA96175	-----	-----	-----	MT	NLT
YP_239632	-----	-----	-----	MN	NLT
SAO24456	-----	-----	-----	MN	NLT
ENL95507	-----	-----	-----	M	NNLT
SGS40384	-----	-----	-----	MT	NLT
SGS40384	-----	-----	-----	MT	NLT
YP_240523	-----	-----	-----	MT	NLT
YP_009204059	-----	-----	-----	MT	NLT
ADI96843	-----	-----	-----	MT	NLT
EZU98143	-----	-----	-----	MT	NLT
EUX77280	-----	-----	-----	MT	NLT
EFG58024	-----	-----	-----	MT	NLT
CAG40530	-----	-----	-----	MT	NLT
ACY11889	-----	-----	-----	MT	NLT
CAI81421	-----	-----	-----	MT	NLA
EW063645	-----	-----	-----	MN	NLT
YP_240297	-----	-----	-----	M	NNLT
SAO04495	-----	-----	-----	MT	NLT
YP_240735	-----	-----	-----	MT	NLT
YP_240809	-----	-----	-----	MT	NLT
EZY57710	-----	-----	-----	MT	NLT
EZY59344	-----	-----	-----	MT	NLT
EZY70921	-----	-----	-----	M	NNLT
YP_002332391	-----	-----	-----	MT	NLT
YP_240887	-----	-----	-----	MT	NLT
EUR29836	-----	-----	-----	MT	NLT
Consistency	0000000000	0000000000	0000000000	0000000096	9799000000

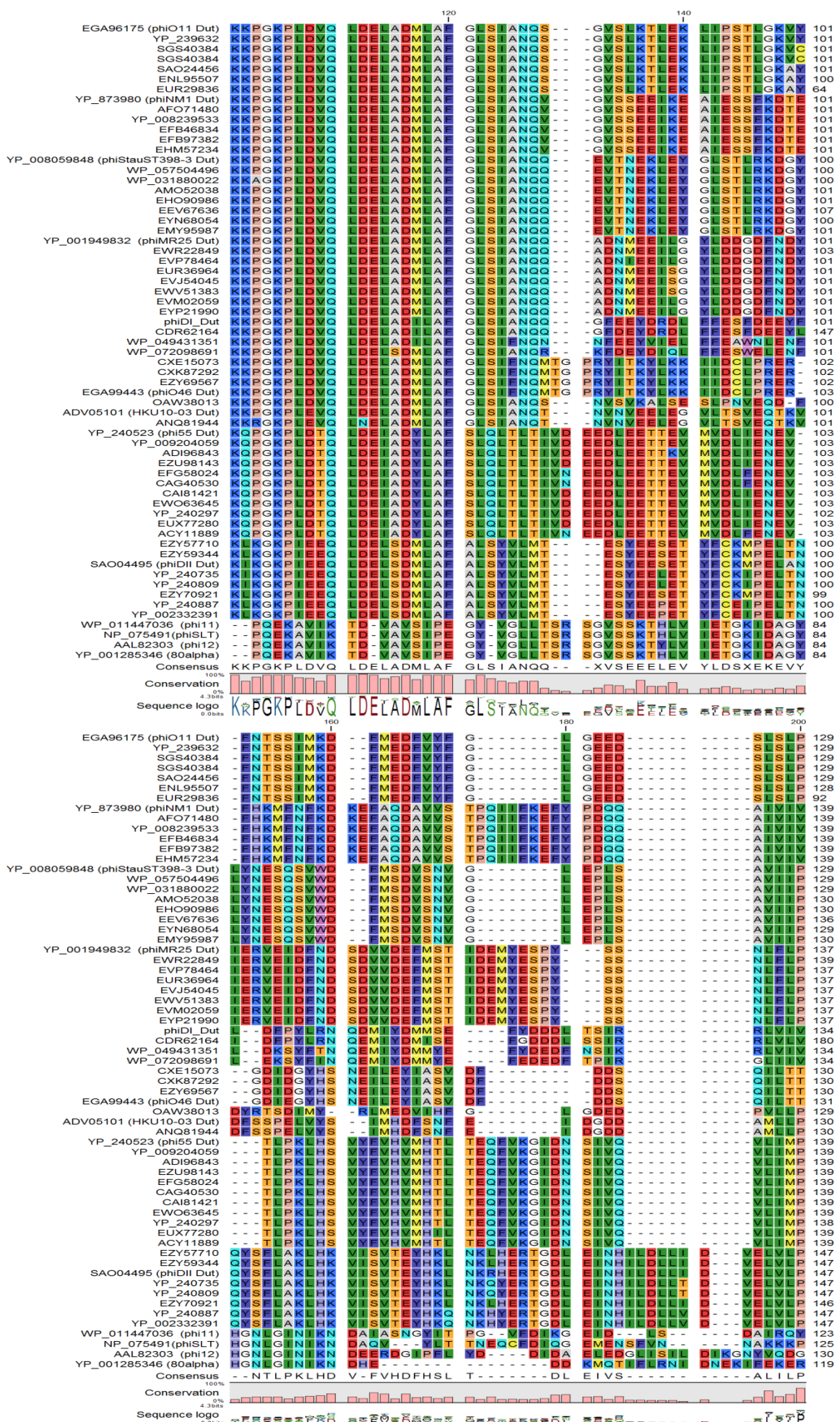
		60	70	80	90	100
YP_873980	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
AFO71480	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
YP_008239533	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EFB46834	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EFB97382	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EHM57234	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
YP_001949832	--	IDQLQELL	QIQKDFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EVP78464	--	IDQLQELL	QIQKDFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EVM02059	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EYP21990	--	VDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EWR22849	--	IDQLQELL	QIQKDFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EUR36964	--	IDQLQELL	QIQKDFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EVJ54045	--	IDQLQELL	QIQKDFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EWV51383	--	IDQLQELL	QIQKDFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
phiDI_Dut	--	TDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
WP_049431351	--	VNQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
WP_072098691	--	VDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EGA99443	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
CXE15073	--	VDQLKELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
CXK87292	--	VDQLKELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EZY69567	--	VDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
CDR62164	LT	TDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
YP_008059848	--	VDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
WP_031880022	--	VDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
AMO52038	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EH090986	--	TDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
WP_057504496	--	VDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EEV67636	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EMY95987	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EYN68054	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
ADV05101	--	VDQLQELL	KIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
OAW38013	--	AEQLKELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
ANQ81944	--	VEQLQELL	KIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
EGA96175	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
YP_239632	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
SAO24456	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
ENL95507	--	VDQLKELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
SGS40384	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
SGS40384	--	IDQLQELL	QIQKEFDDRI	PTLNLRDSKI	AYVVEFFFEWF	NTLETFKNWK
YP_240523	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFVFEWI	NTLEFFKNWK
YP_009204059	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFVFEWI	NTLEFFKNWK
ADI96843	--	TDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFVFEWI	NTLEFFKNWK
EZU98143	--	ADQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFVFEWI	NTLEFFKNWK
EUX77280	--	IDQLQELL	QIQKKFDDRI	PTRNLRDSKI	SMIIEFAEWV	NTLEFFKNWK
EFG58024	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFVFEWI	NTLEFFKNWK
CAG40530	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFVFEWI	NTLEFFKNWK
ACY11889	--	INQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFAEWV	NTLEFFKNWK
CAI81421	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFAEWV	NTLEFFKNWK
EWO63645	--	IDQLQELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFAEWV	NTLEFFKNWK
YP_240297	--	VDQLKELL	QIQKEFDDRI	PTRNLRDSKI	SMIIEFAEWV	NTLEFFKNWK
SAO04495	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
YP_240735	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
YP_240809	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
EZY57710	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
EZY59344	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
EZY70921	--	VDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
YP_002332391	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
YP_240887	--	IDQLQELL	QIQKEFDDRI	ETKNPKDTHK	AYVEEFFFEWY	NTIEPFKNWK
EUR29836	--	-----	-----	-----	-----	-----
Consistency	00	89998999	9999899999	8969759756	8797**7**7	**9*6*****

	110	120	130	140	150
YP_873980	KKPGKPLDVQ	LDELADMLAF	GLSIANQVGV	SSEEIKEAIE	SSFKDTEFHK
AF071480	KKPGKPLDVQ	LDELADMLAF	GLSIANQVGV	SSEEIKEAIE	SSFKDTEFHK
YP_008239533	KKPGKPLDVQ	LDELADMLAF	GLSIANQVGV	SSEEIKEAIE	SSFKDTEFHK
EFB46834	KKPGKPLDVQ	LDELADMLAF	GLSIANQVGV	SSEEIKEAIE	SSFKDTEFHK
EFB97382	KKPGKPLDVQ	LDELADMLAF	GLSIANQVGV	SSEEIKEAIE	SSFKDTEFHK
EHM57234	KKPGKPLDVQ	LDELADMLAF	GLSIANQVGV	SSEEIKEAIE	SSFKDTEFHK
YP_001949832	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEILGYLD	DGDFNDYIER
EVP78464	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NIEEILGYLD	DGDFNDYIER
EVM02059	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEILGYLD	DGDFNDYIER
EYP21990	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEILGYLD	DGDFNDYIER
EWR22849	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEILGYLD	DGDFNDYIER
EUR36964	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEISGYLD	DGDFNDYIER
EVJ54045	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEISGYLD	DGDFNDYIER
EWV51383	KKPGKPLDVQ	LDELADMLAF	GLSIANQQAD	NMEEISGYLD	DGDFNDYIER
phiDI_Dut	KKPGKPLDVQ	LDELADILAF	GLSIANQQGF	-EEYDRDLFF	ESFDEEYFLD
WP_049431351	KKPGKPLDVQ	LDELADILAF	GLSIFNQNNF	-EEYVIELFF	EAWNLENFLD
WP_072098691	KKPGKPLDVQ	LDELSDMMLAF	GLSIANQRKF	-DEYDIQLFF	ESWELENFLE
EGA99443	KKPGKPLDVQ	LDELADMLAF	GLSIFNQMTG	-P---RYITK	YLKKIIDCLP
CXE15073	KKPGKPLDVQ	LDELADMLAF	GLSIFNQMTG	-P---RYITK	YLKKIIDCLP
CXK87292	KKPGKPLDVQ	LDELADMLAF	GLSIFNQMTG	-P---RYITK	YLKKIIDCLP
EZY69567	KKPGKPLDVQ	LDELADMLAF	GLSIFNQMTG	-P---RYITK	YLKKIIDCLP
CDR62164	KKPGKPLDVQ	LDELADILAF	GLSIANQQGF	-DEYDRDLFF	ESFDEEYFLD
YP_008059848	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
WP_031880022	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
AMO52038	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
AHO90986	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
WP_057504496	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
EEV67636	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
EMY95987	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
EYN68054	KKPGKPLDVQ	LDELADMLAF	GLSIANQQEV	-TNEKLEYGL	STLRKDGYLE
ADV05101	KKPGKPLEVQ	LDELADMLAF	GLSIANQTNV	-NVEELEGVL	TSVEQTKVDF
OAW38013	KKPGKPLDVQ	LDELADMLAF	GLSIANQSNV	-SVKALSSESL	PN-VEQDFDY
ANQ81944	KKRGKPLEVQ	LDELADMLAF	GLSIANQTNV	-NVEELEGVL	TSVEQTKVDF
EGA96175	KKPGKPLDVQ	LDELADMLAF	GLSIANQS--	----GVSLKT	LEKLIPSTLG
YP_239632	KKPGKPLDVQ	LDELADMLAF	GLSIANQS--	----GVSLKT	LEKLIPSTLG
SAO24456	KKPGKPLDVQ	LDELADMLAF	GLSIANQS--	----GVSLKT	LEKLIPSTLG
ENL95507	KKPGKPLDVQ	LDELADMLAF	GLSIANQS--	----GVSLKT	LEKLIPSTLG
SGS40384	KKPGKPLDVQ	LDELADMLAF	GLSIANQS--	----GVSLKT	LEKLIPSTLG
SGS40384	KKPGKPLDVQ	LDELADMLAF	GLSIANQS--	----GVSLKT	LEKLIPSTLG
YP_240523	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
YP_009204059	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
ADI96843	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTKV	MVDLIENEVT
EZU98143	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
EUX77280	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
EFG58024	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVN	EEDLEETTEV	MVDLFENEVT
CAG40530	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVN	EEDLEETTEV	MVDLFENEVT
ACY11889	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVN	EEDLEETTEV	MVDLFENEVT
CAI81421	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
EWO63645	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
YP_240297	KQPGKPLDTQ	LDEIADYLAFL	SLQLTLTIVD	EEDLEETTEV	MVDLIENEVT
SAO04495	KIKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEESETYFCK	MPELANQYSF
YP_240735	KIKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEELETYFCK	IPELTNQYSF
YP_240809	KIKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEELETYFCK	IPELTNQYSF
EZY57710	KLKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEESETYFCK	MPELTNQYSF
EZY59344	KLKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEESETYFCK	MPELTNQYSF
EZY70921	KLKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEESETYFCK	MPELTNQYSF
YP_002332391	KLKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEEPETYFCE	IPELTNQYSF
YP_240887	KLKGKPIEEQ	LDELSDMMLAF	ALSYVLMTES	YEEPETYFCE	IPELTNQYSF
EUR29836	KKPGKPLDVQ	LDELADMLAF	GLSIANQSGV	SLKTLEKLIP	STLGKAYFNT
Consistency	*67***986*	*9*99*7***	7*87666433	1343344433	4434344443

	160	170	180	190	200
YP_873980	MFNFKDKEFA	QDAVVSTPQI	IFK-----	EFYPDQQAI	VIVIDIAYNL
AF071480	MFNFKDKEFA	QDAVVSTPQI	IFK-----	EFYPDQQAI	VIVIDIAYNL
YP_008239533	MFNFKDKEFA	QDAVVSTPQI	IFK-----	EFYPDQQAI	VIVIDIAYNL
EFB46834	MFNFKDKEFA	QDAVVSTPQI	IFK-----	EFYPDQQAI	VIVIDIAYNL
EFB97382	MFNFKDKEFA	QDAVVSTPQI	IFK-----	EFYPDQQAI	VIVIDIAYNL
EHM57234	MFNFKDKEFA	QDAVVSTPQI	IFK-----	EFYPDQQAI	VIVIDIAYNL
YP_001949832	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EVP78464	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EVM02059	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EYP21990	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EWR22849	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EUR36964	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EVJ54045	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
EWV51383	VEIDFNDSVD	VDEFMSTIDE	MYE-----	SPYSSNLFL	PFALANNY--
phiDI_Dut	FPYLRNQDMI	YDMMSEFYDD	DLT-----	SIRRLVIVF	KIAEQI----
WP_049431351	KSYFTNQEMI	YDMMYEFYDE	DFN-----	SIKRLVIVF	KIAEQI----
WP_072098691	KSYFINQEMI	YDMMYEFEDE	DFT-----	PIRGLIIVF	KIAEQI----
EGA99443	RERGDIEGYH	SNEILEYIAS	VDF-----	DDSQILTTI	AIARHY----
CXE15073	RERGDIDGYH	SNEILEYIAS	VDF-----	DDSQILTTI	AIARHY----
CXK87292	RERGDIDGYH	SNEILEYIAS	VDF-----	DDSQILTTI	AIARHY----
EZY69567	RERGDIDGYH	SNEILEYIAS	VDF-----	DDSQILTTI	AIARHY----
CDR62164	FPYLRNQEMI	YDMISEFGDD	DLS-----	SIRRLVLVF	KIAEQI----
YP_008059848	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
WP_031880022	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
AMO52038	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
EHO90986	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
WP_057504496	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
EEV67636	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
EMY95987	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
EYN68054	NESQSVWDFM	SDVSNV----	GLE-----	PLSAVIIPPL	DIAYNL----
ADV05101	SSPELVYSIM	HDFS NF----	EID-----	GDDAMLLPF	NFASN L----
OAW38013	RTSDIMYRLM	EDVIHF----	GLG-----	DEDPVLLPF	NIAYNL----
ANQ81944	SSPELVYSIM	HDFS NF----	EID-----	GDDAMLLPF	NFASN L----
EGA96175	KVYFNTSSIM	KDFMEDFVYF	GLG-----	EEDSLSLPL	NIAYNL----
YP_239632	KVYFNTSSIM	KDFMEDFVYF	GLG-----	EEDSLSLPL	NIAYNL----
SAO24456	KAYFNTSSIM	KDFMEDFVYF	GLG-----	EEDSLSLPL	NIAYNL----
ENL95507	KAYFNTSSIM	KDFMEDFVYF	GLG-----	EEDSLSLPL	NIAYNL----
SGS40384	KVCFNTSSIM	KDFMEDFVYF	GLG-----	EEDSLSLPL	NIAYNL----
SGS40384	KVCFNTSSIM	KDFMEDFVYF	GLG-----	EEDSLSLPL	NIAYNL----
YP_240523	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
YP_009204059	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
ADI96843	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
EZU98143	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
EUX77280	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
EFG58024	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
CAG40530	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
ACY11889	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
CAI81421	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
EWO63645	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
YP_240297	LPKLHSVYFV	HVMHTLTEQF	VKG-----	IDNSIVQVL	IMPFLYANTY
SAO04495	LAKLHKVISV	TEYHKLNKRH	ERTGDLEINH	ILDLLIDVEL	VLPFQIAVEY
YP_240735	LAKLHKVISV	TEYHKLNKQY	ERTGDLEINH	ILDLLTDVEL	VLPFQIAVEY
YP_240809	LAKLHKVISV	TEYHKLNKQY	ERTGDLEINH	ILDLLTDVEL	VLPFQIAVEY
EZY57710	LAKLHKVISV	TEYHKLNKLH	ERTGDLEINH	ILDLLIDVEL	VLPFQIAVEY
EZY59344	LAKLHKVISV	TEYHKLNKLH	ERTGDLEINH	ILDLLIDVEL	VLPFQIAVEY
EZY70921	LAKLHKVISV	TEYHKLNKLH	ERTGDLEINH	ILDLLIDVEL	VLPFQIAVEY
YP_002332391	LAKLHKVISV	TEYHKQNKHY	ERTGDLEINH	ILDLLVDVEL	VLPFQIAVEY
YP_240887	LAKLHKVISV	TEYHKQNKHY	ERTGDLEINH	ILDLLVDVEL	VLPFQIAVEY
EUR29836	SSIMKDFMED	FVYFGLG---	-----	EEDSLSLPL	NIAYNL----
Consistency	4443443446	4644443222	3340000000	0344554548	3765452111

