2016

Reconstitution and Mechanistic Studies on the Staphylococcal agr Quorum Sensing Circuit

Boyuan Wang

RECONSTITUTION AND MECHANISTIC STUDIES ON THE STAPHYLOCOCCAL *agr* QUORUM SENSING CIRCUIT

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Boyuan Wang

June 2016

RECONSTITUTION AND MECHANISTIC STUDIES ON THE STAPHYLOCOCCAL *agr* QUORUM SENSING CIRCUIT

Boyuan Wang, Ph.D.

The Rockefeller University 2016

Quorum sensing (QS) plays a central role in virulence induction in the commensal pathogen, *Staphylococcus aureus*. This bacterium secretes an auto-inducer peptide (AIP), a small, cyclic peptide containing a thiolactone linkage as an indicator of its population density, and up-regulates virulence gene expression in response to high extracellular AIP levels. We have investigated two key biochemical events in *S. aureus* QS and revealed several underlying regulatory mechanisms. The first such event, formation of the high-energy thiolactone in AIP, is unusual in that it occurs directly through proteolysis of the precursor peptide, AgrD, without free-energy input from ATP hydrolysis. We showed that this proteolysis is, in line with the thermodynamic prediction, unfavorable and strongly reversible *in vitro*. As a consequence, rapid degradation of the concomitantly released C-terminal fragment of AgrD is required to power efficient AIP production *in vivo*. This observation provides a novel connection between protein homeostasis and QS in *S. aureus*. The second study focused on the AIP-sensing receptor histidine kinase (HK), AgrC, whose auto-phosphorylation exhibits several remarkable properties in our reconstitution system based on nanometer-scale lipid-bilayer discs (nanodiscs). Activation of this receptor by its native activator, for instance, requires a membrane environment enriched of anionic lipids mimicking the electrostatic property of the *S.*

aureus cell membrane. This strong dependence on lipid composition might explain why homologous QS systems exist only in low-GC Gram-positive bacteria, or *Firmicutes*, whose cell membranes are predominantly highly anionic. AgrC also binds to ATP at an exceptionally weak affinity, likely due to its distinct adenine-binding pocket conserved only in a small subfamily of HK receptors existing also exclusively in *Firmicutes*. The low affinity to the nucleotide cofactor likely enables AgrC to sense the energy condition of the bacterium and shut down the QS regardless of the population density when energy starvation drives down the cellular ATP level. Even more intriguing is the plasticity of AgrC auto-kinase activity when bound to different ligands. This behavior contrasts with the generally accepted two-state model of HKs. To understand the plasticity of AgrC, we systematically perturbed the conformation of the AgrC kinase domain using a fusion protein strategy. We demonstrated that the conformational state of a helical linker preceding the kinase domain exercises rheostat-like control over the kinase activity. Using full-length AgrC embedded in nanodiscs, we showed that binding of activator and inhibitor peptides results in twisting of the linker in different directions. These findings provide the first view on molecular motions triggered by ligand binding on a membrane-bound receptor HK. The smooth input-response landscape of the AgrC kinase domain also sheds new light on the mechanism of HK evolution through domain shuffling.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Tom W. Muir, for providing me the opportunity to work in his lab and, specifically, on the project that I am most interested in. The progresses I have made in my graduate school would not be possible without his warm encouragements, his incisive criticisms, his generous support, as well as his enormous patience and forgiveness.

I would like to thank the members of my Faculty Advisor Committee, Professor Seth A. Darst and Tom P. Sakmar for their constant support and insightful suggestions throughout my thesis work. I am grateful to Professor Richard P. Novick, our long-term collaborator at New York University for offering me access to his lab resources. I would also like to thank Professor Jon Clardy from Harvard Medical School for attending my thesis defense as the external examiner.

My thesis research was carried out in close collaboration with several teammates in the so-called "phosphohistidine-*agr*" subgroup of the Muir Lab. Peter M. Moyle and Jeffery G. Johnson have passed down the lab's reagent collection and intellectual legacy to me when I get started. Rob Oslund and Jung-Min Kee have collaboratively developed, and constantly supplied me with, a marvelous phosphohistidine-specific antibody that powers my study on the AgrC receptor histidine kinase. It is my great pleasure to work with Aishan Zhao, a talented and dedicated graduate student since my fourth year, and I am happy to pass down many unfinished businesses to this trustworthy colleague and friend. I would also like to thank Connie (Xinhui) Wang and Ashley Wang for spending their undergraduate research with me on an *agr*-related problem. Stephen (Qian) Xie, a postdoc who joined the team in 2014, also deserves recognition for his aspiration to tackle several long-standing structural problems.

The Muir Lab is a fantastic place to do research in that everyone is willing to share his/her knowledge and skills. I am grateful to Neel H. Shah and Manuel M. Müller for their constant interest in listening to my preliminary research ideas and providing constructive feedbacks. I have also consulted Miquel Vila-Perelló about intein technologies, Beat Fierz about fluorescent spectrophotometry, Disan S. Davis about membrane protein handling, Zack Z. Brown and Zhihua Liu about synthetic chemistry, Glen P. Liszczak about X-ray crystallography, Galia Debelouchina about NMR techniques and Yael David about general genetic questions.

I would like to thank the Dean's Office at The Rockefeller University and Sarah Porter and Megan Krause in the Department of Chemistry at Princeton University for their administrative assistance. I am also indebted to the Muir Lab mangers, Maya Bar-Dagan at Rockefeller and Lenka Bittova at Princeton, who have maintained all instruments and ordered all supplies necessary for my research.

As an international student, it is a long process to acquire necessary English skills and adapt to the lifestyle in United States. Special acknowledgement goes to Rob Oslund, Manuel M. Müller, Galia Debelouchina and Professor Tom Muir for proofreading my

written English and providing valuable suggestions. I would also like to thank Adam Stevens for familiarizing me with baseball and American football and Champak Chatterjee and Uyen T.T. Nguyen for sharing with me their opinions on American and/or European cultures.

I am fortunate to have two friends, Bonan Li and Zhixing Chen living close to me during most of my graduate school years. Interacting with them is full of fun and inspirations. I am indebted to my parents for their unconditional love and support and to my cousins for taking care of my parents during my years of absence. Finally, I would like to thank my wife, Shasha, for organizing a comfortable home and being the communication specialist of our family.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS v

LIST OF FIGURES ix

LIST OF TABLES xiii

Chapter One: Introduction 1

1.1 Quorum sensing in *Staphylococcus aureus* 1

1.1.1 Introduction to quorum sensing 1

1.1.2 Discovery of the *agr* locus 5

1.1.3 Basic architecture of the signaling circuit 8

1.1.4 Interspecies variation of *agr* 10

1.2 Biochemistry of the *agr* autoinduction 12

1.2.1 Translation of Agr proteins 13

1.2.2 Formation of the AIP thiolactone 15

1.2.3 Translocation and maturation of AIP 19

1.2.4 AIP-AgrC interaction 21

1.2.5 AgrA phosphorylation and transcription activation 27

1.3 Reagent development for the manipulation of *agr* 29

1.3.1 AIP scavenging in host defense 30

1.3.2 AgrC-targeting reagents 32

1.3.3 AgrA-targeting reagents 34

1.3.4 AIP-sequestration reagents 35

1.4 Summary: specific aims and significance 36

Chapter Two: Bioinformatic Analysis of the *S. aureus* AgrC **39**

2.1 Identification of close AgrC homologs from genomic databases 41

2.2 Sequence analysis of the HPK10 subfamily 43

2.3 Summary and discussions 46

Chapter Three: Reconstitution of Biochemical Events Related to AgrC **49**

3.1 Functional reconstitution of AgrC to nanodiscs 49

3.2 Recapitulation of AIP-dependent regulation of AgrC kinase activity 55

3.3 The AgrC-I-AgrA phospho-relay and AgrA dephosphorylation 58

3.4 Physicochemical parameters of the AIP-I-AgrC-I interaction 61

3.5 Nucleotide specificity and kinetics of AgrC-I auto-phosphorylation 64

3.6 Summary and discussions 68

Chapter Four: Mechanisms by which AgrC-I Responds to Signals **72**

4.1 Perturbing the conformation of the AgrC-I HK domain 73

4.2 Conformation-activity landscape of the AgrC-I HK domain and the molecular motion induced in full-length AgrC-I by AIP-I or AIP-II binding 77

4.3 A constitutive mutation, R238H, modifies the conformation-activity landscape 82

4.4 Reconstitution of the AIP-II-revertible constitutive mutants 88

4.5 Summary and discussions 90

4.5.1 Auto-kinase regulation of AgrC-I 90

4.5.2 Trans-kinase regulation of AgrC-I 95

4.5.3 Mechanisms of constitutive activity 96

Chapter Five: Key Driving Forces That Power AIP Thiolactone Formation **97**

5.1 Estimation of the free energy cost associated with thiolactone formation 98

5.2 Expression and purification of recombinant AgrB and AgrD constructs 102

5.3 AgrD(1-32)-thiolactone is the major product of the AgrB-catalyzed proteolysis of AgrD 105

5.4 AgrD processing by AgrB is a reversible process 107

5.5 AgrD(1-32)-thiolactone is stabilized by phospholipids 112

5.6 AgrD processing by AgrB is a reversible process *in vivo* 117

5.7 Summary and Discussion 119

5.7.1 Factors that facilitate AgrD proteolytic cyclization 119

5.7.2 Physiological implications of the ΔG^{0}' of thiolactone formation 120

Chapter Six: Outlook **124**

6.1 The signal-response property of the AgrC-I HK domain 124

6.1.1 The trans-kinase activity and the pHis stability in GCN4-AgrC-I chimeras 126

6.1.2 Introduction of scissoring motions to the AgrC-I interdomain linker helices 128

6.1.3 Is the interdomain linker pair an analog-to-digital converter? 130

6.2 Degradation of the AgrD C-terminal cleavage fragment 132

6.3 Detection of AIP thiolactones from other *Firmicutes* species 134

Chapter Seven: Methods **137**

7.1 Materials and Equipments 137

7.1.1 General Materials 137

7.1.2 General Equipments 139

7.2 Molecular cloning 140

7.2.1 cDNA sequence of Agr proteins 140

7.2.2 Expression constructs for AgrC and AgrA 140

7.2.3 Expression plasmids for AgrB and AgrD constructs 141

7.3 Recombinant protein production and reconstitution 143

7.3.1 Soluble proteins 143

7.3.2 Integral membrane proteins 144

7.3.3 AgrD constructs 145

7.3.4 Nanodisc reconstitution and SEC-MALS analysis 149

7.3.5 AgrB proteoliposome assembly 151

7.4 Characterization of the AgrC-AgrA TCS 152

7.4.1 Auto-kinase and trans-kinase assays 152

7.4.2 Kinetic data processing 158

7.4.3 Determination of the dissociation constants between AgrC-I and AIP 163

7.4.4 Other methods related to the AgrC-AgrA TCS 166

7.5 Characterization of the AgrB-AgrD peptidase-substrate pair 171

7.5.1: Biochemical assays based on AgrB-I proteoliposomes 171

7.5.2 Ring-opening of FlagHis$_6$-AgrD-I(1-32)-thiolactone and AIP-I 173

7.5.3 Quantification methods 174

7.5.4 Construction of *S. aureus* strains and western blotting analysis thereof 178

7.5.5 Other methods related to AgrB and AgrD 180

Appendix **182**

References **194**

LIST OF FIGURES

Chapter One

Figure 1.1 Common features of bacterial quorum sensing 2

Figure 1.2 The genetic arrangement of the *S. aureus agr* locus 6

Figure 1.3 Architecture of the agr autoinduction loop and effector regulation 8

Figure 1.4 Allelic variation of *agr* among four *S. aureus* subgroups 11

Figure 1.5 Architecture of RNAII and the possible implication on the translation of Agr proteins. 14

Figure 1.6 Formation of the AIP thiolactone 16

Figure 1.7 The AgrC-AIP interactions 22

Figure 1.8 The crystal structure of AgrA and the regulatory elements of P2 and P3 transcription. 26

Figure 1.9 Mechanisms of AIP scavenging by mammalian hosts. 31

Figure 1.10 Synthetic molecules or natural products that targets *agr* 33

Chapter Two

Figure 2.1 Primary functions of HK homology boxes captured in crystal structures 40