	 210. 220. 230.
YP_873980	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
AFO71480	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_008239533	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EFB46834	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EFB97382	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EHM57234	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_001949832	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EVP78464	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EVM02059	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EYP21990	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EW22849	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EUR36964	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EVJ54045	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EWV51383	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
phiDI_Dut	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
WP_049431351	YTIDQLISAY KKKMERNHAR QDGTADKDKG YV---
WP_072098691	YTIDQLIAAY KKKMERNHAR QDGTADKDKG YV---
EGA99443	YSIDQLIDAY KKKMKRNHER QDGTADTGKG YV---
CXE15073	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
CXK87292	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EZY69567	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
CDR62164	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_008059848	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
WP_031880022	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
AMO52038	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EHO90986	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
WP_057504496	YSIDQLIDAY KKKMKRNHKKR QDGTADAGKG YV---
EEV67636	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EMY95987	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EYN68054	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
ADV05101	YTIDHLIKAY KKKMERNHAR QDGTADAGKG YV---
OAW38013	YTIDQLVIAY KKKMERNHAR QDGTADTEKG YV---
ANQ81944	YTIDQLIEAY KKKMERNHAR QDGTADAIPM MEKRR
EGA96175	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_239632	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
SAO24456	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
ENL95507	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
SGS40384	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
SGS40384	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_240523	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_009204059	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
ADI96843	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EZU98143	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EUX77280	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EFG58024	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
CAG40530	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
ACY11889	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
CAI81421	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EW063645	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_240297	YTIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
SAO04495	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_240735	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_240809	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EZY57710	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EZY59344	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EZY70921	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_002332391	YSINQLIDAY KKKMKRNHER QDGTADAGKG YV---
YP_240887	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
EUR29836	YSIDQLIDAY KKKMKRNHER QDGTADAGKG YV---
Consistency	*8*99*98** *9**9***8* *9***9899 99000

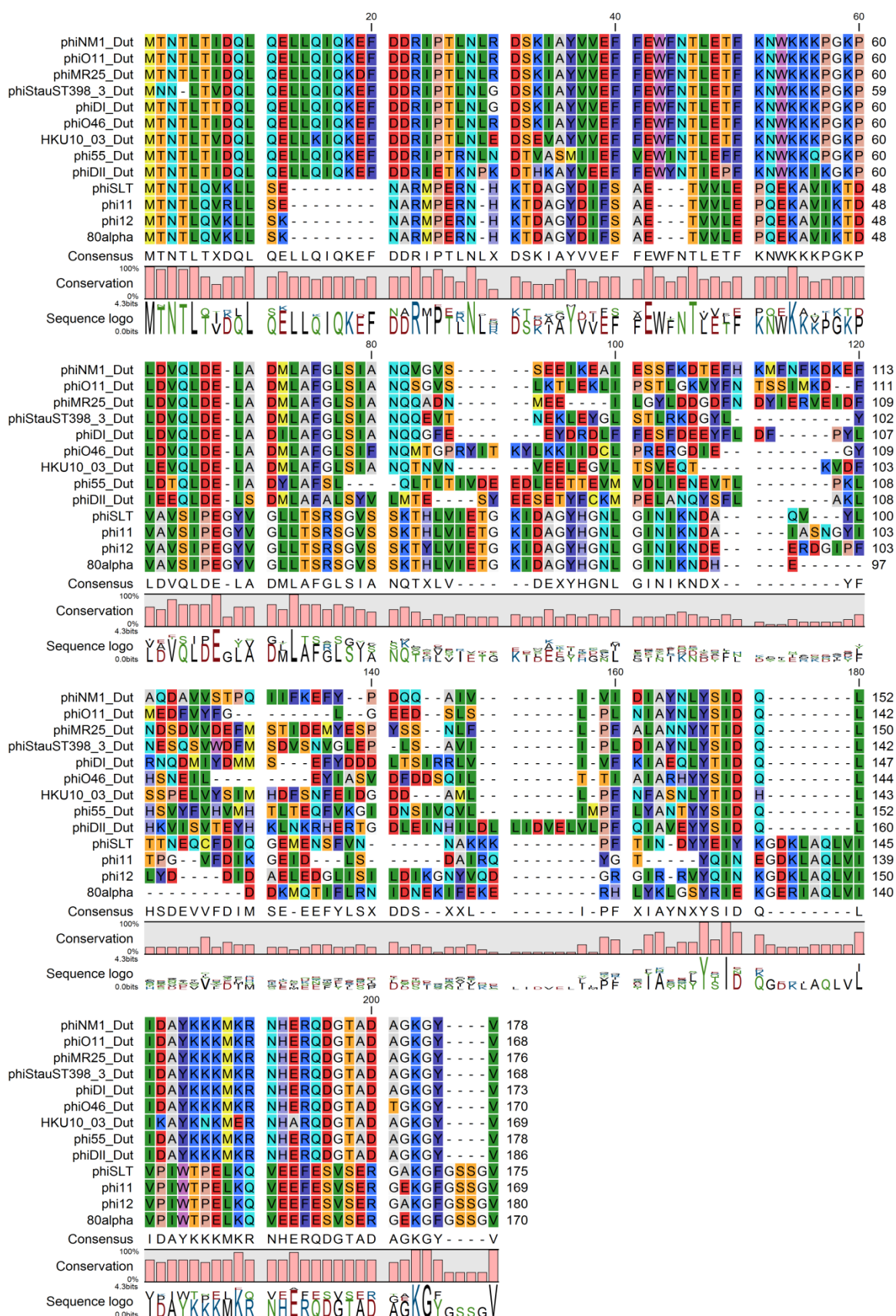






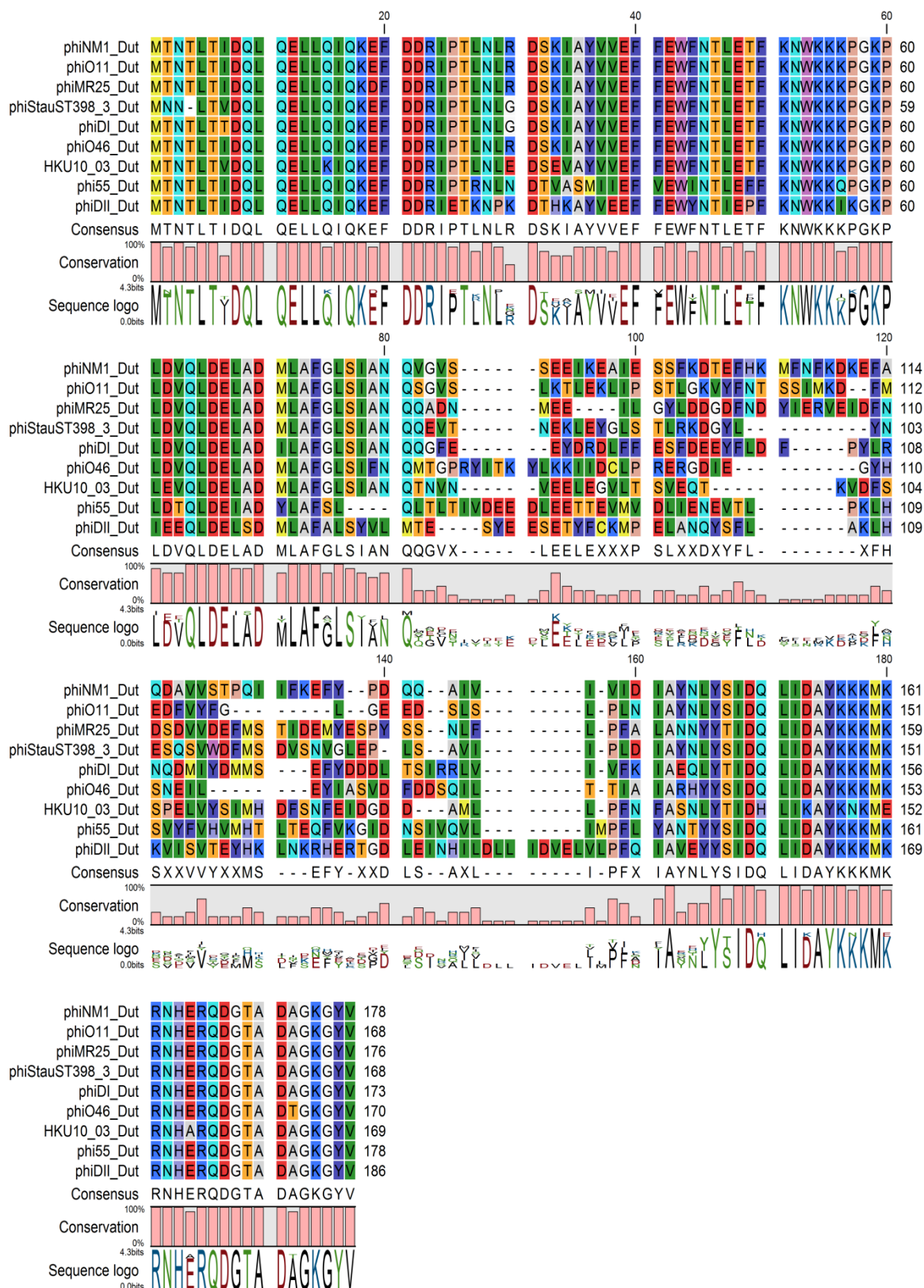
Appendix 3 The 59 phage-encoded dimeric Duts identified by a NCBI protein BLAST. (A) Staphylococcal phage dimeric Duts identified by protein BLAST. **(B)** PRALINE protein alignment of the dimeric Duts from *Staphylococcus aureus* phages. Colours indicate relative sequence conservation at each position, with red being most conserved and blue being least (adapted from alignment generated by PRALINE)^{284,285}. **(C)** Protein sequence alignment from CLC Genomics Workbench 7.5.2 used to construct Fig 3.7. A conservation bar graph of residues is shown below the amino acid sequence alignment, where low and high bars indicate low and high conservation, respectively.

Appendix 4 CLC alignment used to construct Fig 3.9



Appendix 4 Protein sequence alignment from CLC Genomics Workbench 7.5.2 used to construct Fig 3.9. Includes the 9 different phage-encoded dimeric Dut group representatives and 4 phage-encoded trimeric Dut sequences. A conservation bar graph of residues is shown below the amino acid sequence alignment, where low and high bars indicate low and high conservation, respectively.

Appendix 5 CLC alignment used to construct Fig 3.10 and 3.11



Appendix 5 Protein sequence alignment from CLC Genomics Workbench 7.5.2 used to construct Fig 3.10 and Fig 3.11. Includes the 9 different phage-encoded dimeric Dut group representatives. A conservation bar graph of residues is shown below the amino acid sequence alignment, where low and high bars indicate low and high conservation, respectively.

Appendix 7 Crystallographic statistics for Φ DI-dUPNPP and Φ O11-dUPNPP structures

Crystallographic statistics of the Φ O11-dUPNPP structure

Processed data

Beamline	XALOC (ALBA)
Wavelength (Å)	0.98
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	33.47, 88.84, 54.15
α , β , γ (°)	90, 91.68, 90
Resolution (Å)	54.13–2.1 (2.21–2.1)
R _{merge} (%)	0.13 (0.31)
Mean I/σ(I)	6.4 (3.7)
Unique reflections	18124 (2620)
Completeness (%)	98.0 (98.1)
Redundancy	3.0 (3.0)

Refined data

R _{factor} (%)	0.19 (0.20)
R _{free} (%)	0.24 (0.28)
No. non-hydrogen atoms	
Protein	2373
Ligand	56
Water	74
Mg	4
RMSD	
Bond deviation (Å)	0.0077
Angle deviation (°)	1.138
Average B-factor (Å ²)	
Protein	33.31
Ligand	23.37
Water	32.86
Mg	20.23
Ramachandran Map	
Favoured (%)	97.03
Allowed (%)	2.31
Outliers (%)	0.66
PDB accession code	-

Values in parentheses correspond to the data for the highest resolution shell.

$$R_{\text{merge}} = \sum (I - \langle I \rangle) / \sum \langle I \rangle$$

$$R_{\text{factor}} = \sum ||F_o| - |F_c|| / \sum |F_o|$$

R_{free} is the R_{factor} calculated with 5% of the total unique reflections chosen randomly and omitted from refinement.

Crystallographic statistics of the ϕ DI-dUPNPP structure

Processed data

Beamline	DLS-I.04
Wavelength (Å)	0.97942
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	69.89, 81.44, 81.53
α , β , γ (°)	90, 106.92, 90
Resolution (Å)	81.44 - 2.10 (2.21 - 2.10)
R _{merge} (%)	0.048 (0.35)
Mean I/σ(I)	8.4 (2.1)
Unique reflections	50736 (7325)
Completeness (%)	99.24 (98.9)
Redundancy	4.0 (3.7)

Refined data

R _{factor}	0.224
R _{free}	0.265
No. non-hydrogen atoms	
Protein	5388
Ligand	112
Water	210
Mg	8
RMSD	
Bond deviation (Å)	0.016
Angle deviation (°)	1.34
Ramachandran Map	
Favoured (%)	99
Allowed (%)	1
PDB accession code	5MYD

Values in parentheses correspond to the data for the highest resolution shell.

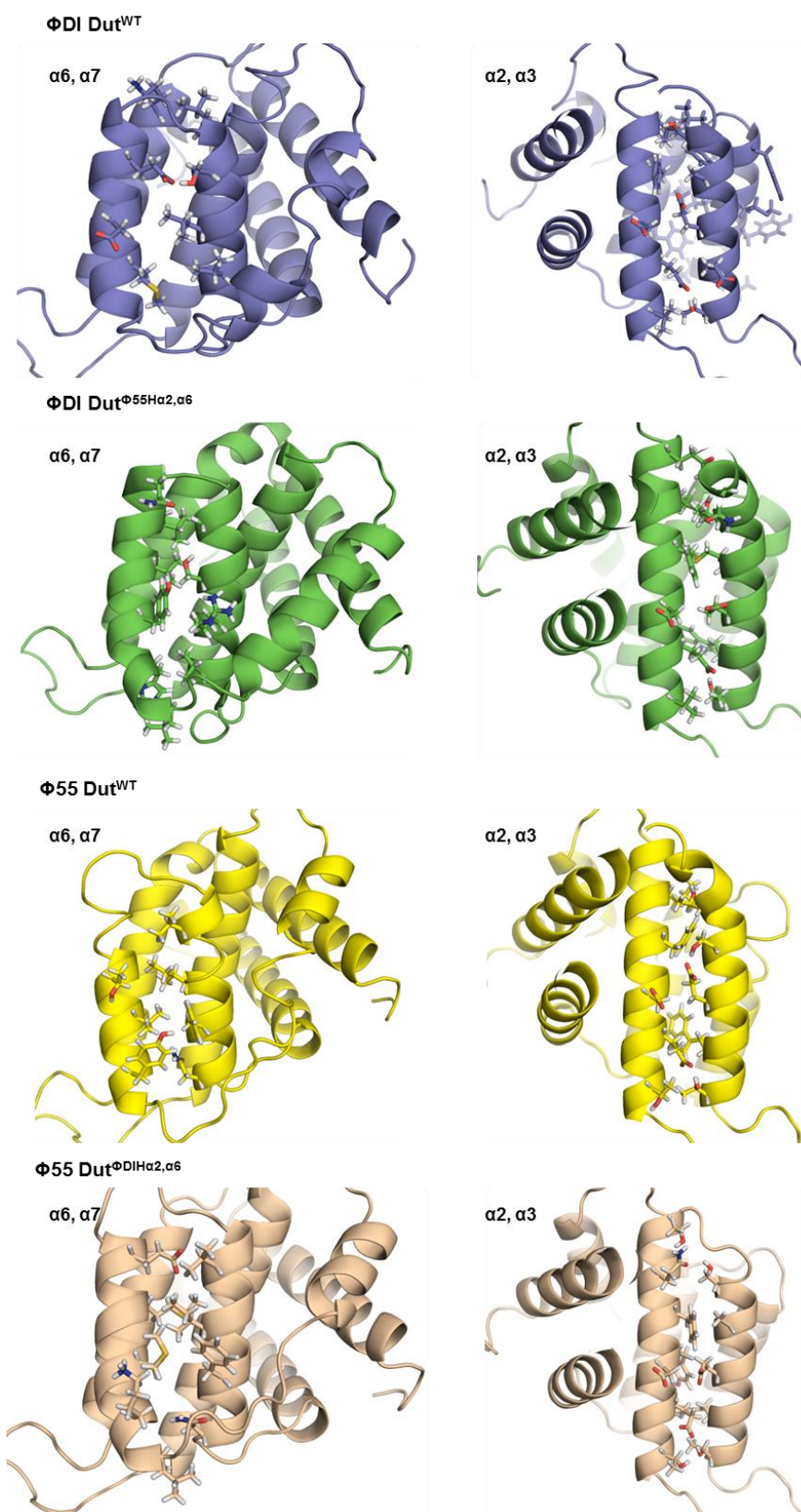
$$R_{\text{merge}} = \sum (|I - \langle I \rangle|) / \sum \langle I \rangle$$

$$R_{\text{factor}} = \sum ||F_o| - |F_c|| / \sum |F_o|$$

R_{free} is the R_{factor} calculated with 5% of the total unique reflections chosen randomly and omitted from refinement.

Appendix 7 Crystallographic statistics of Dut ϕ DI and ϕ O11 dUPNPP-bound structures.

Appendix 8 Predicted structural interference in $\Phi\text{DI}^{\Phi55\alpha2,\alpha6}$ and $\Phi55^{\Phi\text{DI}\alpha2,\alpha6}$

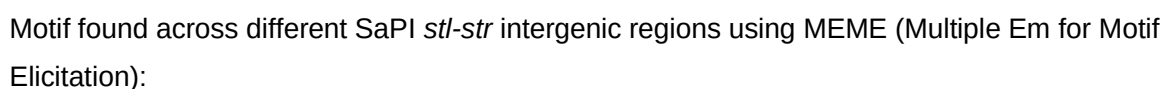


Appendix 8 Cartoon representations of the monomer 3D structures of the $\Phi\text{DI}^{\text{WT}}$, $\Phi\text{DI}^{\Phi55\alpha2,\alpha6}$, $\Phi55^{\text{WT}}$ and $\Phi55^{\Phi\text{DI}\alpha2,\alpha6}$ dimeric Duts as generated by I-Tasser^{369–371}. Predicted interference between $\alpha2$ - $\alpha3$ and $\alpha6$ - $\alpha7$ is shown. Side-chains are displayed, coloured by elements. I-TASSER templates and confidence values can be found in Appendix 3 for wild-type structures and Appendix 9 for mutants.