Figure 2.2 Flow chart of the database search for AgrC homologs 41

Figure 2.3 Alignment of the *E. coli* EnvZ protein to HPK10 subfamily members at the G1 box 43

Figure 2.4 Statistics of the expanded HPK10 subfamily. 44

Figure 2.5 Identification of a less conserved interdomain linker within AgrC 45

Figure 2.6 Comparison between the prototypic *com* and *agr* loci 47

Chapter Three

Figure 3.1 Reconstitution of AgrC-I to nanodiscs composed of DMPC and DMPG 50

Figure 3.2 Dependence of AgrC-I activation on lipid composition 52

Figure 3.3 Nanodisc reconstitution of AgrC-II and -III using POPC and POPG 54

Figure 3.4 Auto-kinase activity and AIP-dependent regulation of AgrC-I 56

Figure 3.5 Comparison of the auto-kinase activity and regulation among *S. aureus* subgroups-I, -II and –III 58

Figure 3.6 Phospho-relay and phosphatase assays of AgrC-I and AgrA 59

Figure 3.7 Ring-opening of the thiolactone is involved in AIP-AgrC interactions 61

Figure 3.8 The stoichiometry and dissociation constants of the interaction between AgrC-I and AIPs 63

Figure 3.9 Nucleotide specificity of AgrC-I 65

Figure 3.10 Kinetics of the forward and reverse auto-kinase reactions in AgrC-I 66

Chapter Four

Figure 4.1 An α-helical linker connecting the AgrC-I sensor and HK domains 74

Figure 4.2 Design principles of the GCN4-AgrC-I chimera series 76

Figure 4.3 Twisting of the interdomain linkers exercises rheostat-like control over the auto-kinase activity of the AgrC-I HK domain. 78

Figure 4.4. AIP binding induces rotational movement of the AgrC-I linker helices 80

Figure 4.5 Auto-kinase activity and ligand responses of AgrC-I^{R238H} 83

Figure 4.6 The R238H mutation renders the AgrC-I HK domain less responsive to the twisting movement of linker helices. 85

Figure 4.7 Auto-kinase activation by the R238H mutation requires the attachment of the AgrC-I construct to nanodiscs by a TM domain 87

Figure 4.8 Ligand responses of the AIP-II-revertible constitutive mutants 89

Figure 4.9 An alternative model of the linker conformation in three ligand states 93

Figure 4.10 New structural evidence on the molecular movement related to RHK activation 91

Chapter Five

Figure 5.1 The AIP biosynthesis pathway and the model for free-energy change estimation for the thiolactone formation 98

Figure 5.2 Preparation of recombinant AgrB and AgrD 103

Figure 5.3. Dimerization of AgrB under native conditions. 104

Figure 5.4 Lipid-composition dependence and subgroup specificity of the AgrB-catalyzed proteolysis of AgrD 106

Figure 5.5 AgrD(1-32)-thiolactone is the major product of the AgrB-catalyzed proteolysis of AgrD 107

Figure 5.6 AgrD processing by AgrB is a reversible process 109

Figure 5.7. Generation of AgrD(1-32) is dependent upon AgrB catalysis 111

Figure 5.8 Ring-opening equilibrium of AIP-I and FlagHis_6-AgrD-I(1-32)-thiolactones 113

Figure 5.9. Nanodisc co-migration assays for AIP-I and AgrD-I constructs 115

Figure 5.10 Hydrophobic side chains on the ring-forming residues are required for the liposome-dependent stabilization of the thiolactone 116

Figure 5.11. Reversibility of the proteolytic cyclization in vivo. 118

Figure 5.12 Summary of the driving forces that power the formation of the AgrD-(1-32)-thiolactone intermediate 121

Figure 5.13 Multiple sequence alignment of Staphylococcal AgrDs 123

Chapter Six

Figure 6.1 The three-component phospho-equilibration system 127

Figure 6.2 Strategies to introduce scissoring movements to the helical interdomain linkers 128

Figure 6.3 Putative conformational distributions of the two modules flanking the interdomain linker 131

Figure 6.4 Identification of $AgrD^{C}$ degradation proteases 133

Figure 6.5 Identification of thiolactone AIPs from complex systems 135

Chapter Seven

Figure 7.1 Preparation of recombinant AgrD constructs. 146

Figure 7.2 Correction of the systematic errors introduced by the Sep-Pak procedure or the ionization bias of the MS 176

Appendix

Appendix Figure 1. RP-HPLC-MS results of all synthetic and recombinantly prepared AgrD constructs described in Chapter 5 189

Appendix Figure 2. MS results of all AgrD products from the biochemical reactions described in Chapter 5. 191

LIST OF TABLES

Chapter Two

Table 2.1 Taxonomic coverage of the HPK10 subfamily 44

Chapter Three

Table 3.1 Kinetic parameters of AgrC-I forward and reverse auto-kinase activities 67

Chapter Five

Table 5.1. Ring-closure equilibrium constants of thiolactones 114

Appendix

Appendix Table 1. All RHK hits from the database search for HPK10 homologs in Chapter 2 182

Appendix Table 2. AgrD constructs prepared as starting materials or detected as products in Chapter 5 188

Chapter One: Introduction

1.1 Quorum Sensing in *Staphylococcus aureus*

1.1.1 Introduction to Quorum Sensing

The concept of quorum sensing (QS) emerged out of the effort to understand the observation that an individual *Vibrio Fischeri* cell, when present within a quorum of cells from the same species, produces more light than in isolation (Figure 1.1a) (1). A series of elegant genetic and biochemical studies led to the conclusion that this bacterium produces a diffusive molecule, termed the autoinducer (AI), as an indicator of its local population density (Figure 1.1a, small triangles) (2-4). Expression of light-production genes requires high extracellular concentration of AI, and therefore occurs only within large populations (Figure 1.1a, right panel) (5). This "autoinduction" behavior, along with the discovery of similar regulatory principles in a spectrum of prokaryotic species, initiated a major paradigm shift in microbiology (6, 7). With the ever-increasing evidence of intercellular communication within and across species, prokayortes are no longer treated as the sum of isolated, unicellular organism but instead a community with constant information exchange (8). The signaling events that enable an individual bacterial cell to communicate with its surrounding quorum via chemical messengers are therefore referred to as QS. Importantly, autoinduction and QS are not used interchangeably in this dissertation. While the outcome of QS in pure cultures is almost always autoinduction, QS can have very a different impact on participant strains/species in mixed cultures or natural habitats (*vide infra*).