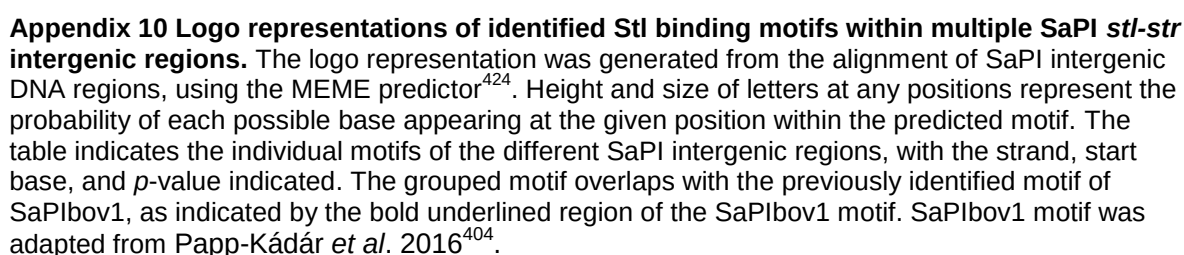
Appendix 9 Templates and confidence values for $\Phi\text{DI}^{\Phi55\alpha2,\alpha6}$ and $\Phi55^{\Phi\text{DI}\alpha2,\alpha6}$ generated by I-TASSER^a for Appendix 8

Server	Dimeric Dut	Top threading templates	Closest structural analog	Top final model		
				C-score ^b	TM-score ^c	RMSD (Å) ^d
I-Tasser	$\Phi\text{DI}^{\Phi55\alpha2,\alpha6}$ Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.14	0.73±0.11	4.8±3.1
	$\Phi55^{\Phi\text{DI}\alpha2,\alpha6}$ Dut	1w2yA, 2yayA	1w2yA, Deoxyuridine 5'-triphosphate nucleotide hydrolase (<i>Campylobacter jejuni</i>)	0.36	0.76±0.10	4.3±2.9

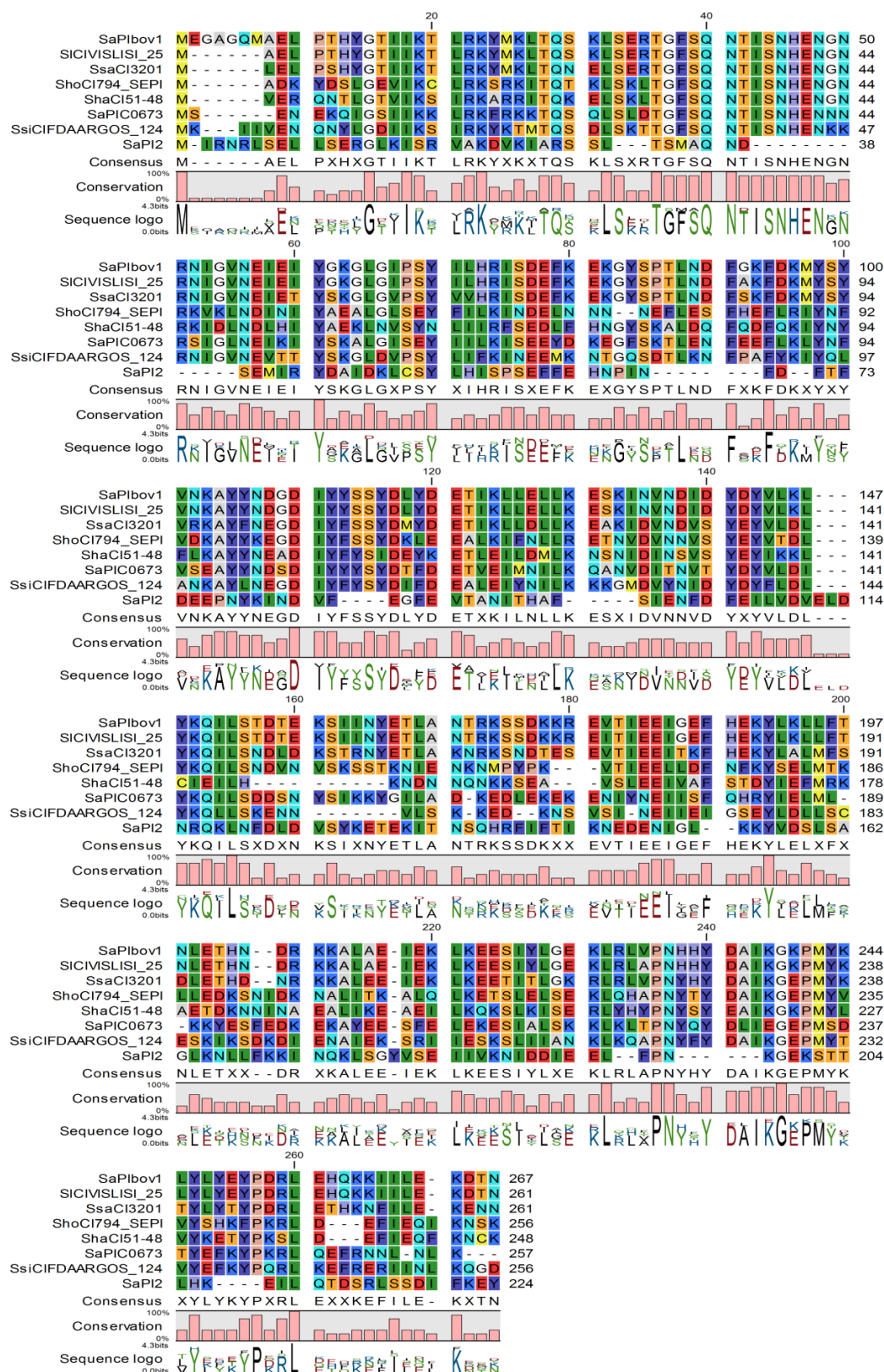
Appendix 9 I-TASSER templates and values for the generated phage-encoded dimeric Dut $\alpha2,\alpha6$ mutant models. ^aSee reference for details³⁷⁰. ^bC-score: confidence score estimating the quality of predicted models, typically ranging between -5 and 2, a higher value signifies a model with higher confidence. ^cTM-score; a scale for measuring the structural similarity between two structures. TM-score >0.5 indicates a model of correct topology, TM-score <0.17 means a random similarity. ^dRMSD; average distance of all residue pairs in two structures.



Motif found in SaPIbov1 *stl-str* intergenic region using MEME (Multiple Em for Motif Elicitation) in Papp-Kádár *et al.* 2016⁴⁰⁴:



Appendix 11 CLC alignment used to construct Fig 5.2 and 5.3



Appendix 12 Templates and confidence values for predicted SaPIbov1 StI generated by I-TASSER^a and Phyre2^b

Server	Top threading templates	Closest structural analog	Top final model		
			C-score ^c	TM-score ^d	RMSD (Å) ^e
I-Tasser	2qfcA, 5jufA, 5d50A, 4rykA, 2awiA, 5jubA, 3u3wA	5jufA, ComR Transcriptional regulator (<i>S. thermophilus</i>)	-2.13	0.46±0.15	10.9±4.6
Server	PDB code	Template info	Aligned Residues	Confidence (%) ^f	Identity (%)
Phyre2	5JUB	c5jubA, ComR transcriptional regulator (<i>S. thermophilus</i>)	10-261	99.8	10
	2GRM	c2grmB, PrgX/ICF10 transcriptional protein (<i>E. faecalis</i>)	14-261	99.8	12
	4RYK	c4rykA, Lmo0325 putative transcriptional regulator (<i>L. monocytogenes</i>)	13-260	99.8	12
	2AXZ	c2axzC, PrgX/cCF10 transcriptional protein (<i>E. faecalis</i>)	14-261	99.7	13
	4YV9	c4yv9C, Rgg2 transcriptional regulator (<i>S. dysgalactiae</i>)	14-263	99.7	12
	5FD4	c5fd4B, ComR transcriptional regulator (<i>S. suis</i>)	11-261	99.7	13
	2QFC	c2qfcB, PlcR/PapR transcriptional regulator (<i>B. thuringiensis</i>)	12-261	99.7	13

Appendix 12 I-TASSER and Phyre2 templates and values for the generated SaPIbov1 StI repressor models. ^{a,b}See references for details^{370,388}. ^cC-score: confidence score estimating the quality of predicted models, typically ranging between -5 and 2, a higher value signifies a model with higher confidence. ^dTM-score; a scale for measuring the structural similarity between two structures. TM-score >0.5 indicates a model of correct topology, TM-score <0.17 means a random similarity. ^eRMSD; average distance of all residue pairs in two structures. ^fConfidence represents the probability (from 0 to 100) that the match between the query sequence and the corresponding template is a true homology. A match with confidence >90%, generally should indicate that the query sequence adopts the overall fold shown by the template and that the core of the protein is modelled at high accuracy.

Glossary

Prophage: phage integrated into the bacterial chromosome.

Temperate phage: a phage that can undergo either a lytic or lysogenic life cycle.

Lysogen: a bacterium with a prophage.

Horizontal gene transfer: The transfer of foreign DNA among unicellular or multicellular organisms that are not parents and offspring (i.e. not vertical).

Mobile genetic element (MGE): a DNA sequence that is able to promote its transfer between organisms.

Genomic island (GI): discrete sections of DNA, which are found inserted into bacterial genomes and are thought to be acquired via horizontal gene transfer.

Staphylococcal pathogenicity island (SaPI): a GI present in *S. aureus* encoding phage-related genes that is mobilised by induction of a helper phage. The SaPIs are part of the PICI family and were the first PICIs identified.

Phage-inducible chromosomal island (PICI): a GI encoding phage-related genes equivalent to those of the SaPIs and that is mobilised after induction by a helper phage.

Satellite virus: a genetic element whose life cycle depends on a helper virus, which it parasitizes.

Terminase complex: an enzyme complex composed of two subunits, small (TerS) and large (TerL) terminase, responsible for the DNA packaging in phages.

***pac* site:** a specific DNA sequence recognised by the terminase complex of a *pac* phage to initiate packaging.

***cos* site:** a cohesive end site is a DNA sequence recognised by the terminase complex of a *cos* phage to initiate and terminate packaging.

Headful packaging: DNA packaging mechanism of *pac* phages in which concatemeric DNA is cleaved at the *pac* site and encapsidated into a preformed prohead. The complete filling of the prohead is the signal for a nonspecific cleavage of the DNA, which finishes the packaging process.

Cos unit-length packaging: DNA packaging mechanism of cos phages in which concatemeric DNA is cleaved at the cos site and encapsidated into a preformed prohead. The recognition of the cos site at a precise unit-length distance is the signal for the second cleavage, which finishes the packaging process.

Molecular piracy: a biological process in which a replicon takes over the structural proteins of another replicon to package and spread its own genome.

SOS: bacterial response to DNA damage, detaining the cell cycle and triggering DNA repair. The SOS response includes the RecA-mediated cleavage of the LexA repressor and derepression of repair response genes.

Campbell mechanism: an integrase mediated site-specific integration of a circularised genetic element into a bacterial chromosome by a specific reciprocal crossover between shared sequences.

Concatemeric DNA: a large, non-interrupted molecule of DNA, which encompasses numerous copies of the same genome linked end to end in tandem.

Procapsid: the empty precursor of the capsid particle made up of several copies of capsid protein subunits, which is normally constituted by internal scaffolding proteins.

Capsid: a polyhedral protein complex, or shell of a virus, which will harbour the virus genome; usually icosahedral in bacteriophages.

Virion: a virus infective particle.

List of references

1. Edwards, A. M., Massey, R. C. & Clarke, S. R. Molecular mechanisms of *Staphylococcus aureus* nasopharyngeal colonization. *Molecular Oral Microbiology* **27**, 1–10 (2012).
2. Herron-Olson, L., Fitzgerald, J. R., Musser, J. M. & Kapur, V. Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS One* **2**, e1120 (2007).
3. Lekkerkerk, W. S. N., Van Wamel, W. J. B., Snijders, S. V., Willems, R. J., Van Duijkeren, E., Broens, E. M., Wagenaar, J. A., Lindsay, J. A. & Vos, M. C. What is the origin of livestock-associated methicillin-resistant *Staphylococcus aureus* clonal complex 398 isolates from humans without livestock contact? An epidemiological and genetic analysis. *Journal of Clinical Microbiology*. **53**, 1836–1841 (2015).
4. Lowder, B. V., Guinane, C. M., Ben Zakour, N. L., Weinert, L. A., Conway-Morris, A., Cartwright, R. A., Simpson, A. J., Rambaut, A., Nubel, U. & Fitzgerald, J. R. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*. **106**, 19545–19550 (2009).
5. Peacock, S. J., De Silva, I. & Lowy, F. D. What determines nasal carriage of *Staphylococcus aureus*? *Trends in Microbiology* **9**, 605–610 (2001).
6. Kluytmans, J., Van Belkum, A. & Verbrugh, H. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clinical Microbiology Reviews* **10**, 505–520 (1997).
7. Gorwitz, R. J., Kruszon-Moran, D., McAllister, S. K., McQuillan, G., McDougal, L. K., Fosheim, G. E., Jensen, B. J., Killgore, G., Tenover, F. C. & Kuehnert, M. J. Changes in the prevalence of nasal colonization with *Staphylococcus aureus* in the United States, 2001–2004. *Journal of Infectious Diseases* **197**, 1226–1234 (2008).
8. Yu, V. L., Goetz, A., Wagener, M., Smith, P. B., Rihs, J. D., Hanchett, J. & Zuravleff, J. J. *Staphylococcus aureus* nasal carriage and infection in patients on hemodialysis. Efficacy of antibiotic prophylaxis. *New England Journal of Medicine*. **315**, 91–96 (1986).
9. Rocha, L. A., Marques Ribas, R., da Costa Darini, A. L. & Gontijo Filho, P. P. Relationship between nasal colonization and ventilator-associated pneumonia and the role of the environment in transmission of *Staphylococcus aureus* in intensive care units. *American Journal of Infection Control*. **41**, (2013).
10. Wertheim, H. F. L., Melles, D. C., Vos, M. C., van Leeuwen, W., van Belkum, A., Verbrugh, H. A. & Nouwen, J. L. The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infectious Diseases*. **5**, 751–62 (2005).
11. von Eiff, C., Becker, K., Machka, K., Stammer, H. & Peters, G. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. Study Group. *New England Journal of Medicine*. **344**, 11–6 (2001).
12. Safdar, N. & Bradley, E. A. The risk of infection after nasal colonization with *Staphylococcus aureus*. *American Journal of Medicine* **121**, 310–315 (2008).
13. Tattevin, P., Schwartz, B. S., Graber, C. J., Volinski, J., Bhukhen, A., Bhukhen, A., Mai, T. T., Vo, N. H., Dang, D. N., Phan, T. H., Basuino, L., Perdreau-Remington, F., Chambers, H. F. & Diep, B. A. Concurrent epidemics of skin and soft tissue infection and bloodstream infection due to community-associated methicillin-resistant *Staphylococcus aureus*. *Clinical Infectious Diseases*. **55**, 781–788 (2012).
14. Hasty, M. B., Klasner, A., Kness, S., Denmark, T. K., Ellis, D., Herman, M. I. & Brown, L. Cutaneous community-associated methicillin-resistant *Staphylococcus aureus* among all skin and soft-tissue infections in two geographically distant pediatric emergency departments. *Academic Emergency Medicine*. **14**, 35–40 (2007).
15. Holland, T. L., Arnold, C. & Fowler, V. G. Clinical management of *Staphylococcus aureus* bacteremia. *Journal of the American Medical Association*. **312**, 1330 (2014).
16. Frimodt-Møller, N., Espersen, F., Skinhøj, P., Rosdahl, V. T., Dalton, H., Allison, M., Spengler, R., Greenough, I. W., Stolley, P., Gatell, J., Trilla, A., Latorre, X., Al, E., Bruun, B., Sloth, K., Bentzon, M., Frederiksen, W., Peltola, H., Salomaa, T., *et al.* Epidemiology of *Staphylococcus aureus* bacteremia in Denmark from 1957 to 1990. *Clinical Microbiology*

- and *Infection*. **3**, 297–305 (1997).
17. Bishara, J., Goldberg, E., Leibovici, L., Samra, Z., Shaked, H., Mansur, N. & Paul, M. Healthcare-associated vs. hospital-acquired *Staphylococcus aureus* bacteremia. *International Journal of Infectious Diseases*. **16**, e457–e463 (2012).
 18. Nienaber, J. J. C., Sharma Kuinkel, B. K., Clarke-Pearson, M., Lamletthson, S., Park, L., Rude, T. H., Barriere, S., Woods, C. W., Chu, V. H., Marín, M., Bukovski, S., Garcia, P., Corey, G. R., Korman, T., Doco-Leconte, T., Murdoch, D. R., Reller, L. B. & Fowler, V. G. Methicillin-susceptible *Staphylococcus aureus* endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. *Journal of Infectious Diseases*. **204**, 704–713 (2011).
 19. Federspiel, J., Stearns, S., Peppercorn, A., Chu, V. & Fowler, V. Increasing US rates of endocarditis with *Staphylococcus aureus*: 1999–2008. *Archives of Internal Medicine*. **172**, 363–365 (2012).
 20. Lina, G., Piemont, Y., Godail-Gamot, F., Bes, M., Peter, M.-O., Gauduchon, V., Vandenesch, F. & Etienne, J. Involvement of Panton-Valentine leukocidin-producing *Staphylococcus aureus* in primary skin infections and pneumonia. *Clinical Infectious Diseases*. **29**, 1128–1132 (1999).
 21. Switalski, L. M., Patti, J. M., Butcher, W., Gristina, A. G., Speziale, P. & Höök, M. A collagen receptor on *Staphylococcus aureus* strains isolated from patients with septic arthritis mediates adhesion to cartilage. *Molecular Microbiology*. **7**, 99–107 (1993).
 22. Frazee, B. W., Fee, C. & Lambert, L. How common is MRSA in adult septic arthritis? *Annals of Emergency Medicine*. **54**, 695–700 (2009).
 23. Kalinka, J., Hachmeister, M., Geraci, J., Sordelli, D., Hansen, U., Niemann, S., Oetermann, S., Peters, G., Löffler, B. & Tuchscher, L. *Staphylococcus aureus* isolates from chronic osteomyelitis are characterized by high host cell invasion and intracellular adaptation, but still induce inflammation. *International Journal of Medical Microbiology*. **304**, 1038–1049 (2014).
 24. Inoue, S., Moriyama, T., Horinouchi, Y., Tachibana, T., Okada, F., Maruo, K. & Yoshiya, S. Comparison of clinical features and outcomes of *Staphylococcus aureus* vertebral osteomyelitis caused by methicillin-resistant and methicillin-sensitive strains. *Springerplus* **2**, 283 (2013).
 25. Salgado-Pabón, W., Breshears, L., Spaulding, A. R., Merriman, J. A., Stach, C. S., Horswill, A. R., Peterson, M. L. & Schlievert, P. M. Superantigens are critical for *Staphylococcus aureus* infective endocarditis, sepsis, and acute kidney injury. *MBio* **4**, e00494-13 (2013).
 26. Spaulding, A. R., Satterwhite, E. A., Lin, Y.-C., Chuang-Smith, O. N., Frank, K. L., Merriman, J. A., Schaefer, M. M., Yarwood, J. M., Peterson, M. L. & Schlievert, P. M. Comparison of *Staphylococcus aureus* strains for ability to cause infective endocarditis and lethal sepsis in rabbits. *Frontiers in Cellular and Infection Microbiology*. **2**, 1–9 (2012).
 27. Argudín, M. Á., Mendoza, M. C. & Rodicio, M. R. Food poisoning and *Staphylococcus aureus* enterotoxins. *Toxins* **2**, 1751–1773 (2010).
 28. Kreiswirth, B. N., Projan, S. J., Schlievert, P. M. & Novick, R. P. Toxic shock syndrome toxin 1 is encoded by a variable genetic element. *Clinical Infectious Diseases*. **11**, S83–S89 (1989).
 29. Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Molecular Microbiology*. **29**, 527–543 (1998).
 30. Gillet, Y., Issartel, B., Vanhems, P., Fournet, J. C., Lina, G., Bes, M., Vandenesch, F., Piémont, Y., Brousse, N., Floret, D. & Etienne, J. Association between *Staphylococcus aureus* strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet* **359**, 753–759 (2002).
 31. Handler, M. Z. & Schwartz, R. A. Staphylococcal scalded skin syndrome: diagnosis and management in children and adults. *Journal of the European Academy of Dermatology and Venereology* **28**, 1418–1423 (2014).
 32. Hermans, K., Devriese, L. A. & Haesebrouck, F. Rabbit staphylococcosis: difficult solutions

- for serious problems. *Veterinary Microbiology*. **91**, 57–64 (2003).
33. Bergonier, D., de Crémoux, R., Rupp, R., Lagriffoul, G. & Berthelot, X. Mastitis of dairy small ruminants. *Veterinary Research*. **34**, 689–716 (2003).
 34. Viana, D., Blanco, J., Tormo-Más, M. Á., Selva, L., Guinane, C. M., Baselga, R., Corpa, J. M., Lasa, Í., Novick, R. P., Fitzgerald, J. R. & Penadés, J. R. Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. *Molecular Microbiology*. **77**, 1583–1594 (2010).
 35. Holden, M. T. G., Hsu, L.-Y., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Zemlickova, H., Coombs, G., Kearns, A. M., Hill, R. L. R., Edgeworth, J., Gould, I., Gant, V., et al. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Research*. **23**, 653–664 (2013).
 36. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clinical Microbiology Reviews*. **28**, 603–661 (2015).
 37. Köck, R., Becker, K., Cookson, B., van Gemert-Pijnen, J. E., Harbarth, S., Kluytmans, J., Mielke, M., Peters, G., Skov, R. L., Struelens, M. J., Tacconelli, E., Navarro Torné, A., Witte, W. & Friedrich, A. W. Methicillin-resistant *Staphylococcus aureus* (MRSA): burden of disease and control challenges in Europe. *Euro surveillance: bulletin européen sur les maladies transmissibles = European communicable disease bulletin* **15**, 19688 (2010).
 38. Chambers, H. F. & Deleo, F. R. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nature Reviews Microbiology*. **7**, 629–41 (2009).
 39. Planet, P. J., Larussa, S. J., Dana, A., Smith, H., Xu, A., Ryan, C., Uhlemann, A. C., Boundy, S., Goldberg, J., Narechania, A., Kulkarni, R., Ratner, A. J., Geoghegan, J. A., Kolokotronis, S. O. & Prince, A. Emergence of the epidemic methicillin-resistant *Staphylococcus aureus* strain USA300 coincides with horizontal transfer of the arginine catabolic mobile element and speG-mediated adaptations for survival on skin. *MBio*. **4**, e00889-13 (2013).
 40. Chang, S., Sievert, D. M., Hageman, J. C., Boulton, M. L., Tenover, F. C., Downes, F. P., Shah, S., Rudrik, J. T., Pupp, G. R., Brown, W. J., Cardo, D. & Fridkin, S. K. Infection with vancomycin-resistant *Staphylococcus aureus* containing the vanA resistance gene. *New England Journal of Medicine*. **348**, 1342–1347 (2003).
 41. Howden, B. P., Davies, J. K., Johnson, P. D. R., Stinear, T. P. & Grayson, M. L. Reduced vancomycin susceptibility in *Staphylococcus aureus*, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: Resistance mechanisms, laboratory detection, and clinical implications. *Clinical Microbiology Reviews*. **23**, 99–139 (2010).
 42. Diekema, D. J., Pfaller, M. A., Schmitz, F. J., Smayevsky, J., Bell, J., Jones, R. N. & Beach, M. Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific Region for the SENTRY Antimicrobial Surveillance Program, 1997-1999. *Clinical Infectious Diseases*. **32**, S114–S132 (2001).
 43. Fluit, A. C., Verhoef, J. & Schmitz, F. J. Frequency of isolation and antimicrobial resistance of gram-negative and gram-positive bacteria from patients in intensive care units of 25 European university hospitals participating in the European arm of the SENTRY Antimicrobial Surveillance Program 19. *European Journal of Clinical Microbiology & Infectious Diseases*. **20**, 617–625 (2001).
 44. Lindsay, J. A., Knight, G. M., Budd, E. L. & McCarthy, A. J. Shuffling of mobile genetic elements (MGEs) in successful healthcare-associated MRSA (HA-MRSA). *Mobile Genetic Elements*. **2**, 239–243 (2012).
 45. Kennedy, A. D., Otto, M., Braughton, K. R., Whitney, A. R., Chen, L., Mathema, B., Mediavilla, J. R., Byrne, K. A., Parkins, L. D., Tenover, F. C., Kreiswirth, B. N., Musser, J. M. & DeLeo, F. R. Epidemic community-associated methicillin-resistant *Staphylococcus aureus*: Recent clonal expansion and diversification. *Proceedings of the National Academy of Sciences*. **105**, 1327–1332 (2008).
 46. Fridkin, S. K., Hageman, J. C., Morrison, M., Sanza, L. T., Como-Sabetti, K., Jernigan, J. A.,

- Harriman, K., Harrison, L. H., Lynfield, R. & Farley, M. M. Methicillin-resistant *Staphylococcus aureus* disease in three communities. *New England Journal of Medicine*. **352**, 1436–1444 (2005).
47. Ito, T., Hiramatsu, K., Oliveira, D. C., De Lencastre, H., Zhang, K., Westh, H., O'Brien, F., Giffard, P. M., Coleman, D., Tenover, F. C., Boyle-Vavra, S., Skov, R. L., Enright, M. C., Kreiswirth, B., Kwan, S. K., Grundmann, H., Laurent, F., Sollid, J. E., Kearns, A. M., *et al.* Classification of Staphylococcal Cassette Chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. *Antimicrobial Agents and Chemotherapy*. **53**, 4961–4967 (2009).
 48. Sakoulas, G. & Moellering, Jr., R. C. Increasing antibiotic resistance among methicillin-resistant *Staphylococcus aureus* strains. *Clinical Infectious Diseases*. **46**, S360–S367 (2008).
 49. Hisata, K., Kuwahara-Arai, K., Yamamoto, M., Ito, T., Nakatomi, Y., Cui, L., Baba, T., Terasawa, M., Sotozono, C., Kinoshita, S., Yamashiro, Y. & Hiramatsu, K. Dissemination of methicillin-resistant *staphylococci* among healthy Japanese children. *Journal of Clinical Microbiology*. **43**, 3364–3372 (2005).
 50. Vandenesch, F. F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.-E. & Etienne, J. Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. *Emerging Infectious Diseases*. **9**, (2003).
 51. Li, M., Du, X., Villaruz, A. E., Diep, B. A., Wang, D., Song, Y., Tian, Y., Hu, J., Yu, F., Lu, Y. & Otto, M. MRSA epidemic linked to a quickly spreading colonization and virulence determinant. *Nature Medicine*. **18**, 816–819 (2012).
 52. Rooijackers, S. H. M., Ruyken, M., van Roon, J., van Kessel, K. P. M., van Strijp, J. A. G. & van Wamel, W. J. B. Early expression of SCIN and CHIPS drives instant immune evasion by *Staphylococcus aureus*. *Cellular Microbiology*. **8**, 1282–1293 (2006).
 53. Lindsay, J. A. & Holden, M. T. G. Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Functional & Integrative Genomics*. **6**, 186–201 (2006).
 54. Alibayov, B., Baba-Moussa, L., Sina, H., Zdeňková, K. & Demnerová, K. *Staphylococcus aureus* mobile genetic elements. *Molecular Biology Reports*. **41**, 5005–5018 (2014).
 55. Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*. **3**, (2005).
 56. Lindsay, J. A. *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *International Journal of Medical Microbiology*. **304**, 103–109 (2014).
 57. Novick, R. P. Mobile genetic elements and bacterial toxinoses: The superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid*. **49**, 93–105 (2003).
 58. Argemi, X., Martin, V., Loux, V., Dahyot, S., Lebeurre, J., Guffroy, A., Martin, M., Velay, A., Keller, D., Riegel, P., Hansmann, Y., Paul, N. & Prévost, G. Whole-genome sequencing of seven strains of *Staphylococcus lugdunensis* allows identification of mobile genetic elements. *Genome Biology and Evolution*. **9**, 1183–1189 (2017).
 59. Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences*. **98**, 8821–8826 (2001).
 60. Mašlaňová, I., Doškař, J., Varga, M., Kuntová, L., Mužík, J., Malúšková, D., Růžicková, V. & Pantůček, R. Bacteriophages of *Staphylococcus aureus* efficiently package various bacterial genes and mobile genetic elements including SCCmec with different frequencies. *Environmental Microbiology Reports*. **5**, 66–73 (2013).
 61. Suttle, C. A. Viruses in the sea. *Nature*. **437**, 356–361 (2005).
 62. Wichels, A., Biel, S. S., Gelderblom, H. R., Brinkhoff, T., Muyzer, G. & Schütt, C. Bacteriophage diversity in the North Sea. *Applied and Environmental Microbiology*. **64**, 4128–4133 (1998).

63. Marsh, P. & Wellington, E. M. H. Phage-host interactions in soil. *FEMS Microbiology Ecology*. **15**, 99–107 (1994).
64. Mirzaei, M. K. & Maurice, C. F. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nature Reviews Microbiology*. **15**, 397–408 (2017).
65. Clokie, M. R., Millard, A. D., Letarov, A. V & Heaphy, S. Phages in nature. *Bacteriophage*. **1**, 31–45 (2011).
66. Ashelford, K. E., Day, M. J. & Fry, J. C. Elevated abundance of bacteriophage infecting bacteria in soil. *Applied and Environmental Microbiology*. **69**, 285–289 (2003).
67. Sulakvelidze, A. Bacteriophage. *Bacteriophage*. **1**, 1–2 (2011).
68. Rosenwasser, S., Ziv, C., Creveld, S. G. van & Vardi, A. Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean. *Trends in Microbiology*. **24**, 821–832 (2016).
69. McCarthy, A. J., Witney, A. A. & Lindsay, J. A. *Staphylococcus aureus* temperate bacteriophage: carriage and horizontal gene transfer is lineage associated. *Frontiers in Cellular and Infection Microbiology*. **2**, 6 (2012).
70. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature*. **399**, 541–8 (1999).
71. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*. **5**, 801–812 (2007).
72. Chibani-Chennoufi, S., Bruttin, A., Dillmann, M. L. & Brüssow, H. Phage-host interaction: An ecological perspective. *Journal of Bacteriology*. **186**, 3677–3686 (2004).
73. Boyd, E. F. Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions. *Advances in Virus Research*. **82**, 91–118 (2012).
74. Sumby, P. & Waldor, M. K. Transcription of the toxin genes present within the staphylococcal phage ϕ Sa3ms is intimately linked with the phage's life cycle. *Journal of Bacteriology*. **185**, 6841–6851 (2003).
75. Nakayama, K., Takashima, K., Ishihara, H., Shinomiya, T., Kageyama, M., Kanaya, S., Ohnishi, M., Murata, T., Mori, H. & Hayashi, T. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Molecular Microbiology*. **38**, 213–231 (2000).
76. Coulthurst, S. J. The Type VI secretion system – a widespread and versatile cell targeting system. *Research in Microbiology*. **164**, 640–654 (2013).
77. Scholl, D. & Martin, D. W. Antibacterial efficacy of R-type pyocins towards *Pseudomonas aeruginosa* in a murine peritonitis model. *Antimicrobial Agents and Chemotherapy*. **52**, 1647–1652 (2008).
78. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews*. **68**, 560–602 (2004).
79. Kwan, T., Liu, J., DuBow, M., Gros, P. & Pelletier, J. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proceedings of the National Academy of Sciences*. **102**, 5174–5179 (2005).
80. Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K. I., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K. & Hiramatsu, K. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet*. **359**, 1819–1827 (2002).
81. Waldor, M. K. & Friedman, D. I. Phage regulatory circuits and virulence gene expression. *Current Opinion in Microbiology*. **8**, 459–465 (2005).
82. Ross, J. & Topp, E. Abundance of antibiotic resistance genes in bacteriophage following soil fertilization with dairy manure or municipal biosolids, and evidence for potential transduction. *Applied and Environmental Microbiology*. **81**, 7905–7913 (2015).
83. Betley, M. & Mekalanos, J. Staphylococcal enterotoxin A is encoded by phage. *Science*.

- 229**, 185–187 (1985).
84. Coleman, D. C., Sullivan, D. J., Russell, R. J., Arbuthnott, J. P., Carey, B. F. & Pomeroy, H. M. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. *Journal of General Microbiology*. **135**, 1679–97 (1989).
 85. Narita, S., Kaneko, J., Chiba, J. ichi, Piémont, Y., Jarraud, S., Etienne, J. & Kamio, Y. Phage conversion of Pantón-Valentine leukocidin in *Staphylococcus aureus*: Molecular analysis of a PVL-converting phage, ϕ SLT. *Gene*. **268**, 195–206 (2001).
 86. Zhang, X., McDaniel, A. D., Wolf, L. E., Keusch, G. T., Waldor, M. K. & Acheson, D. W. Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *Journal of Infectious Diseases*. **181**, 664–670 (2000).
 87. Brüssow, H. & Hendrix, R. W. Phage genomics: small is beautiful. *Cell*. **108**, 13–16 (2002).
 88. Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. Prophage genomics. *Microbiology and Molecular Biology Reviews*. **67**, 238–276 (2003).
 89. Lee, C. Y. & Iandolo, J. J. Mechanism of bacteriophage conversion of lipase activity in *Staphylococcus aureus*. *Journal of Bacteriology*. **164**, 288–293 (1985).
 90. Day, J., Fernández, R. E. & Maurelli, A. T. Pathoadaptive mutations that enhance virulence: genetic organization of the cadA regions of *Shigella* spp. *Infection and Immunity*. **69**, 7471–7480 (2001).
 91. Susskind, M. M., Botstein, D. & Wright, A. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. III. Failure of superinfecting phage DNA to enter sieA⁺ lysogens. *Virology*. **62**, 350–366 (1974).
 92. Kliem, M. & Dreiseikelmann, B. The superimmunity gene sim of bacteriophage P1 causes superinfection exclusion. *Virology*. **171**, 350–355 (1989).
 93. Berngruber, T. W., Weissing, F. J. & Gandon, S. Inhibition of superinfection and the evolution of viral latency. *Journal of Virology*. **84**, 10200–10208 (2010).
 94. Selva, L., Viana, D., Regev-Yochay, G., Trzcinski, K., Corpa, J. M., Lasa, I., Novick, R. P. & Penadés, J. R. Killing niche competitors by remote-control bacteriophage induction. *Proceedings of the National Academy of Sciences. U. S. A.* **106**, 1234–8 (2009).
 95. Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A. S., Doran, K. S., Salamon, P., Youle, M. & Rohwer, F. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences*. **110**, 10771–10776 (2013).
 96. Ebel-Tsipis, J., Botstein, D. & Fox, M. S. Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. *Journal of Molecular Biology*. **71**, 433–448 (1972).
 97. Sandra, R. M. & Berger, H. Bacteriophage P1-mediated generalized transduction in *Escherichia coli*: Structure of abortively transduced DNA. *Virology*. **106**, 30–40 (1980).
 98. O'Shea, Y. A. & Boyd, E. F. Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction. *FEMS Microbiology Letters*. **214**, 153–157 (2002).
 99. Morse, M. L., Lederberg, E. M. & Lederberg, J. Transduction in *Escherichia Coli* K-12. *Genetics*. **41**, 142–56 (1956).
 100. Del Campillo-Campbell, A., Kayajanian, G., Campbell, A. & Adhya, S. Biotin-requiring mutants of *Escherichia coli* K-12. *Journal of Bacteriology*. **94**, 2065–2066 (1967).
 101. Novick, R. P. & Ram, G. The floating (pathogenicity) island: a genomic dessert. *Trends in Genetics*. **32**, 114–126 (2016).
 102. Faruque, S. M., Asadulghani, Kamruzzaman, M., Nandi, R. K., Ghosh, A. N., Balakrish Nair, G., Mekalanos, J. J. & Sack, D. A. RS1 element of *Vibrio cholerae* can propagate horizontally as a filamentous phage exploiting the morphogenesis genes of CTX Φ . *Infection and Immunity*. **70**, 163–170 (2002).