At high population densities, i.e., when a microorganism seizes dominance of a habitat, it may undertake a variety of collective actions to bolster its population fitness. For instance, some bacteria develop extracellular structures such as biofilms and/or capsules for adhesion and protective purposes to increase their survival under stressed conditions,

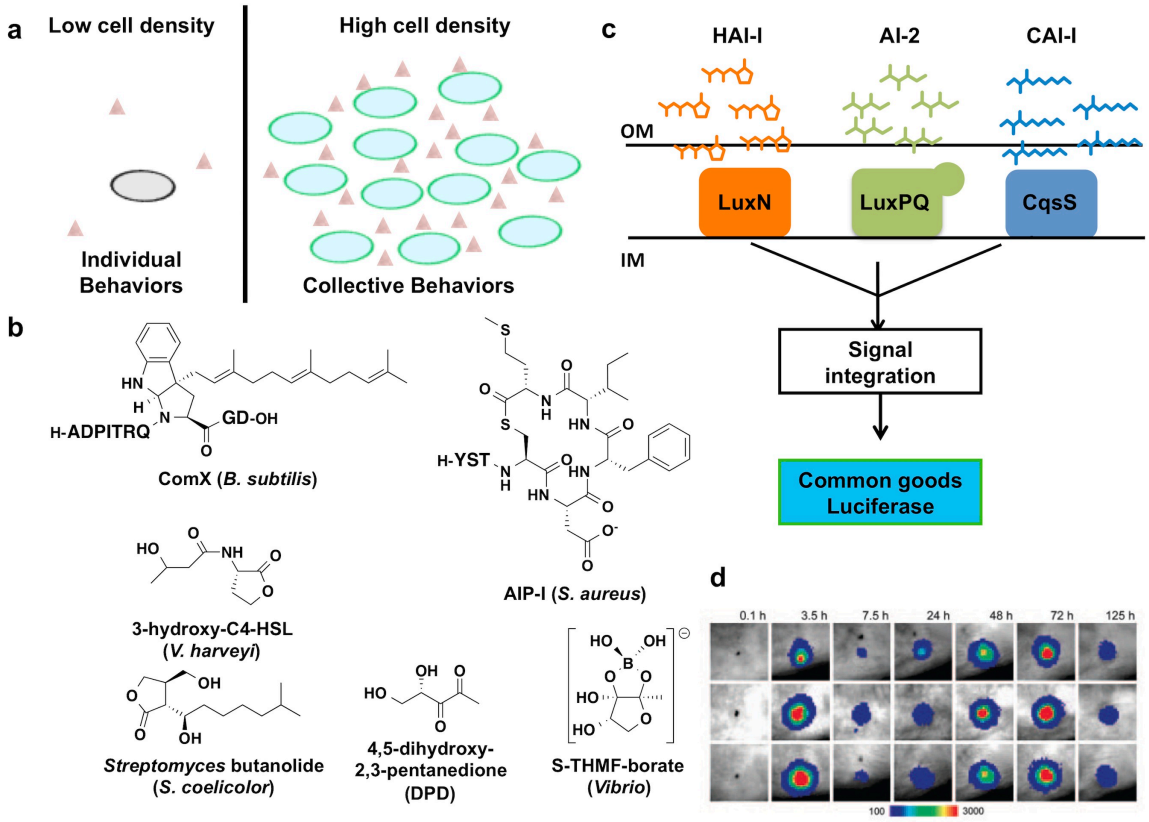


Figure 1.1 Common features of bacterial quorum sensing
(a) Population-dependent collective behaviors exemplified by the light production in *V. fischeri*. Diagram shows the situation at low or high cell densities. Cells and AI molecules are depicted by ovals and small triangles, respectively. Light production correlates to the AI level and occurs only at high population density. **(b)** Chemical diversity of bacterial AIs. See text for details. Note that DPD cyclizes into S-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran (S-THMF), which condenses with boric acid in marine bacteria to yield the bicyclic S-THMF-borate. **(c)** QS in *Vibrio harveyi* integrates signals from self, kin and distantly related species. See text for details. IM, inner membrane; OM, outer membrane. **(d)** Autoinduction of *S. aureus* experiences an eclipse phase in a murine infection model. An autoinduction-dependent luciferase reporter was implemented in the bacteria. Panels show the intensity of bioluminescence at the indicated time points in a pseudocolor scale. Note that the population size was roughly constant between the 3.5-h and 72-h time points. The panel is adapted from citation (29).

while others breakdown existing biofilm to let individual cells progress to the planktonic stage of their life cycles (9). A spectrum of "common goods" are also under control by autoinduction, including cytolytic toxins and hydrolases for further exploitation of nutrients from the environment, secondary metabolites for defensive or metabolic purposes, or factors that benefit the bacterium's mutualistic host (10-12). Besides these uniform behaviors, some bacteria also exhibit, at least partly in response to autoinduction, stochastic decision making to increase the phenotypic diversity and hence the robustness of the isogenic population. The most notable example exists in the community of *Bacillus subtilis*, in which high-density-related phenotypes including competence, motility and cannibalism are all controlled by stochasticity (13). Notably, however, for most collective behaviors listed above, premature actions at insufficient population would be energetically wasteful or counterproductive. As a result, the ability to sense the population density and make timing decisions for collective behaviors confers substantial evolutionary advantage to microorganisms, and is central to the phenotypic diversity of the prokaryotic world. Furthermore, QS-related behavior is also considered the driving force of speciation and the emergence of multicellularity (14-16).