103. Six, E. W. & Klug, C. A. Bacteriophage P4: a satellite virus depending on a helper such as prophage P2. *Virology*. **51**, 327–44 (1973).
104. Sano, E., Carlson, S., Wegley, L. & Rohwer, F. Movement of viruses between biomes. *Applied and Environmental Microbiology*. **70**, 5842–5846 (2004).
105. Kenzaka, T., Tani, K. & Nasu, M. High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME Journal*. **4**, 648–659 (2010).
106. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature*. **499**, 219–222 (2013).
107. Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., Lewis, J. D. & Bushman, F. D. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research*. **21**, 1616–1625 (2011).
108. Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M. B. & Petit, M.-A. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME Journal*. **11**, 237–247 (2017).
109. Adriaenssens, E. M. & Rodney Brister, J. How to name and classify your phage: An informal guide. *Viruses*. **9**, 70 (2017).
110. Nelson, D. Phage taxonomy: we agree to disagree. *Journal of Bacteriology*. **186**, 7029–31 (2004).
111. Lima-Mendez, G., Van Helden, J., Toussaint, A. & Leplae, R. Reticulate representation of evolutionary and functional relationships between phage genomes. *Molecular Biology and Evolution*. **25**, 762–777 (2008).
112. Rohwer, F. & Edwards, R. The phage proteomic tree: a genome-based taxonomy for phage. *Journal of Bacteriology*. **184**, 4529–4535 (2002).
113. Ackermann, H.-W. Tailed bacteriophages: the order caudovirales. *Advances in Virus Research*. **51**, 135–201 (1998).
114. Lavigne, R. & Ceyssens, P.-J. Family – Myoviridae. in *Virus Taxonomy*. 46–62 (Elsevier, San Diego, US, 2012).
115. Jarvis, A. W., Collins, L. J. & Ackermann, H. W. A study of five bacteriophages of the Myoviridae family which replicate on different gram-positive bacteria. *Archives of Virology*. **133**, 75–84 (1993).
116. Brüssow, H. & Desiere, F. Comparative phage genomics and the evolution of siphoviridae: Insights from dairy phages. *Molecular Microbiology*. **39**, 213–222 (2001).
117. Hendrix, R. W., Casjens, S. R. & Lavigne, R. Family – Siphoviridae. in *Virus Taxonomy*. 86–98 (Elsevier, San Diego, US, 2012).
118. Lavigne, R. & Kropinski, A. Family – Podoviridae. in *Virus Taxonomy*. 63–85 (Elsevier, San Diego, 2012).
119. Trun, N. & Trempey, J. Bacteriophage. in *Fundamental Bacterial Genetics*. 105–126 (John Wiley & Sons, Oxford, UK, 2009).
120. Young, R. Bacteriophage lysis: mechanism and regulation. *Microbiological Reviews*. **56**, 430–481 (1992).
121. Young, R. & Bläsi, U. Holins: form and function in bacteriophage lysis. *FEMS Microbiological Reviews*. **17**, 195–205 (1995).
122. Nelson, D. C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D. G., Dong, S. & Donovan, D. M. *Endolysins as antimicrobials*. *Advances in virus research* **83**, (Elsevier Inc., 2012).
123. Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I. & Herskovits, A. A. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nature Reviews Microbiology*. **13**, 641–650 (2015).
124. Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobián-Güemes, A. G., Coutinho, F. H., Dinsdale, E. A., Felts, B., Furby, K. A., George, E. E., Green, K. T.,

- Gregoracci, G. B., Haas, A. F., Haggerty, J. M., Hester, E. R., Hisakawa, N., Kelly, L. W., Lim, Y. W. Corrigendum: lytic to temperate switching of viral communities. *Nature*. **539**, 123–123 (2016).
125. Fortier, L.-C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*. **4**, 354–365 (2013).
 126. Maiques, E., Úbeda, C., Campoy, S., Salvador, N., Lasa, Í., Novick, R. P., Barbé, J. & Penadés, J. R. β -lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *Journal of Bacteriology*. **188**, 2726–2729 (2006).
 127. Úbeda, C., Maiques, E., Knecht, E., Lasa, Í., Novick, R. P. & Penadés, J. R. Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in *staphylococci*. *Molecular Microbiology*. **56**, 836–844 (2005).
 128. Miller, C. SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. *Science*. **305**, 1629–1631 (2004).
 129. Favre, A., Chams, V. & Caldeira de Araujo, A. Photosensitized UVA light induction of the SOS response in *Escherichia coli*. *Biochimie*. **68**, 857–864 (1986).
 130. Phillips, I., Culebras, E., Moreno, F. & Baquero, F. Induction of the SOS response by new 4-quinolones. *Journal of Antimicrobial Chemotherapy*. **20**, 631–638 (1987).
 131. Kim, B. & Little, J. W. LexA and λ CI repressors as enzymes: specific cleavage in an intermolecular reaction. *Cell*. **73**, 1165–1173 (1993).
 132. Łoś, M. & Węgrzyn, G. Pseudolysogeny. *Advances in Virus Research*. **82**, 339–349 (2012).
 133. Campbell, A. Comparative molecular biology of lambdoid phages. *Annual Review of Microbiology*. **48**, 193–222 (1994).
 134. Botstein, D. A theory of modular evolution for bacteriophages. *Annals of the New York Academy of Sciences*. **354**, 484–90 (1980).
 135. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proceedings of the National Academy of Sciences*. **96**, 2192–2197 (1999).
 136. Casjens, S. Prophages and bacterial genomics: What have we learned so far? *Molecular Microbiology*. **49**, 277–300 (2003).
 137. Clark, A. J., Inwood, W., Cloutier, T. & Dhillon, T. S. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *Journal of Molecular Biology*. **311**, 657–679 (2001).
 138. Miold, S., Ehrbar, K., Weissmüller, A., Prager, R., Tschäpe, H., Rüssmann, H. & Hardt, W. D. Salmonella host cell invasion emerged by acquisition of a mosaic of separate genetic elements, including *Salmonella* pathogenicity island 1 (SPI1), SPI5, and sopE2. *Journal of Bacteriology*. **183**, 2348–2358 (2001).
 139. De Paepe, M., Hutinet, G., Son, O., Amarir-Bouhram, J., Schbath, S. & Petit, M.-A. A. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genetics*. **10**, e1004181 (2014).
 140. Martinsohn, J. T., Radman, M. & Petit, M.-A. A. The λ red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genetics*. **4**, e1000065 (2008).
 141. Swenson, K. M., Guertin, P., Deschênes, H. & Bergeron, A. Reconstructing the modular recombination history of *Staphylococcus aureus* phages. *BMC Bioinformatics*. **14**(Suppl 15):S17, (2013).
 142. Baker, J., Limberger, R., Schneider, S. J. & Campbell, A. Recombination and modular exchange in the genesis of new lambdoid phages. *New Biologist*. **3**, 297–308 (1991).
 143. Casjens, S. R. Comparative genomics and evolution of the tailed-bacteriophages. *Current Opinion in Microbiology*. **8**, 451–458 (2005).
 144. Friedman, D. I. & Gottesman, M. Lytic Mode of Lambda Development. *Lambda II*. **13**, 21-51

- (1983).
145. Ptashne, M. Phage repressors. In: strategy of the viral genome. *Ciba Foundation Symposium*. 155–68 (1971).
 146. Benson, N. & Youderian, P. Phage lambda Cro protein and cI repressor use two different patterns of specific protein-DNA interactions to achieve sequence specificity in vivo. *Genetics*. **121**, 5–12 (1989).
 147. Little, J. W. Autodigestion of *lexA* and phage lambda repressors. *Proceedings of the National Academy of Sciences*. **81**, 1375–1379 (1984).
 148. Gottesman, M. E. & Weisberg, R. A. Prophage insertion and excision. in *The Bacteriophage Lambda*. **2**, 113–138 (1971).
 149. Campbell, A. M. Chromosomal insertion sites for phages and plasmids. *Journal of Bacteriology*. **174**, 7495–7499 (1992).
 150. Weisberg, R. A. & Landy, A. Site-specific recombination in phage Lambda. *Lambda II*. **13**, 211–250 (1983).
 151. Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G. D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Research*. **41**, 8341–8356 (2013).
 152. Seah, N. E., Warren, D., Tong, W., Laxmikanthan, G., Van Duyne, G. D. & Landy, A. Nucleoprotein architectures regulating the directionality of viral integration and excision. *Proceedings of the National Academy of Sciences*. **111**, 12372–12377 (2014).
 153. Gottesman, M. E. & Yarmolinsky, M. B. Integration-negative mutants of bacteriophage lambda. *Journal of Molecular Biology*. **31**, 487–505 (1968).
 154. Schindler, D. & Echols, H. Retroregulation of the *int* gene of bacteriophage lambda: control of translation completion. *Proceedings of the National Academy of Sciences*. **78**, 4475–4479 (1981).
 155. Miller, H. I., Abraham, J., Benedik, M., Campbell, A., Court, D., Echols, H., Fischer, R., Galindo, J. M., Guarneros, G., Hernandez, T., Mascarenhas, D., Montanez, C., Schindler, D., Schmeissner, U. & Sosa, L. Regulation of the integration-excision reaction by bacteriophage lambda. *Cold Spring Harbor Symposia on Quantitative Biology*. **45 Pt 1**, 439–445 (1981).
 156. Guarneros, G. & Echols, H. New mutants of bacteriophage λ with a specific defect in excision from the host chromosome. *Journal of Molecular Biology*. **47**, 565–574 (1970).
 157. Kobiler, O., Oppenheim, A. B. & Herman, C. Recruitment of host ATP-dependent proteases by bacteriophage λ . in *Journal of Structural Biology*. **146**, 72–78 (2004).
 158. Craig, N. L. & Nash, H. A. *E. coli* integration host factor binds to specific sites in DNA. *Cell*. **39**, 707–716 (1984).
 159. Ball, C. A. & Johnson, R. C. Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. *Journal of Bacteriology*. **173**, 4027–31 (1991).
 160. Skalka, A., Poonian, M. & Bartl, P. Concatemers in DNA replication: electron microscopic studies of partially denatured intracellular lambda DNA. *Journal of Molecular Biology*. **64**, 541–50 (1972).
 161. Wake, R. G., Kaiser, A. D. & Inman, R. B. Isolation and structure of phage lambda head-mutant DNA. *Journal of Molecular Biology*. **64**, 519–40 (1972).
 162. Hwang, Y., Hang, J. Q., Neagle, J., Duffy, C. & Feiss, M. Endonuclease and helicase activities of bacteriophage lambda terminase: changing nearby residue 515 restores activity to the gpA K497D mutant enzyme. *Virology*. **277**, 204–14 (2000).
 163. Dodson, M., Echols, H., Wickner, S., Alfano, C., Mensa-Wilmot, K., Gomes, B., LeBowitz, J., Roberts, J. D. & McMacken, R. Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda: localized unwinding of duplex DNA by a six-protein reaction. *Proceedings of the National Academy of Sciences*. **83**, 7638–42 (1986).
 164. Bouche, J. P., Rowen, L. & Kornberg, A. The RNA primer synthesized by primase to initiate

- phage G4 DNA replication. *Journal of Biological Chemistry*. **253**, 765–769 (1978).
165. Espeli, O. & Marians, K. J. Untangling intracellular DNA topology. *Molecular Microbiology*. **52**, 925–931 (2004).
 166. Weigel, C., Seitz, H., Abril, A., Salas, M., Andreu, J., Hermoso, J., Rivas, G., Abril, A., Marco, S., Carrascosa, J., Salas, M., Hermoso, J., Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., Carniel, E., Ahmad, S., *et al.* Bacteriophage replication modules. *FEMS Microbiological Reviews*. **30**, 321–81 (2006).
 167. Takahashi, S. The starting point and direction of rolling-circle replicative intermediates of coliphage λ DNA. *MGG Molecular & General Genetics*. **142**, 137–153 (1975).
 168. Yarnell, W. S. Mechanism of intrinsic transcription termination and antitermination. *Science*. **284**, 611–615 (1999).
 169. Ferrer, M. D., Quiles-Puchalt, N., Harwich, M. D., Tormo-Más, M. Á., Campoy, S., Barbé, J., Lasa, Í., Novick, R. P., Christie, G. E. & Penadés, J. R. RinA controls phage-mediated packaging and transfer of virulence genes in Gram-positive bacteria. *Nucleic Acids Research*. **39**, 5866–5878 (2011).
 170. Quiles-Puchalt, N., Tormo-Más, M. Á., Campoy, S., Toledo-Arana, A., Monedero, V., Lasa, Í., Novick, R. P., Christie, G. E. & Penadés, J. R. A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. *Nucleic Acids Research*. **41**, 7260–7275 (2013).
 171. Desiere, F., Pridmore, R. D. & Brüssow, H. Comparative genomics of the late gene cluster from *Lactobacillus* phages. *Virology*. **275**, 294–305 (2000).
 172. Rixon, F. J. & Schmid, M. F. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. *Current Opinion in Virology*. **5**, 105–110 (2014).
 173. Oliveira, L., Tavares, P. & Alonso, J. C. Headful DNA packaging: bacteriophage SPP1 as a model system. *Virus Research*. **173**, 247–259 (2013).
 174. Casjens, S. R. & Hendrix, R. W. Bacteriophage lambda: early pioneer and still relevant. *Virology*. **479–480**, 310–330 (2015).
 175. Hendrix, R. W. & Johnson, J. E. Bacteriophage HK97 capsid assembly and maturation. *Advances in Experimental Medicine and Biology*. **726**, 351–363 (2012).
 176. Chang, J. R., Spilman, M. S. & Dokland, T. Assembly of bacteriophage P2 capsids from capsid protein fused to internal scaffolding protein. *Virus Genes*. **40**, 298–306 (2010).
 177. Aksyuk, A. A. & Rossmann, M. G. Bacteriophage assembly. *Viruses*. **3**, 172–203 (2011).
 178. Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V., Arisaka, F. & Rossmann, M. G. Structure and morphogenesis of bacteriophage T4. *Cellular and Molecular Life Sciences*. **60**, 2356–2370 (2003).
 179. Feiss, M. & Rao, V. B. The bacteriophage DNA packaging machine. *Advances in Experimental Medicine and Biology*. **726**, 489–509 (2012).
 180. Kaiser, D., Syvanen, M. & Masuda, T. DNA packaging steps in bacteriophage lambda head assembly. *Journal of Molecular Biology*. **91**, 175–86 (1975).
 181. Valpuesta, J. M. & Carrascosa, J. L. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Quarterly Reviews of Biophysics*. **27**, 107 (1994).
 182. Agirrezabala, X., Martín-Benito, J., Valle, M., González, J. M., Valencia, A., Valpuesta, J. M. & Carrascosa, J. L. Structure of the connector of bacteriophage T7 at 8 Å resolution: Structural homologies of a basic component of a DNA translocating machinery. *Journal of Molecular Biology*. **347**, 895–902 (2005).
 183. Lander, G. C. The Structure of an Infectious P22 Virion Shows the Signal for Headful DNA Packaging. *Science*. **312**, 1791–1795 (2006).
 184. Doan, D. N. P. & Dokland, T. The gpQ portal protein of bacteriophage P2 forms dodecameric connectors in crystals. *Journal of Structural Biology*. **157**, 432–436 (2007).

185. Morais, M. C., Koti, J. S., Bowman, V. D., Reyes-Aldrete, E., Anderson, D. L. & Rossmann, M. G. Defining molecular and domain boundaries in the bacteriophage ϕ 29 DNA packaging motor. *Structure*. **16**, 1267–1274 (2008).
186. Olson, N. H., Gingery, M., Eiserling, F. A. & Baker, T. S. The structure of isometric capsids of bacteriophage T4. *Virology*. **279**, 385–391 (2001).
187. Chen, D.-H., Baker, M. L., Hryc, C. F., DiMaio, F., Jakana, J., Wu, W., Dougherty, M., Haase-Pettingell, C., Schmid, M. F., Jiang, W., Baker, D., King, J. A. & Chiu, W. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proceedings of the National Academy of Sciences*. **108**, 1355–1360 (2011).
188. Oh, B., Moyer, C. L., Hendrix, R. W. & Duda, R. L. The delta domain of the HK97 major capsid protein is essential for assembly. *Virology*. **456–457**, 171–178 (2014).
189. Huet, A., Conway, J. F., Letellier, L. & Boulanger, P. In vitro assembly of the T=13 procapsid of bacteriophage T5 with its scaffolding domain. *Journal of Virology*. **84**, 9350–9358 (2010).
190. Rao, V. B. & Feiss, M. Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annual Review of Virology*. **2**, 351–378 (2015).
191. Dokland, T. Scaffolding proteins and their role in viral assembly. *Cellular and Molecular Life Sciences*. **56**, 580–603 (1999).
192. Weigele, P. R., Sampson, L., Winn-Stapley, D. & Casjens, S. R. Molecular genetics of bacteriophage P22 scaffolding protein's functional domains. *Journal of Molecular Biology*. **348**, 831–844 (2005).
193. Black, L.W., Showe, M.K., and Steven, A. . Morphogenesis of the T4 head. in *Molecular Biology of Bacteriophage T4*. (ed. Karam, J.) 218–258 (Am. Soc. Microbiol, 1994).
194. Rossmann, M. G., Mesyanzhinov, V. V, Arisaka, F. & Leiman, P. G. The bacteriophage T4 DNA injection machine. *Current Opinion in Structural Biology* **14**, 171–180 (2004).
195. Němeček, D., Gilcrease, E. B., Kang, S., Prevelige, P. E., Casjens, S. & Thomas, G. J. Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. *Journal of Molecular Biology*. **374**, 817–836 (2007).
196. Goetzinger, K. R. & Rao, V. B. Defining the ATPase center of bacteriophage T4 DNA packaging machine: Requirement for a catalytic glutamate residue in the large terminase protein gp17. *Journal of Molecular Biology*. **331**, 139–154 (2003).
197. Zhao, H., Christensen, T. E., Kamau, Y. N. & Tang, L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proceedings of the National Academy of Sciences*. **110**, 8075–80 (2013).
198. Oliveira, L., Henriques, A. O. & Tavares, P. Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency. *Journal of Biological Chemistry*. **281**, 21914–21923 (2006).
199. Hwang, Y. & Feiss, M. Mutations affecting the high affinity ATPase center of gpA, the large subunit of bacteriophage lambda terminase, inactivate the endonuclease activity of terminase. *Journal of Molecular Biology*. **261**, 524–35 (1996).
200. Smits, C., Chechik, M., Kovalevskiy, O. V, Shevtsov, M. B., Foster, A. W., Alonso, J. C. & Antson, A. A. Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Reports*. **10**, 592–598 (2009).
201. Cornilleau, C., Atmane, N., Jacquet, E., Smits, C., Alonso, J. C., Tavares, P. & Oliveira, L. The nuclease domain of the SPP1 packaging motor coordinates DNA cleavage and encapsidation. *Nucleic Acids Research*. **41**, 340–354 (2013).
202. Mitchell, M. S. Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. *Nucleic Acids Research*. **30**, 4009–4021 (2002).
203. Casjens, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schicklmaier, P., Schmieger, H., Pedulla, M. L., Ford, M. E., Houtz, J. M., Hatfull, G. F. & Hendrix, R. W. The generalized transducing *Salmonella* bacteriophage ES18: Complete genome sequence and DNA packaging strategy. *Journal of Bacteriology*. **187**, 1091–1104 (2005).