Because of the enormous diversity of prokaryotes, it is essential for individual species to distinguish AIs produced by its diverse neighbors and respond (or not respond) in such a way that benefits its own fitness. To achieve this specificity, bacteria from different taxonomic lineages have implemented AIs of an equivalent level of diversity, as well as receptor proteins to sense them in a highly specific manner. Most AIs known to date fall in one of a handful of categories, each featuring a characteristic "core" structure

generated by a distinct set of biosynthetic genes. Best characterized AI categories include linear, prenylated and (thio)lactone peptides in *Firmicutes* (17-20), *a*cyl-*h*omoserine *l*actones (AHLs) , furanone derivatives and alkylquinolones in *Proteobacteria* (2, 21, 22), and butyrolactones in *Actinobacteria* (Figure 1.1b) (23). While the core structure provides rough distinction of AIs at the phylum level, a much greater source of diversity is conferred by the starting-material variation in the biosynthesis. The structure of the ribosomally-synthesized, post-translational modified AI peptides in *Firmicutes* is dictated by their precursor sequence and the proteolytic processing sites, while that of AHLs largely depends on the fatty-acyl group preference of their respective LuxI synthase. Importantly, in both cases, the cognate receptors that sense these AIs are so specific that minimal alteration of the AI structure at certain positions may alter its potency or efficacy (24, 25). It is the strict specificity of the receptors that allows the bacteria to avoid the interference of ever-existing environmental noises.

Despite the diversity of AI structures, inter-species crosstalk mediated by these chemical messengers is prevalent. *Vibrio harveyi*, a free-living marine bacterium, produces three distinct AIs and integrates their signals through a convergent phospho-relay signal transduction system (Figure 1.1c) (26). Intriguingly, while the HSL-type HAI-I is highly specific for detection by *V. harveyi*, the hydroxyketone signal CAI-I and the furanone-type AI-2 are also produced in the closely-related *Vibrio* kins and an even broader spectrum of bacteria, respectively. Converging three signal-transduction systems detecting these AIs onto a central regulator enables this bacterium to weigh differentially the signals produced by self, kin and distantly related species for its decision-making

process. By contrast to the cooperative crosstalk in the genus *Vibrio*, quorum quenching (QQ), defined as the antagonism of QS by foreign chemicals, is also frequently described as the means of interspecies competition (6, 27). The antagonizing reagent varies from analogs of the native AI to biosynthetically/structurally unrelated compounds. Overall, it is generally accepted that interspecies communication plays a central role in shaping the ecological niche and the evolutionary trajectory of microbial organisms (28).

Finally, it is noteworthy that the outcome of QS-related signaling is not solely dependent upon the population density or the extracellular AI concentration. Since QS induction often leads to a profound change in gene-expression pattern and hence an often more energetically expensive lifestyle, it is not surprising that checkpoints for cellular fitness are implemented upstream to the QS-mediated gene-expression regulation. In line with this notion, down-regulation of QS-induced behaviors after entrance of the stationary phase has been observed in a few staphylococcal and streptococcal species when cultured in liquid medium (29). In a murine *Staphylococcus aureus* infection model, activity of a QS-regulated reporter gene experienced biphasic induction and was eclipsed between the two induction phases due to the action of the host macrophages (Figure 1.1d) (30). Possible mechanisms of such checkpoints have been proposed, however, no consensus has been reached for any taxonomic phylum.

1.1.2 Discovery of the *agr* locus

Staphylococcus aureus is one of the low-GC gram-positive species in which autoinduction behaviors were earliest described. This commensal pathogen has been

known, since 1970s, to abruptly increase its extracellular protein secretion while decrease its surface protein production in the late-exponential phase when cultured in liquid medium (31). The chromosomal locus responsible for the global regulation of these proteins was first identified in 1982 using a transposon-insertion strategy (See Figure 1.2) (32). The mutant strain was found to possess an unaffected gene encoding alpha-cytotoxin but to be incapable of secreting this toxin. Further investigation of this mutant revealed defect in up-regulation of a few other toxins and down-regulation of surface protein-A, and the locus of transposon insertion was therefore named *agr* (for accessory gene regulator) (33). The open reading frame (ORF) disrupted by the insertion encodes a DNA-binding protein, AgrA, whose expression was once thought to be driven by a weak promoter, P1, located immediately upstream to the ORF (Figure 1.2) (34). In 1988, this ORF was also found in a larger, QS-inducible polycistronic mRNA, in which three additional ORFs upstream to AgrA were identified to encode AgrB, AgrD and AgrC proteins (Figure 1.2) (35). Meanwhile, the locus that encodes the delta-hemolysin (*hld*), an *agr*-regulated effector toxin, was mapped in close proximity to *agr*, and surprisingly, the mRNA transcript of the 26-amino acid (aa) toxin measures ~500-nucleotide (nt) in

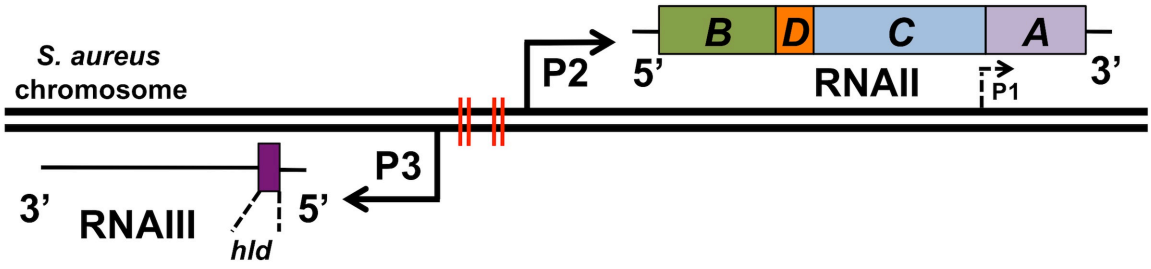


Figure 1.2 The genetic arrangement of the *S. aureus agr* locus.
The DNA double strand of *S. aureus* chromosome is shown in two parallel lines, with the location of promoters indicated with arrows. ORFs in RNAII and RNAIII are colored boxes (not drawn to scale). Red bars on the DNA segment between P2 and P3 promoters indicate the AgrA-binding DNA elements.

length (Figure 1.2) (36). Intriguingly, the pleiotropic phenotype observed from the *hld* transcript deletion was similar to that of *agr* deletion, suggesting that this RNA plays a role in the global regulation of accessory-gene production (37). Following numerical order, the QS-inducible promoters governing transcription of the polycistronic *agr* operon and the *hld* transcript were named P2 and P3 and the their transcripts termed RNAII and RNAIII, respectively. These two operons are transcribed in back-to-back directions and promoter elements are located merely 41-nt apart from each other. Both operons and the intergenic regulatory elements are collectively referred to as the *agr* locus, due to their essentiality to the autoinduction behavior in *S. aureus*.