204. Camacho, A. G., Gual, A., Lurz, R., Tavares, P. & Alonso, J. C. *Bacillus subtilis* bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. *Journal of Biological Chemistry*. **278**, 23251–23259 (2003).
205. Ghosh-Kumar, M., Alam, T. I., Draper, B., Stack, J. D. & Rao, V. B. Regulation by interdomain communication of a headful packaging nuclease from bacteriophage T4. *Nucleic Acids Research*. **39**, 2742–2755 (2011).
206. Roy, A., Bhardwaj, A., Datta, P., Lander, G. C. & Cingolani, G. Small terminase couples viral DNA binding to genome-packaging ATPase activity. *Structure*. **20**, 1403–1413 (2012).
207. De Beer, T., Ortega, M., Berton, N., Yang, Q., Overduin, M. & Catalano, C. E. Letter to the editor: assignment of the ¹H, ¹³C, and ¹⁵N resonances of the DNA binding domain of gpNu1, a genome packaging protein from bacteriophage λ. *Journal of Biomolecular NMR*. **18**, 69–70 (2000).
208. Sun, S., Kondabagil, K., Draper, B., Alam, T. I., Bowman, V. D., Zhang, Z., Hegde, S., Fokine, A., Rossmann, M. G. & Rao, V. B. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell*. **135**, 1251–1262 (2008).
209. Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N. & Novick, R. P. Bacteriophage-mediated spread of bacterial virulence genes. *Current Opinion in Microbiology*. **23**, 171–178 (2015).
210. Rentas, F. J. & Rao, V. B. Defining the bacteriophage T4 DNA packaging machine: Evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17. *Journal of Molecular Biology*. **334**, 37–52 (2003).
211. García, P., Martínez, B., Obeso, J. M., Lavigne, R., Lurz, R. & Rodríguez, A. Functional genomic analysis of two *Staphylococcus aureus* phages isolated from the dairy environment. *Applied and Environmental Microbiology*. **75**, 7663–7673 (2009).
212. Casjens, S. R. & Gilcrease, E. B. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods in Molecular Biology*. **502**, 91–111 (2009).
213. Kenzaka, T., Tani, K., Sakotani, A., Yamaguchi, N. & Nasu, M. High-frequency phage-mediated gene transfer among *Escherichia coli* cells, determined at the single-cell level. *Applied and Environmental Microbiology*. **73**, 3291–3299 (2007).
214. Ziermann, R. & Calendar, R. Characterization of the cos sites of bacteriophages P2 and P4. *Gene*. **96**, 9–15 (1990).
215. Rao, V. B. & Feiss, M. The bacteriophage DNA packaging motor. *Annual Review of Genetics*. **42**, 647–681 (2008).
216. Sippy, J. & Feiss, M. Initial cos cleavage of bacteriophage λ concatemers requires proheads and gpFI in vivo. *Molecular Microbiology*. **52**, 501–513 (2004).
217. Chen, J., Carpena, N., Quiles-Puchalt, N., Ram, G., Novick, R. P. & Penadés, J. R. Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. *ISME Journal*. **9**, 1260–1263 (2014).
218. King, J. & Casjens, S. Catalytic head assembling protein in virus morphogenesis. *Nature*. **251**, 112–119 (1974).
219. Fane, B. A. & Prevelige, P. E. Mechanism of scaffolding-assisted viral assembly. *Advances in Protein Chemistry*. **64**, 259–299 (2003).
220. Spilman, M. S., Damle, P. K., Dearborn, A. D., Rodenburg, C. M., Chang, J. R., Wall, E. A., Christie, G. E. & Dokland, T. Assembly of bacteriophage 80α capsids in a *Staphylococcus aureus* expression system. *Virology*. **434**, 242–250 (2012).
221. Steven, A. C., Heymann, J. B., Cheng, N., Trus, B. L. & Conway, J. F. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Current Opinion in Structural Biology*. **15**, 227–236 (2005).
222. Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W. & Johnson, J. E. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*. **289**, 2129–33 (2000).

223. Banerjee, R., Nath, S., Ranjan, A., Khamrui, S., Pani, B., Sen, R. & Sen, U. The first structure of polarity suppression protein, Psu from Enterobacteria phage P4, reveals a novel fold and a knotted dimer. *Journal of Biological Chemistry*. **287**, 44667–44675 (2012).
224. Pani, B., Ranjan, A. & Sen, R. Interaction surface of bacteriophage P4 protein Psu required for complex formation with the transcription terminator Rho. *Journal of Molecular Biology*. **389**, 647–660 (2009).
225. Gilcrease, E. B., Winn-Stapley, D. A., Hewitt, F. C., Joss, L. & Casjens, S. R. Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. *Journal of Bacteriology*. **187**, 2050–2057 (2005).
226. Tang, L., Gilcrease, E. B., Casjens, S. R. & Johnson, J. E. Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. *Structure*. **14**, 837–845 (2006).
227. Lander, G. C., Evilevitch, A., Jeembaeva, M., Potter, C. S., Carragher, B. & Johnson, J. E. Bacteriophage Lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by Cryo-EM. *Structure*. **16**, 1399–1406 (2008).
228. Qin, L., Fokine, A., O'Donnell, E., Rao, V. B. & Rossmann, M. G. Structure of the small outer capsid protein, Soc: a clamp for stabilizing capsids of T4-like phages. *Journal of Molecular Biology*. **395**, 728–741 (2010).
229. King, J., Lenk, E. V & Botstein, D. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. *Journal of Molecular Biology*. **80**, 697–731 (1973).
230. Orlova, E. V, Gowen, B., Dröge, A., Stiege, A., Weise, F., Lurz, R., Van Heel, M. & Tavares, P. Structure of a viral DNA gatekeeper at 10 Å resolution by cryo-electron microscopy. *EMBO Journal*. **22**, 1255–1262 (2003).
231. Lhuillier, S., Gallopin, M., Gilquin, B., Brasiles, S., Lancelot, N., Letellier, G., Gilles, M., Dethan, G., Orlova, E. V, Couprie, J., Tavares, P. & Zinn-Justin, S. Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proceedings of the National Academy of Sciences*. **106**, 8507–8512 (2009).
232. Camacho, A., Jiménez, F., Viñuela, E. & Salas, M. Order of assembly of the lower collar and the tail proteins of *Bacillus subtilis* bacteriophage phi 29. *Journal of Virology*. **29**, 540–5 (1979).
233. Lander, G. C., Khayat, R., Li, R., Prevelige, P. E., Potter, C. S., Carragher, B. & Johnson, J. E. The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit. *Structure*. **17**, 789–799 (2009).
234. Casjens S, H. R. Control mechanisms in dsDNA bacteriophage assembly. in *The Bacteriophages*. 15–75. (1988).
235. Sciara, G., Bebeacua, C., Bron, P., Tremblay, D., Ortiz-Lombardia, M., Lichiere, J., van Heel, M., Campanacci, V., Moineau, S. & Cambillau, C. Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proceedings of the National Academy of Sciences*. **107**, 6852–6857 (2010).
236. Katsura, I. Mechanism of length determination in bacteriophage lambda tails. *Advances in Biophysics*. **26**, 1–18 (1990).
237. Auzat, I., Dröge, A., Weise, F., Lurz, R. & Tavares, P. Origin and function of the two major tail proteins of bacteriophage SPP1. *Molecular Microbiology*. **70**, 557–569 (2008).
238. Aksyuk, A. A., Leiman, P. G., Kurochkina, L. P., Shneider, M. M., Kostyuchenko, V. A., Mesyanzhinov, V. V & Rossmann, M. G. The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO Journal*. **28**, 821–829 (2009).
239. Zhao, L., Kanamaru, S., Chaidirek, C. & Arisaka, F. P15 and P3, the tail completion proteins of bacteriophage T4, both form hexameric rings. *Journal of Bacteriology*. **185**, 1693–1700 (2003).
240. Vianelli, A., Wang, G. R., Gingery, M., Duda, R. L., Eiserling, F. A. & Goldberg, E. B. Bacteriophage T4 self-assembly: localization of gp3 and its role in determining tail length. *Journal of Bacteriology*. **182**, 680–688 (2000).
241. Wang, J., Hofnung, M. & Charbit, A. The C-terminal portion of the tail fiber protein of

- bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of *Escherichia coli* K-12. *Journal of Bacteriology*. **182**, 508–512 (2000).
242. Chaturongakul, S. & Ounjai, P. Phage-host interplay: examples from tailed phages and Gram-negative bacterial pathogens. *Frontiers in Microbiology*. **5**, (2014).
 243. Heller, K. J. Identification of the phage gene for host receptor specificity by analyzing hybrid phages of T5 and BF23. *Virology*. **139**, 11–21 (1984).
 244. Coombs, D.H., and Arisaka, F. T4 tail structure and function. in *Molecular Biology of Bacteriophage T4*. 259–281 (1994).
 245. Sass, P. & Bierbaum, G. Lytic activity of recombinant bacteriophage 11 and 12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. *Applied and Environmental Microbiology*. **73**, 347–352 (2007).
 246. Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology*. **2**, 414–424 (2004).
 247. Hacker, J. & Carniel, E. Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. *EMBO Reports*. **2**, 376–381 (2001).
 248. Schmidt, H. & Hensel, M. Pathogenicity islands in bacterial pathogenesis. *Clinical Microbiology Reviews* **17**, 14–56 (2004).
 249. Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R. & Goebel, W. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extra intestinal *Escherichia coli* isolates. *Microbial Pathogenesis*. **8**, 213–225 (1990).
 250. Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschape, H. & Hacker, J. Excision of large DNA regions termed pathogenicity islands from tRNA- specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infection and Immunity*. **62**, 606–614 (1994).
 251. Groisman, E. A. & Ochman, H. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell*. **87**, 791–794 (1996).
 252. Hacker, J., Hochhut, B., Middendorf, B., Schneider, G., Buchrieser, C., Gottschalk, G. & Dobrindt, U. Pathogenomics of mobile genetic elements of toxigenic bacteria. *International Journal of Medical Microbiology*. **293**, 453–461 (2004).
 253. Dobrindt, U., Agerer, F., Michaelis, K., Janka, A., Buchrieser, C., Samuelson, M., Svanborg, C., Gottschalk, G., Karch, H. & Hacker, J. Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *Journal of Bacteriology*. **185**, 1831–1840 (2003).
 254. Lee, C. A. Pathogenicity islands and the evolution of bacterial pathogens. *Infectious Agents and Disease*. **5**, 1–7 (1996).
 255. Sullivan, J. T., Trzebiatowski, J. R., Cruickshank, R. W., Gouzy, J., Brown, S. D., Elliot, R. M., Fleetwood, D. J., McCallum, N. G., Rossbach, U., Stuart, G. S., Weaver, J. E., Webby, R. J., De Bruijn, F. J. & Ronson, C. W. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *Journal of Bacteriology*. **184**, 3086–3095 (2002).
 256. Gaillard, M., Vallaes, T., Vorhölter, F. J., Minoia, M., Werlen, C., Sentchilo, V., Pühler, A. & Van Meer, J. R. Der. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. *Journal of Bacteriology*. **188**, 1999–2013 (2006).
 257. Larbig, K. D., Christmann, A., Johann, A., Klockgether, J., Hartsch, T., Merkl, R., Wiehlmann, L., Fritz, H. J. & Tümmler, B. Gene islands integrated into tRNAGly genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *Journal of Bacteriology*. **184**, 6665–6680 (2002).
 258. Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L., Beanan, M., Dodson, R. J., Daugherty, S. C., Madupu, R., Angiuoli, S. V., Durkin, A. S., Haft, D. H., Vamathevan, J., Khouri, H., Utterback, T., et al. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *Journal of Bacteriology*. **187**, 2426–2438 (2005).

259. Ito, T., Okuma, K., Ma, X. X., Yuzawa, H. & Hiramatsu, K. Insights on antibiotic resistance of *Staphylococcus aureus* from its whole genome: genomic island SCC. *Drug Resistance Updates*. **6**, 41–52 (2003).
260. Baba, T., Bae, T., Schneewind, O., Takeuchi, F. & Hiramatsu, K. Genome sequence of *Staphylococcus aureus* strain newman and comparative analysis of staphylococcal genomes: Polymorphism and evolution of two major pathogenicity islands. *Journal of Bacteriology*. **190**, 300–10 (2008).
261. Lindsay, J. A. & Holden, M. T. G. *Staphylococcus aureus*: superbug, super genome? *Trends in Microbiology*. **12**, 378–385 (2004).
262. Novick, R. P., Christie, G. E. & Penadés, J. R. The phage-related chromosomal islands of Gram-positive bacteria. *Nature Reviews Microbiology*. **8**, 541–551 (2010).
263. Briani, F., Dehò, G., Forti, F. & Ghisotti, D. The plasmid status of satellite bacteriophage P4. *Plasmid*. **45**, 1–17 (2001).
264. Arnold, H. P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., Kristjansson, J. K., Garrett, R. & Zillig, W. The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Molecular Microbiology*. **34**, 217–226 (1999).
265. Úbeda, C., Tormo, M. Á., Cucarella, C., Trotonda, P., Foster, T. J., Lasa, Í. & Penadés, J. R. Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Molecular Microbiology*. **49**, 193–210 (2003).
266. Międzybrodzki, R., Borysowski, J., Weber-Dąbrowska, B., Fortuna, W., Letkiewicz, S., Szufnarowski, K., Pawełczyk, Z., Rogóż, P., Kłak, M., Wojtasik, E. & Górski, A. Clinical aspects of phage therapy. *Advances in virus research*. **83**, (2012).
267. Quiles-Puchalt, N., Martínez-Rubio, R., Ram, G., Lasa, Í. & Penadés, J. R. Unravelling bacteriophage ϕ 11 requirements for packaging and transfer of mobile genetic elements in *Staphylococcus aureus*. *Molecular Microbiology*. **91**, 423–437 (2014).
268. Maiques, E., Ubeda, C., Tormo, M. A., Ferrer, M. D., Lasa, I., Novick, R. P. & Penadés, J. R. Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. *Journal of Bacteriology*. **189**, 5608–16 (2007).
269. Mir-Sanchis, I., Martínez-Rubio, R., Martí, M., Chen, J., Lasa, Í., Novick, R. P., Tormo-Más, M. Á. & Penadés, J. R. Control of *Staphylococcus aureus* pathogenicity island excision. *Molecular Microbiology*. **85**, 833–845 (2012).
270. Ruzin, A., Lindsay, J. & Novick, R. P. Molecular genetics of SaPI1--a mobile pathogenicity island in *Staphylococcus aureus*. *Molecular Microbiology*. **41**, 365–77 (2001).
271. Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I. & Penades, J. R. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *Journal of Bacteriology*. **183**, 2888–2896 (2001).
272. Chen, H. J., Chang, Y. C., Tsai, J. C., Hung, W. C., Lin, Y. T., You, S. J., Tseng, S. P. & Teng, L. J. New structure of phage-related islands carrying fusB and a virulence gene in fusidic acid-resistant *Staphylococcus epidermidis*. *Antimicrobial Agents and Chemotherapy*. **57**, 5737–5739 (2013).
273. O’Neill, A. J., Larsen, A. R., Skov, R., Henriksen, A. S. & Chopra, I. Characterization of the epidemic European fusidic acid-resistant impetigo clone of *Staphylococcus aureus*. *Journal of Clinical Microbiology*. **45**, 1505–1510 (2007).
274. Scott, J., Nguyen, S. V., King, C. J., Hendrickson, C. & McShan, W. M. Phage-like *Streptococcus pyogenes* chromosomal islands (SpyCi) and mutator phenotypes: Control by growth state and rescue by a SpyCi-encoded promoter. *Frontiers in Microbiology*. **3**, 317 (2012).
275. Matos, R. C., Lapaque, N., Rigottier-Gois, L., Debarbieux, L., Meylheuc, T., Gonzalez-Zorn, B., Repoila, F., Lopes, M. de F. & Serror, P. *Enterococcus faecalis* Prophage Dynamics and Contributions to Pathogenic Traits. *PLoS Genetics*. **9**, e1003539 (2013).
276. Roser Martínez-Rubio, Nuria Quiles-Puchalt, Miguel Martí, Geeta Ram, Davida Smyth,

- John Chen, Richard P Novick, J. R. P. Phage-inducible islands in the coccal sea. *The ISME Journal*. (2016).
277. Martínez-Rubio, R., Quiles-Puchalt, N., Martí, M., Humphrey, S., Ram, G., Smyth, D., Chen, J., Novick, R. P. & Penadés, J. R. Phage-inducible islands in the Gram-positive cocci. *ISME Journal*. **11**, 1–14 (2016).
 278. Úbeda, C., Maiques, E., Barry, P., Matthews, A., Tormo, M. Á., Lasa, Í., Novick, R. P. & Penadés, J. R. SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Molecular Microbiology*. **67**, 493–503 (2008).
 279. Ram, G., Chen, J., Kumar, K., Ross, H. F., Ubeda, C., Damle, P. K., Lane, K. D., Penadés, J. R., Christie, G. E., Novick, R. P., Penades, J. R., Christie, G. E. & Novick, R. P. Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proceedings of the National Academy of Sciences*. **109**, 16300–16305 (2012).
 280. Tormo-Más, M. Á. A., Mir, I., Shrestha, A., Tallent, S. M., Campoy, S., Lasa, Í., Barbé, J., Novick, R. P., Christie, G. E., Penadés, J. R., Lasa, I., Barbé, J., Novick, R. P., Christie, G. E., Penadés, J. R., Tormo-Mas, M. A., Mir, I., Shrestha, A., Tallent, S. M., *et al.* Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature*. **465**, 779–782 (2010).
 281. Frígols, B., Quiles-Puchalt, N., Mir-Sanchis, I., Donderis, J., Elena, S. F., Buckling, A., Novick, R. P., Marina, A. & Penadés, J. R. Virus Satellites Drive Viral Evolution and Ecology. *PLoS Genetics*. **11**, e1005609 (2015).
 282. Lane, K. Transcriptional crosstalk between helper bacteriophages and Staphylococcal aureus pathogenicity islands. A dissertation submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy at Virginia Commonwealth University. (2013).
 283. Penadés, J. R. & Christie, G. E. The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. *Annual Review of Virology*. **2**, 181–201 (2015).
 284. Simossis, V. A. & Heringa, J. The PRALINE online server: optimising progressive multiple alignment on the web. *Computational Biology and Chemistry*. **27**, 511–519 (2003).
 285. Simossis, V. A. & Heringa, J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Research*. **33**, W289–W294 (2005).
 286. Ubeda, C., Olivarez, N. P., Barry, P., Wang, H., Kong, X., Matthews, A., Tallent, S. M., Christie, G. E. & Novick, R. P. Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Molecular Microbiology*. **72**, 98–108 (2009).
 287. Campbell, A. Phage integration and chromosome structure. A personal history. *Annual Review of Genetics*. **41**, 1–11 (2007).
 288. Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science*. **323**, 139–141 (2009).
 289. Ubeda, C., Barry, P., Penadés, J. R. & Novick, R. P. A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proceedings of the National Academy of Sciences*. **104**, 14182–8 (2007).
 290. Ubeda, C., Tormo-Más, M. Á., Penadés, J. R. & Novick, R. P. Structure-function analysis of the SaPI_{bov1} replication origin in *Staphylococcus aureus*. *Plasmid*. **67**, 183–90 (2012).
 291. Tallent, S. M., Langston, T. B., Moran, R. G. & Christie, G. E. Transducing particles of *Staphylococcus aureus* pathogenicity island SaPI1 are comprised of helper phage-encoded proteins. *Journal of Bacteriology*. **189**, 7520–4 (2007).
 292. Tormo, M. Á., Ferrer, M. D., Maiques, E., Úbeda, C., Selva, L., Lasa, Í., Calvete, J. J., Novick, R. P. & Penadés, J. R. *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *Journal of Bacteriology*. **190**, 2434–2440 (2008).
 293. Úbeda, C., Maiques, E., Tormo, M., Campoy, S., Lasa, Í., Barbé, J., Novick, R. P. &

- Penadés, J. R. SaPI operon I is required for SaPI packaging and is controlled by LexA. *Molecular Microbiology*. **65**, 41–50 (2007).
294. Damle, P. K., Wall, E. A., Spilman, M. S., Dearborn, A. D., Ram, G., Novick, R. P., Dokland, T. & Christie, G. E. The roles of SaPI1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. *Virology*. **432**, 277–282 (2012).
 295. Poliakov, A., Chang, J. R., Spilman, M. S., Damle, P. K., Christie, G. E., Mobley, J. A. & Dokland, T. Capsid size determination by *Staphylococcus aureus* pathogenicity island SaPI1 involves specific incorporation of SaPI1 proteins into procapsids. *Journal of Molecular Biology*. **380**, 465–475 (2008).
 296. Ram, G., Chen, J., Ross, H. F. & Novick, R. P. An insight into staphylococcal pathogenicity island-mediated interference with phage late gene transcription. *Bacteriophage*. **5**, e1028608 (2015).
 297. Fitzgerald, J. R., Monday, S. R., Foster, T. J., Bohach, G. A., Hartigan, P. J., Meaney, W. J. & Smyth, C. J. Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *Journal of Bacteriology*. **183**, 63–70 (2001).
 298. Diep, B. A., Gill, S. R., Chang, R. F., Phan, T. H. Van, Chen, J. H., Davidson, M. G., Lin, F., Lin, J., Carleton, H. A., Mongodin, E. F., Sensabaugh, G. F. & Perdreau-Remington, F. Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet*. **367**, 731–739 (2006).
 299. Fitzgerald, J. R., Hartigan, P. J., Meaney, W. J. & Smyth, C. J. Molecular population and virulence factor analysis of *Staphylococcus aureus* from bovine intramammary infection. *Journal of Applied Microbiology*. **88**, 1028–1037 (2000).
 300. Malachowa, N. & DeLeo, F. R. Mobile genetic elements of *Staphylococcus aureus*. *Cellular and Molecular Life Sciences*. **67**, 3057–71 (2010).
 301. Yamaguchi, T., Nishifuji, K., Sasaki, M., Fudaba, Y., Aepfelbacher, M., Takata, T., Ohara, M., Komatsuzawa, H., Amagai, M. & Sugai, M. Identification of the *Staphylococcus aureus* etd pathogenicity island which encodes a novel exfoliative toxin, ETD, and EDIN-B. *Infection and Immunity*. **70**, 5835–5845 (2002).
 302. Van Wamel, W. J. B., Rooijackers, S. H. M., Ruyken, M., Van Kessel, K. P. M. & Van Strijp, J. A. G. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on β -hemolysin-converting bacteriophages. *Journal of Bacteriology*. **188**, 1310–1315 (2006).
 303. Kuroda, M., Yamashita, A., Hirakawa, H., Kumano, M., Morikawa, K., Higashide, M., Maruyama, A., Inose, Y., Matoba, K., Toh, H., Kuhara, S., Hattori, M. & Ohta, T. Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection. *Proceedings of the National Academy of Sciences*. **102**, 13272–7 (2005).
 304. Maiques, E., Quiles-Puchalt, N., Donderis, J., Ciges-Tomas, J. R., Alite, C., Bowring, J. Z., Humphrey, S., Penadés, J. R. & Marina, A. Another look at the mechanism involving trimeric dUTPases in *Staphylococcus aureus* pathogenicity island induction involves novel players in the party. *Nucleic Acids Research*. **44**, 5457–5469 (2016).
 305. Tormo-Más, M. Á., Donderis, J., García-Caballer, M., Alt, A., Mir-Sanchis, I., Marina, A. & Penadés, J. R. Phage dUTPases control transfer of virulence genes by a proto-oncogenic G protein-like mechanism. *Molecular Cell*. **49**, 947–958 (2013).
 306. Leveles, I., Németh, V., Szabó, J. E., Harmat, V., Nyíri, K., Bendes, Á. Á., Papp-Kádár, V., Zagyva, I., Róna, G., Ozohanics, O., Vékey, K., Tóth, J. & Vértessy, B. G. Structure and enzymatic mechanism of a moonlighting dUTPase. *Acta Crystallographica Section D*. **69**, 2298–2308 (2013).
 307. Christie, G. E., Matthews, A. M., King, D. G., Lane, K. D., Olivarez, N. P., Tallent, S. M., Gill, S. R. & Novick, R. P. The complete genomes of *Staphylococcus aureus* bacteriophages 80 and 80 α —Implications for the specificity of SaPI mobilization. *Virology*. **407**, 381–390 (2010).
 308. Quiles-Puchalt, N., Carpena, N., Alonso, J. C., Novick, R. P., Marina, A. & Penades, J. R.

- Staphylococcal pathogenicity island DNA packaging system involving cos-site packaging and phage-encoded HNH endonucleases. *Proceedings of the National Academy of Sciences*. **111**, 6016–6021 (2014).
309. Neamah, M. M., Mir-Sanchis, I., Lopez-Sanz, M., Acosta, S., Baquedano, I., Haag, A. F., Marina, A., Ayora, S. & Penadés, J. R. Sak and Sak4 recombinases are required for bacteriophage replication in *Staphylococcus aureus*. *Nucleic Acids Research*. **45**, 6507–6519 (2017).
 310. Bowring, J., Neamah, M. M., Donderis, J., Mir-Sanchis, I., Alite, C., Ciges-Tomas, J. R., Maiques, E., Medmedov, I., Marina, A. & Penadés, J. R. Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer. *Elife*. **6**, e26487 (2017).
 311. Alite, C., Humphrey, S., Donderis, J., Maiques, E., Ciges-Tomas, J. R., Penadés, J. R. & Marina, A. Dissecting the link between the enzymatic activity and the SaPI inducing capacity of the phage 80 α dUTPase. *Scientific Reports*. **7**, 11234 (2017).
 312. Penadés, J. R., Donderis, J., García-Caballer, M., Tormo-Más, M. Á. & Marina, A. dUTPases, the unexplored family of signalling molecules. *Current Opinion in Microbiology*. **16**, 163–170 (2013).
 313. Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T., McCarty, J., Srikumar, R., Williams, D., Wu, J. J., Gros, P., Pelletier, J. & DuBow, M. Antimicrobial drug discovery through bacteriophage genomics. *Nature Biotechnology*. **22**, 185–191 (2004).
 314. Dearborn, A. D., Spilman, M. S., Damle, P. K., Chang, J. R., Monroe, E. B., Saad, J. S., Christie, G. E. & Dokland, T. The *Staphylococcus aureus* pathogenicity island 1 protein gp6 functions as an internal scaffold during capsid size determination. *Journal of Molecular Biology*. **412**, 710–722 (2011).
 315. Shore, D., Dehò, G., Tsipis, J. & Goldstein, R. Determination of capsid size by satellite bacteriophage P4. *Proceedings of the National Academy of Sciences*. **75**, 400–4 (1978).
 316. Spilman, M. S., Dearborn, A. D., Chang, J. R., Damle, P. K., Christie, G. E. & Dokland, T. A conformational switch involved in maturation of *Staphylococcus aureus* bacteriophage 80 α capsids. *Journal of Molecular Biology*. **405**, 863–876 (2011).
 317. Dearborn, A. D. & Dokland, T. Mobilization of pathogenicity islands by *Staphylococcus aureus* strain Newman bacteriophages. *Bacteriophage*. **2**, 70–78 (2012).
 318. Kala, S., Cumby, N., Sadowski, P. D., Hyder, B. Z., Kanelis, V., Davidson, A. R. & Maxwell, K. L. HNH proteins are a widespread component of phage DNA packaging machines. *Proceedings of the National Academy of Sciences*. **111**, 6022–6027 (2014).
 319. Bento, J. C., Lane, K. D., Read, E. K., Cerca, N. & Christie, G. E. Sequence determinants for DNA packaging specificity in the *S. aureus* pathogenicity island SaPI1. *Plasmid*. **71**, 8–15 (2014).
 320. Carpena, N., Manning, K. A., Dokland, T., Marina, A. & Penadés, J. R. Convergent evolution of pathogenicity islands in helper cos phage interference. *Philosophical Transactions of the Royal Society B Biological Sciences*. **371**, 20150505 (2016).
 321. Lagos, R., Jiang, R. Z., Kim, S. & Goldstein, R. Rho-dependent transcription termination of a bacterial operon is antagonized by an extrachromosomal gene product. *Proceedings of the National Academy of Sciences*. **83**, 9561–9565 (1986).
 322. Ram, G., Chen, J., Ross, H. F. & Novick, R. P. Precisely modulated pathogenicity island interference with late phage gene transcription. *Proceedings of the National Academy of Sciences*. **111**, 14536–14541 (2014).
 323. Rountree, P. M. The Phenomenon of lysogenicity in *Staphylococci*. *Journal of General Microbiology*. **3**, 153–163 (1949).
 324. Tzipilevich, E., Habusha, M. & Ben-Yehuda, S. Acquisition of phage sensitivity by bacteria through exchange of phage receptors. *Cell*. **168**, 186–199.e12 (2017).
 325. Chen, J., Ram, G., Penadés, J. R., Brown, S. & Novick, R. P. Pathogenicity island-directed transfer of unlinked chromosomal virulence genes. *Molecular Cell*. **57**, 138–149 (2015).

326. Scott, J., Thompson-Mayberry, P., Lahmamsi, S., King, C. J. & McShan, W. M. Phage-associated mutator phenotype in group A *Streptococcus*. *Journal of Bacteriology*. **190**, 6290–6301 (2008).
327. Nguyen, S. V & McShan, W. M. Chromosomal islands of *Streptococcus pyogenes* and related *Streptococci*: molecular switches for survival and virulence. *Frontiers in Cellular and Infection Microbiology*. **4**, 109 (2014).
328. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*. **494**, 489–491 (2013).
329. Vértessy, B. G. & Tóth, J. Keeping Uracil Out of DNA: Physiological role, structure and catalytic mechanism of dUTPases. *Accounts of Chemical Research*. **42**, 97–106 (2009).
330. Ariza, M.-E., Glaser, R., Kaumaya, P. T. P., Jones, C. & Williams, M. V. The EBV-encoded dUTPase activates NF-kappa B through the TLR2 and MyD88-dependent signaling pathway. *Journal of Immunology*. **182**, 851–9 (2009).
331. Leang, R. S., Wu, T. T., Hwang, S., Liang, L. T., Tong, L., Truong, J. T. & Sun, R. The anti-interferon activity of conserved viral dUTPase ORF54 is essential for an effective MHV-68 infection. *PLoS Pathogens*. **7**, e1002292 (2011).
332. Madrid, A. S. & Ganem, D. Kaposi's Sarcoma-associated herpesvirus ORF54/dUTPase downregulates a ligand for the NK activating receptor NKp44. *Journal of Virology*. **86**, 8693–8704 (2012).
333. Ariza, M.-E. & Williams, M. V. A human endogenous retrovirus K dUTPase triggers a TH1, TH17 cytokine response: does it have a role in psoriasis? *Journal of Investigative Dermatology*. **131**, 2419–2427 (2011).
334. Chu, R., Lin, Y., Rao, M. S. & Reddy, J. K. Cloning and identification of rat deoxyuridine triphosphatase as an inhibitor of peroxisome proliferator-activated receptor. *Journal of Biological Chemistry*. **271**, 27670–27676 (1996).
335. Krokan, H. E., Drabløs, F. & Slupphaug, G. Uracil in DNA – occurrence, consequences and repair. *Oncogene*. **21**, 8935–8948 (2002).
336. Slupphaug, G., Eftedal, I., Kavli, B., Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Hellet, N. M., Haug, T., Krokan, H. E. & Levine, D. W. Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry*. **34**, 128–138 (1995).
337. Seeberg, E., Eide, L. & Bjørås, M. The base excision repair pathway. *Trends in Biochemical Sciences*. **20**, 391–397 (1995).
338. Ahmad, S. I., Kirk, S. H. & Eisenstark, A. Thymine metabolism and thymineless death in prokaryotes and eukaryotes. *Annual Review of Microbiology*. **52**, 591–625 (1998).
339. O'Donovan, G. A., Edlin, G., Fuchs, J. A., Neuhard, J. & Thomassen, E. Deoxycytidine triphosphate deaminase: characterization of an *Escherichia coli* mutant deficient in the enzyme. *Journal of Bacteriology*. **105**, 666–672 (1971).
340. Bianchi, V., Pontis, E. & Reichard, P. Regulation of pyrimidine deoxyribonucleotide metabolism by substrate cycles in dCMP deaminase-deficient V79 hamster cells. *Molecular Cell Biology*. **7**, 4218–24 (1987).
341. el-Hajj, H. H., Zhang, H. & Weiss, B. Lethality of a dut (deoxyuridine triphosphatase) mutation in *Escherichia coli*. *Journal of Bacteriology*. **170**, 1069–1075 (1988).
342. Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO Journal*. **12**, 4425–31 (1993).
343. Baldo, A. M. & McClure, M. A. Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. *Journal of Virology*. **73**, 7710–21 (1999).
344. Tóth, J., Varga, B., Kovács, M., Málnási-Csizmadia, A. & Vértessy, B. G. Kinetic mechanism of human dUTPase, an essential nucleotide pyrophosphatase enzyme. *Journal of Biological Chemistry*. **282**, 33572–33582 (2007).