Given the functions of Agr proteins predicted by bioinformatic analysis and the precedence of AIs in Gram-negative bacteria, the regulatory function of *agr* was expected to occur through the production and sensing of a diffusive chemical messenger (38). In 1995, six years after characterization of the P2 operon, the first description of this AI was achieved upon biochemical fractionation and HPLC-MS analysis (39). This, 8-aa small peptide was named the AIP (*a*uto*i*nducer *p*eptide). Edman analysis of this AIP revealed a sequence embedded in the AgrD ORF, while MS of the native peptide as well as treatment with iodoacetic acid and hydroxylamine suggested that the AIP contains a thiolactone macrocycle formed by the condensation between the carboxyl group at the C-terminus and the thiol group on the side chain of a internal cysteine residue (see Figure 1.1b) (20). In 1998 the Muir Group unambiguously confirmed this distinctive cyclic structure through chemical synthesis, thus completing the preliminary characterization of all *agr* gene products (25).

1.1.3 Basic architecture of the signaling circuit

It is particularly remarkable that the genetic interaction among the four Agr proteins and the RNAIII had been correctly unraveled, through a combination of elegant genetic experiments and bioinformatics analyses, even prior to the characterization of the AIP (Figure 1.3) (38). In this early model, AgrB and AgrD produce or activate the production of the AI, while the AI activates the two-component, phospho-relay signal transduction system involving AgrC and AgrA. The active form of AgrA up-regulates the transcription of both P2 and P3 operons, thus closing a positive feedback circuit (through the activation of P2) and increasing the production of RNAIII. Function of AgrD and AgrB were specified later on as the AIP precursor and the AgrD-processing peptidase, respectively, however the basic architecture of the signaling circuit is retained.

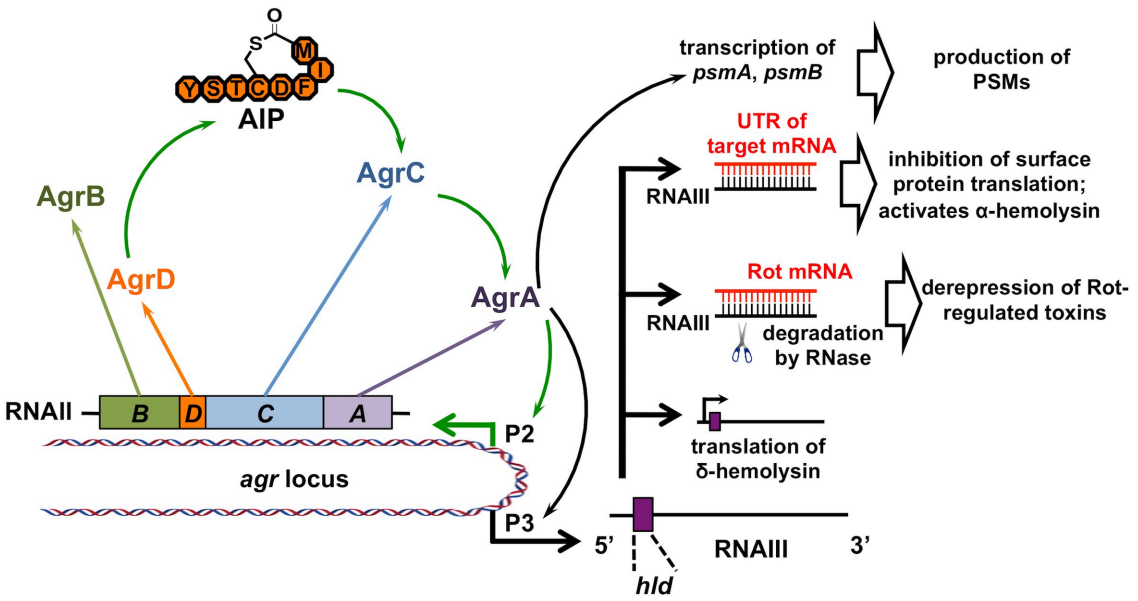


Figure 1.3 Architecture of the *agr* autoinduction loop and effector regulation.
Left half: a diagram of the *agr* autoinduction loop. Colored arrows indicate enhancement of activity or abundance through transcription, translation, enzymatic reactions and/or biophysical interactions. **Right half:** regulation of *agr* effector protein production. See text for details.

For over a decade since its discovery, RNAIII had been considered as the only direct effector of *agr* autoinduction. Ironically, its mechanism of action had been enigmatic for an even longer period until Huntzinger et al. showed that it inhibits the translation of *s*urface *p*rotein *A* (Spa) through interacting with, and mediating the degradation of, the *spa* mRNA (Figure 1.3) (40). Translation of several other *agr*-regulated proteins was subsequently found activated or inhibited by RNAIII employing similar mechanisms (41, 42). Among them is the *r*epressor *o*f *t*oxins (Rot), a transcription repressor targeting genes of a series of secreted toxins and proteases (Figure 1.3) (43). Hence, there are at least three known mechanisms by which RNAIII exercises control over *agr* effector protein production: (i) acting as the coding mRNA (ii) blocking translation and/or promoting degradation of the effector mRNA and (iii) indirectly influencing effector transcription by down-regulation of *rot*.

Within the past decade, nucleic-acid array analysis extended our scope of *agr*-autoinduction effectors beyond those under RNAIII regulation (44, 45). Employing a DNA-array strategy, Shu et al. identified, in addition to P2 and P3, three operons containing AgrA-binding elements in their promoter regions (Figure 1.3) (45). Two of these three encode *p*henol-*s*oluble *m*odulins (PSMs) whose leukocidal activities are essential for the bacterium's immune evasion. In contrast to the DNA-array approach for AgrA binding, hundreds of *S. aureus* genes were found influenced by *agr* autoinduction in a comparative transcriptomic study (44). While this study provides no information on the mechanism by which individual genes are affected, it assisted the identification of a few genes carrying *agr*-related functions. As a notable example, PmtABCD, the only

ABC transporter up-regulated upon induction of the *agr* system, was later shown to involve the exportation of PSMs (46).

1.1.4 Interspecies variation of *agr*

One of the most intriguing features of the *agr* locus is its polymorphism within single species. Within five years of the first *agr* locus being cloned, four allelic variants of this locus were reported (20, 47). The variable region spans half the length of the P2 operon, covering the main body of AgrB, the entire AgrD and the N-terminal half of AgrC ORFs. This setting allows each *agr* variant to specifically produce, and mediate autoinduction in response to, its own AIP (Figure 1.4a). In *S. aureus* strains carrying different *agr* variants, the vast majority of conserved, structural genes (excluding mobile genetic elements) are predominantly identical, suggesting that the variation occurs at a sub-species level. Strains harboring each *agr* allele are therefore classified as a pherotype or a specificity subgroup.

While *S. aureus* from all four groups are capable of qualitatively similar autoinduction when cultured alone, the effect of AIPs on the induction of a heterologous *agr* system is, in most cases, strongly inhibitory (Figure 1.4a) (48). The only exception lies between the two most closely related groups, I and IV—AIPs from these groups share 7 identical residues out of 8 positions. Despite this level of similarity, AIP-I activates the induction of group-IV strain three orders of magnitude less potently than AIP-IV. By contrast, the thiolactone-forming cysteine is the only identical residue shared between AIP-I and AIP-II. Clinical isolates of *S. aureus* from one infection site rarely exhibit variegation in the

agr locus, primarily because an *agr*-heterologous cell population cannot achieve cooperative autoinduction to support the fitness of all participant subgroups (49).

A long-standing puzzle of *agr* polymorphism concerns the evolutionary advantage offered by individual *agr* alleles. Indeed, a correlation has been observed between *agr* variants and infection types (49). For instance, group-III strains are overrepresented in menstrual toxic shock syndromes, while the exfoliatin-producing strains causing scalded

a

	AIP-I	AIP-II	AIP-III	AIP-IV
agr-I	activator	inhibitor	inhibitor	activator
agr-II	inhibitor	activator	inhibitor	inhibitor
agr-III	inhibitor	inhibitor	activator	inhibitor
agr-IV	activator (weak)	inhibitor	inhibitor	activator

b

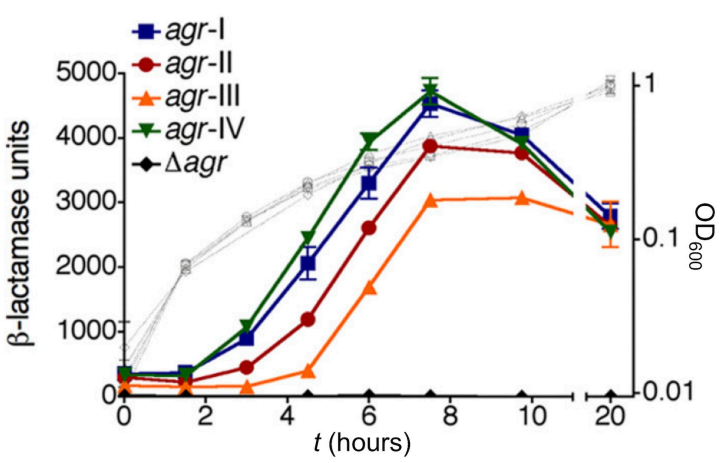


Figure 1.4 Allelic variation of *agr* among four *S. aureus* subgroups
(a) Structure and efficacy of AIPs from all four subgroups. **(b)** Effect of *agr* variants on the autoinduction timing when introduced to an isogenic, *agr*-null background. The background strain carrying the indicated *agr* allele and a P3-driven β-lactamase reporter was assayed for the β-lactamase activity in the medium supernatant at indicated time points. Plot shows a time courses of the β-lactamase activity (closed symbols) and a growth curve (open, gray symbols) for each *agr* allele tested. The figure was adapted from citation (48).

skin syndrome are predominantly group-IV. In a seminal study performed by Geisinger et al., all four *agr* alleles were introduced, one at a time, into an *agr*-null background strain through chromosomal insertion at an identical attachment site (50). Side-by-side comparison of these alleles on an isogenic background revealed major differences in the temporal control of autoinduction: induction was achieved earliest with group-I and group-IV and latest with group-III alleles. This observation argues for a model in which *agr* variants act as alarm clocks for the onset of autoinduction, and their correlation to infection types originated from the infection site's selection for the most favorable induction timing. The mechanism underlying the differential autoinduction timing and dynamics mediated by different *agr* variants, however, requires further investigation.

1.2 Biochemistry of the *agr* autoinduction

As was alluded to in section 1.1.3, central to *agr* autoinduction is a positive feedback circuit consisting of two mutually enhancing pathways, namely the production and sensing of AIP. AIP production begins from the translation of RNAII into four Agr proteins. AgrD, the precursor of AIP, is first proteolytically processed by AgrB, a membrane-integrated peptidase to generate a thiolactone intermediate. This intermediate is then exported across the membrane and is subject to a second proteolysis step to release the mature AIP pheromones into the extracellular space. This AIP-production pathway enhances the sensing pathway by contributing the activation signal. AIP sensing is achieved through its detection by the membrane-bound *r*eceptor *h*istidine *k*inase (RHK), AgrC. This interaction activates a phospho-relay cascade that leads to the phosphorylation of the cognate response regulator (RR), AgrA. Upon phosphorylation,

AgrA binds to the P2 promoter and up-regulates the transcription of RNAII, which encodes all four Agr proteins. This enhances the synthesis of AgrB and AgrD proteins and confers positive feedback to the AIP synthesis pathway.

This dissertation will primarily focus on the biochemical and biophysical characterization of protein components involved in the above feedback loop. Thus, an in-depth, stepwise review from biochemical perspectives will be provided in this section, covering works published before and during my matriculation at the Rockefeller University, with the exception of my own work, which will be presented in Chapters 2 through 5.