345. Kerepesi, C., Szabó, J. E., Papp-Kádár, V., Dobay, O., Szabó, D., Grolmusz, V. & Vértessy, B. G. Life without dUTPase. *Frontiers in Microbiology*. **7**, 1768 (2016).
346. Maul, R. W. & Gearhart, P. J. AID and somatic hypermutation. *Advances in Immunology*. **105**, 159–191 (2010).
347. Yan, N., O'Day, E., Wheeler, L. A., Engelman, A. & Lieberman, J. HIV DNA is heavily uracilated, which protects it from autointegration. *Proceedings of the National Academy of Sciences*. **108**, 9244–9249 (2011).
348. Muha, V., Horváth, A., Békési, A., Pukáncsik, M., Hodoscsek, B., Merényi, G., Róna, G., Batki, J., Kiss, I., Jankovics, F., Vilmos, P., Erdélyi, M. & Vértessy, B. G. Uracil-containing DNA in *Drosophila*: stability, stage-specific accumulation, and developmental involvement. *PLoS Genetics*. **8**, e1002738 (2012).
349. Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. Crystal structure of a dUTPase. *Nature*. **355**, 740–3 (1992).
350. Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. Human dUTP pyrophosphatase: uracil recognition by a β hairpin and active sites formed by three separate subunits. *Structure*. **4**, 1077–1092 (1996).
351. Vértessy, B. G., Larsson, G., Persson, T., Bergman, A. C., Persson, R. & Nyman, P. O. The complete triphosphate moiety of non-hydrolyzable substrate analogues is required for a conformational shift of the flexible C-terminus in *E. coli* dUTP pyrophosphatase. *FEBS Letters*. **421**, 83–88 (1998).
352. Varga, B., Barabás, O., Kovári, J., Tóth, J., Hunyadi-Gulyás, É., Klement, É., Medzihradsky, K. F., Tölgyesi, F., Fidy, J. & Vértessy, B. G. Active site closure facilitates juxtaposition of reactant atoms for initiation of catalysis by human dUTPase. *FEBS Letters*. **581**, 4783–4788 (2007).
353. Pecs, I., Szabo, J. E., Adams, S. D., Simon, I., Sellers, J. R., Vértessy, B. G. & Toth, J. Nucleotide pyrophosphatase employs a P-loop-like motif to enhance catalytic power and NDP/NTP discrimination. *Proceedings of the National Academy of Sciences*. **108**, 14437–14442 (2011).
354. Freeman, L., Buisson, M., Tarbouriech, N., Van der Heyden, A., Labbé, P. & Burmeister, W. P. The flexible motif V of Epstein-Barr virus deoxyuridine 5'-Triphosphate pyrophosphatase is essential for catalysis. *Journal of Biological Chemistry*. **284**, 25280–25289 (2009).
355. Hemsworth, G. R., Moroz, O. V., Fogg, M. J., Scott, B., Bosch-Navarrete, C., González-Pacanowska, D. & Wilson, K. S. The crystal structure of the *Leishmania major* deoxyuridine triphosphate nucleotidohydrolase in complex with nucleotide analogues, dUMP, and deoxyuridine. *Journal of Biological Chemistry*. **286**, 16470–16481 (2011).
356. Harkiolaki, M., Dodson, E. J., Bernier-Villamor, V., Turkenburg, J. P., González-Pacanowska, D. & Wilson, K. S. The crystal structure of *Trypanosoma cruzi* dUTPase reveals a novel dUTP/dUDP binding fold. *Structure*. **12**, 41–53 (2004).
357. Hemsworth, G. R. R., González-Pacanowska, D. & Wilson, K. S. S. On the catalytic mechanism of dimeric dUTPases. *Biochemical Journal*. **456**, (2013).
358. Moroz, O. V., Harkiolaki, M., Galperin, M. Y., Vagin, A. A., González-Pacanowska, D. & Wilson, K. S. The crystal structure of a complex of *Campylobacter jejuni* dUTPase with substrate analogue sheds light on the mechanism and suggests the 'basic module' for dimeric d(C/U)TPases. *Journal of Molecular Biology*. **342**, 1583–1597 (2004).
359. Hill, R. L. L. & Dokland, T. The type 2 dUTPase of bacteriophage ϕ NM1 initiates mobilization of *Staphylococcus aureus* bovine pathogenicity island 1. *Journal of Molecular Biology*. **428**, 142–152 (2016).
360. Ariza, M. E., Glaser, R. & Williams, M. V. Human herpesviruses-encoded dUTPases: a family of proteins that modulate dendritic cell function and innate immunity. *Frontiers in Microbiology*. **5**, 504 (2014).
361. Tarbouriech, N., Buisson, M., Seigneurin, J.-M., Cusack, S., Burmeister, W. P., Jones, A. F., Berry, C., Benito, A., Turkenburg, J. P., Dodson, E. J. & Al., E. The monomeric dUTPase from Epstein-Barr virus mimics trimeric dUTPases. *Structure*. **13**, 1299–1310 (2005).

362. Benedek, A., Horváth, A., Hirmondó, R., Ozohanics, O., Békési, A., Módos, K., Révész, Á., Vékey, K., Nagy, G. N. & Vértessy, B. G. Potential steps in the evolution of a fused trimeric all- β dUTPase involve a catalytically competent fused dimeric intermediate. *FEBS Journal*. **283**, 3268–3286 (2016).
363. Chan, S., Segelke, B., Legin, T., Krupka, H., Cho, U. S., Kim, M., So, M., Kim, C.-Y., Naranjo, C. M., Rogers, Y. C., Park, M. S., Waldo, G. S., Pashkov, I., Cascio, D., Perry, J. L. & Sawaya, M. R. Crystal structure of the *Mycobacterium tuberculosis* dUTPase: insights into the catalytic mechanism. *Journal of Molecular Biology*. **341**, 503–517 (2004).
364. Békési, A., Zagya, I., Hunyadi-Gulyás, É., Pongrácz, V., Kovári, J., Nagy, Á. O., Erdei, A., Medzihradsky, K. F. & Vértessy, B. G. Developmental regulation of dUTPase in *Drosophila melanogaster*. *Journal of Biological Chemistry*. **279**, 22362–22370 (2004).
365. Pecsí, I., Hirmondo, R., Brown, A. C., Lopata, A., Parish, T., Vértessy, B. G. & Tóth, J. The dutpase enzyme is essential in *Mycobacterium smegmatis*. *PLoS One*. **7**, e37461 (2012).
366. Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., Matsuoka, M., Taylor, G. M., Donoghue, H. D., Bouwman, A., Mays, S., Watson, C., Lockwood, D., Khamispour, A., Dowlati, Y., Jianping, S., Rea, T. H., Vera-Cabrera, L., Stefani, M. M., *et al.* Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nature Genetics*. **41**, 1282–1289 (2009).
367. Juganaru, M., Reina, R., Bertolotti, L., Stella, M. C., Profiti, M., Armentano, M., Bollo, E., Amorena, B. & Rosati, S. In vitro properties of small ruminant lentivirus genotype E. *Virology*. **410**, 88–95 (2011).
368. Szabo, J. E., Nemeth, V., Papp-Kadar, V., Nyiri, K., Leveles, I., Bendes, A. A., Zagya, I., Rona, G., Palinkas, H. L., Besztercei, B., Ozohanics, O., Vekey, K., Liliom, K., Toth, J. & Vértessy, B. G. Highly potent dUTPase inhibition by a bacterial repressor protein reveals a novel mechanism for gene expression control. *Nucleic Acids Research*. **42**, 11912–11920 (2014).
369. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. **9**, 40 (2008).
370. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nature Methods*. **12**, 7–8 (2014).
371. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. **5**, 725–738 (2010).
372. Moroz, O. V., Murzin, A. G., Makarova, K. S., Koonin, E. V., Wilson, K. S. & Galperin, M. Y. Dimeric dUTPases, HisE, and MazG belong to a new superfamily of all- α NTP Pyrophosphohydrolases with potential ‘house-cleaning’ functions. *Journal of Molecular Biology*. **347**, 243–255 (2005).
373. Donderis, J., Bowring, J., Maiques, E., Ciges-Tomas, J. R., Alite, C., Mehmedov, I., Tormo-Mas, M. A., Penadés, J. R. & Marina, A. Convergent evolution involving dimeric and trimeric dUTPases in pathogenicity island mobilization. *PLoS Pathogens*. **13**, e1006581 (2017).
374. Hidalgo-Zarco, F. & Gonzalez-Pacanowska, D. Trypanosomal dUTPases as potential targets for drug design. *Current Protein & Peptide Science. Sci*. **2**, 389–397 (2001).
375. Charpentier, E., Anton, A. I., Barry, P., Alfonso, B., Fang, Y. & Novick, R. P. Novel cassette-based shuttle vector system for gram-positive bacteria. *Applied and Environmental Microbiology*. **70**, 6076–85 (2004).
376. Kreiswirth, B., Löfdahl, S., Betley, M., O'Reilly, M., Schlievert, P., Bergdoll, M. & Novick, R. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature*. **305**, 709–12 (1983).
377. Evans, G. A. Molecular cloning: A laboratory manual. Second edition. Volumes 1, 2, and 3. Current protocols in molecular biology. Volumes 1 and 2. *Cell*. **61**, 17–18 (Cold Spring Harbor Laboratory Press, New York, US, 1990).
378. Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, G. J., Smith, J. A., Struhl, K. Wiley: Current protocols in molecular biology (John Wiley & Sons, Ltd, Hoboken, US, 1990).

379. Kim, K.-J., Kim, H.-E., Lee, K.-H., Han, W., Yi, M.-J., Jeong, J. & Oh, B.-H. Two-promoter vector is highly efficient for overproduction of protein complexes. *Protein Science*. **13**, 1698–1703 (2004).
380. Straus, D. & Ausubel, F. M. Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proceedings of the National Academy of Sciences*. **87**, 1889–93 (1990).
381. Lindsay, J. A. *Staphylococcus* : molecular genetics. (Caister Academic Press, 2008).
382. Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V, Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. & Mann, M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences*. **93**, 14440–14445 (1996).
383. West, B. & Kelly, F. C. Relationship of sodium chloride tolerance and serological group of staphylococcal phages. *Journal of General Microbiology*. **28**, 367–373 (1962).
384. Arnaud, M., Chastanet, A. & Debarbouille, M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, Gram-positive bacteria. *Applied and Environmental Microbiology*. **70**, 6887–6891 (2004).
385. Stemmer, K., Burghardt, I., Mayer, C., Reinicke, G. B., Wägele, H., Tollrian, R. & Leese, F. Morphological and genetic analyses of xeniid soft coral diversity (Octocorallia; Alcyonacea). *Organisms Diversity & Evolution*. **13**, 135–150 (2013).
386. Schroedinger, L. The {PyMOL} molecular graphics system, version 1.7.2.1. (2015).
387. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*. **42**, W320-4 (2014).
388. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. **10**, 845–858 (2015).
389. Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pašić, L., Thingstad, T. F., Rohwer, F. & Mira, A. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology*. **7**, 828–836 (2009).
390. Cordero, O. X. & Polz, M. F. Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology*. **12**, 263–273 (2014).
391. Salmond, G. P. C. & Fineran, P. C. A century of the phage: past, present and future. *Nature Reviews Microbiology*. **13**, 777–786 (2015).
392. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*. **41**, W349-57 (2013).
393. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices, edited by G. Von Heijne. *Journal of Molecular Biology*. **292**, 195–202 (1999).
394. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology*. **215**, 403–410 (1990).
395. Hood, I. V & Berger, J. M. Viral hijacking of a replicative helicase loader and its implications for helicase loading control and phage replication. *Elife*. **5**, (2016).
396. Via, A., Uyar, B., Brun, C. & Zanzoni, A. How pathogens use linear motifs to perturb host cell networks. *Trends in Biochemical Sciences*. **40**, 36–48 (2015).
397. Hyman, P. & Abedon, S. T. Bacteriophage host range and bacterial resistance. *Advances in Applied Microbiology*. **70**, 217–248 (2010).
398. Christo-Foroux, E., Vallaes, T., Loux, V., Dassa, E., Deutscher, J., Wandersman, C., Livernois, A., Hot, C., Criscuolo, A., Dauga, C., Clermont, D. & Chesneau, O. Manual and expert annotation of the nearly complete genome sequence of *Staphylococcus sciuri* strain ATCC 29059: a reference for the oxidase-positive *staphylococci* that supports the atypical phenotypic features of the species group. *Systematic and Applied Microbiology*. (2017).
399. Nyíri, K., Kohegyi, B., Micsonai, A., Kardos, J. & Vertessy, B. G. Evidence-based structural model of the staphylococcal repressor protein: Separation of functions into different domains. *PLoS One*. **10**, e0139086 (2015).

400. Deghorain, M., Bobay, L.-M., Smeesters, P. R., Bousbata, S., Vermeersch, M., Perez-Morga, D., Drèze, P.-A., Rocha, E. P. C., Touchon, M. & Van Melderen, L. Characterization of novel phages isolated in coagulase-negative *staphylococci* reveals evolutionary relationships with *Staphylococcus aureus* phages. *Journal of Bacteriology*. **194**, 5829–39 (2012).
401. Gutiérrez, D., Adriaenssens, E. M., Martínez, B., Rodríguez, A., Lavigne, R., Kropinski, A. M. & García, P. Three proposed new bacteriophage genera of staphylococcal phages: ‘3alikevirus’, ‘77likevirus’ and ‘Phietalikevirus’. *Archives of Virology*. **159**, 389–398 (2014).
402. Zhang, J., Zhang, Y. & Inouye, M. *Thermotoga maritima* MazG protein has both nucleoside triphosphate pyrophosphohydrolase and pyrophosphatase activities. *Journal of Biological Chemistry*. **278**, 21408–21414 (2003).
403. Zhang, J. & Inouye, M. MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with era, an essential GTPase in *Escherichia coli*. *Journal of Bacteriology*. **184**, 5323–5329 (2002).
404. Papp-Kádár, V., Szabó, J. E., Nyíri, K. & Vertessy, B. G. In Vitro analysis of predicted DNA-binding sites for the Stl repressor of the *Staphylococcus aureus* SaPIBov1 pathogenicity island. *PLoS One*. **11**, e0158793 (2016).
405. Kozłowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*. **13**, 111 (2012).
406. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. Protein disorder prediction: Implications for structural proteomics. *Structure*. **11**, 1453–1459 (2003).
407. Ishida, T. & Kinoshita, K. PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*. **35**, W460–W464 (2007).
408. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. RONN: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*. **21**, 3369–3376 (2005).
409. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P. & Dunker, A. K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. in *Proteins: Structure, Function and Genetics*. **61**, 176–182 (2005).
410. McGeehan, J. E., Streeter, S. D., Thresh, S. J., Ball, N., Ravelli, R. B. G. & Kneale, G. G. Structural analysis of the genetic switch that regulates the expression of restriction-modification genes. *Nucleic Acids Research*. **36**, 4778–4787 (2008).
411. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*. **3**, 711–721 (2005).
412. Winstel, V., Liang, C., Sanchez-Carballo, P., Steglich, M., Munar, M., Bröker, B. M., Penadés, J. R., Nübel, U., Holst, O., Dandekar, T., Peschel, A. & Xia, G. Wall teichoic acid structure governs horizontal gene transfer between major bacterial pathogens. *Nature Communications*. **4**, (2013).
413. Lindsay, J. A. Genomic variation and evolution of *Staphylococcus aureus*. *International Journal of Medical Microbiology*. **300**, 98–103 (2010).
414. Brown-Jaque, M., Calero-Cáceres, W. & Muniesa, M. Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid*. **79**, 1–7 (2015).
415. Bae, T., Baba, T., Hiramatsu, K. & Schneewind, O. Prophages of *Staphylococcus aureus* Newman and their contribution to virulence. *Molecular Microbiology*. **62**, 1035–47 (2006).
416. Hizi, A. & Herzig, E. dUTPase: the frequently overlooked enzyme encoded by many retroviruses. *Retrovirology*. **12**, 70 (2015).
417. Williams, M., Cox, B. & Ariza, M. Herpesviruses dUTPases: a new family of pathogen-associated molecular pattern (PAMP) proteins with implications for human disease. *Pathogens*. **6**, 2 (2016).
418. Glaser, R., Litsky, M. L., Padgett, D. A., Baiocchi, R. A., Yang, E. V., Chen, M., Yeh, P. E., Green-Church, K. B., Caligiuri, M. A. & Williams, M. V. EBV-encoded dUTPase induces immune dysregulation: Implications for the pathophysiology of EBV-associated disease. *Virology*. **346**, 205–218 (2006).

419. Ariza, M. E., Rivaller, P., Glaser, R., Chen, M. & Williams, M. V. Epstein-Barr virus encoded dUTPase containing exosomes modulate innate and adaptive immune responses in human dendritic cells and peripheral blood mononuclear cells. *PLoS One*. **8**, e69827 (2013).
420. Glaser, R., Litsky, M. L., Padgett, D. A., Baiocchi, R. A., Yang, E. V., Chen, M., Yeh, P. E., Green-Church, K. B., Caligiuri, M. A. & Williams, M. V. EBV-encoded dUTPase induces immune dysregulation: Implications for the pathophysiology of EBV-associated disease. *Virology*. **346**, 205–218 (2006).
421. Aubrecht, T. G., Weil, Z. M., Ariza, M. E., Williams, M., Reader, B. F., Glaser, R., Sheridan, J. F. & Nelson, R. J. Epstein-Barr virus (EBV)-encoded dUTPase and chronic restraint induce impaired learning and memory and sickness responses. *Physiology & Behavior*. **137**, 18–24 (2014).
422. Foerster, J., Nolte, I., Junge, J., Bruinenberg, M., Schweiger, S., Spaar, K., Van Der Steege, G., Ehlert, C., Mulder, M., Kalscheuer, V., Blumenthal-Barby, E., Winter, J., Seeman, P., Ständer, M., Sterry, W. & Te Meerman, G. Haplotype sharing analysis identifies a retroviral dUTPase as candidate susceptibility gene for psoriasis. *Journal of Investigative Dermatology*. **124**, 99–102 (2005).
423. Lai, O. Y., Chen, H., Michaud, H.-A., Hayashi, G., Kuebler, P. J., Hultman, G. K., Ariza, M.-E., Williams, M. V., Batista, M. D., Nixon, D. F., Foerster, J., Bowcock, A. M. & Liao, W. Protective effect of human endogenous retrovirus K dUTPase variants on psoriasis susceptibility. *Journal of Investigative Dermatology*. **132**, 1833–1840 (2012).
424. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology*. **2**, 28–36 (1994).

Publications related to this work

Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer (2017) **Bowring J***, Neamah MM*, Donderis J*, Mir-Sanchis I, Alite C, J Ciges-Tomas R, Maiques E, Mehmedov I, Marina A, Penadés JR. *eLIFE*. DOI: 10.7554/eLife.26487.

Convergent evolution involving dimeric and trimeric dUTPases in pathogenicity island mobilization (2017) Donderis J*, **Bowring J***, Maiques E*, Ciges-Tomas R, Alite C, Mehmedov I, Angeles Tormo-Mas M, Penadés JR, Marina A. *PLoS Pathogens*. DOI: 10.1371/journal.ppat.1006581.

Another look at the mechanism involving trimeric dUTPases in *Staphylococcus aureus* pathogenicity island induction involves novel players in the party (2016) Maiques E*, Quiles-Puchalt N*, Donderis J, Ciges-Tomas R, Alite C, **Bowring J**, Humphrey S, Penadés JR, Marina A. *Nucleic Acids Research*. DOI: 10.1093/nar/gkw317.

Bacteriophage Moonlighting Proteins in the Control of Bacterial Pathogenicity, in Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections (2017) **Bowring JZ**, Marina A, Penadés JR, Quiles-Puchalt N (ed B. Henderson), John Wiley & Sons, Inc., Hoboken, NJ, USA. DOI: 10.1002/9781118951149.ch22