1.2.1 Translation of Agr proteins

Although the coding sequence of all four Agr proteins are co-transcribed in the polycistronic RNAII, this transcript is exquisitely designed to finely coordinate their translation. Discernible *r*ibosomal *b*inding *s*equences (RBS) are found immediately upstream to each ORF (Figure 1.5a) (38). The RBS preceding the coding sequence of AgrD and AgrA closely matches the consensus sequence, suggesting highly efficient translation initiation (51). Furthermore, the AgrB and AgrD ORFs are concatenated in such a way that the RBS of AgrD locates upstream to the AgrB stop codon. This design harnesses the translation activity within the AgrB ORF to prevent secondary structure formation that might otherwise occlude the translation initiation at the AgrD RBS, further increasing the production efficiency of AgrD (52). By contrast, the ORFs of AgrB and AgrC are expected to be less active due to (i) the less favorable RBS and (ii) the sub-optimal initiating codons, i.e., UUG for AgrB and GUG for AgrC. These features are

expected to result in the enzymatic substrates, AgrD and AgrA, being produced in excess to their respective enzymes, AgrB and AgrC. Moreover, two out of four proteins encoded by RNAII are integral membrane proteins. Since membrane insertion occurs co-translationally, the polysome complex of RNAII, which also translates AgrD, must localize in the proximity of the cell membrane (Figure 1.5b) (53). Because the 46-aa

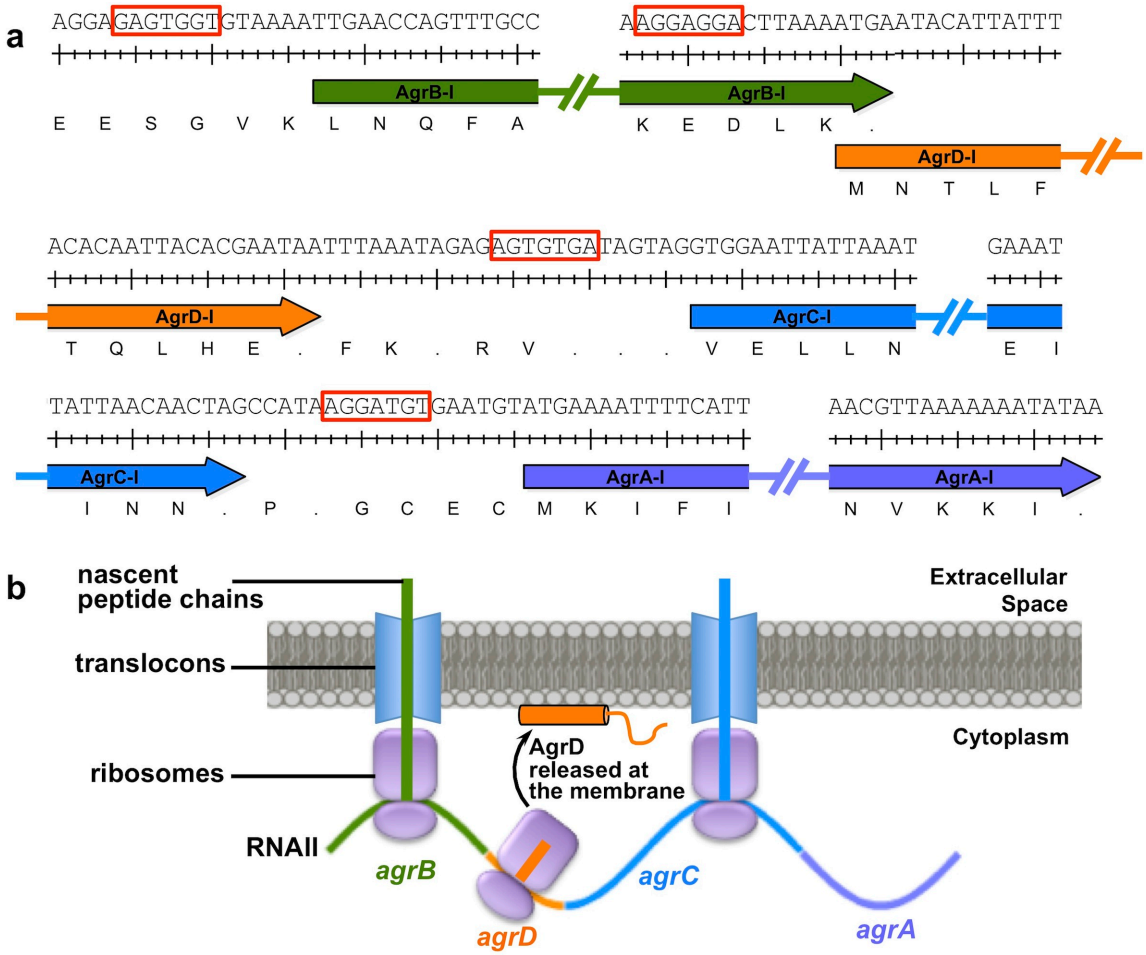


Figure 1.5 Architecture of RNAII and the possible implication on the translation of Agr proteins.
(a) The coding-strand DNA sequence of the polycistronic RNAII: four RBSs are indicated with red boxes. ORFs are indicated with arrows following the color code in Figure 1.2, accompanied with the translated sequence. Due to the space limit, only the coding region of the first and last five amino acids is shown for each ORF. **(b)** Diagram showing the putative membrane localization of the RNAII polysome as a consequence of membrane-protein translocation.

AgrD is highly hydrophobic and likely unstructured unless associated with the lipid bilayer (*vide infra,* section 1.2.2), it might be an important protective measure to synthesize AgrD in proximity to the cell membrane to prevent its aggregation or degradation.

1.2.2 Formation of the AIP thiolactone

One of the most fascinating features of the *agr* system is the thiolactone structure of the AIP. In reality, however, this thiolactone motif is generated by a single proteolysis reaction involving AgrB and AgrD (54, 55). The substrate of this proteolysis, AgrD, contains the mature AIP sequence sandwiched by an N-terminal leader peptide and a C-terminal recognition sequence (Figure 1.6a). The AgrB-catalyzed proteolysis clips off the recognition sequence as a linear peptide and concomitantly installs the thiolactone to the N-fragment, herein referred to as the thiolactone intermediate. The N-terminal 20 residues of the leader form an amphipathic helix and anchors AgrD to the cell membrane putatively by lateral association (56). This positioning is required for AgrD processing, as the substitution of its N-terminal region with an artificial amphipathic helical sequence, but not a hydrophobic trans-membrane domain, is tolerated. In all staphylococcal AgrDs, the amphipathic sequence is followed by an "IG" motif that might act as a helicity breaker to facilitate the proteolytic release of mature AIP from the thiolactone intermediate (57). The recognition sequence, on the other hand, is enriched in acidic residues and highly conserved in staphylococcal AgrDs. The hydrophilicity of this region has been thought to prevent full-length AgrD from premature exportation (Jeffery G. Johnson, unpublished work). The significance of the sequence conservation is not yet

fully understood despite some evidence showing a couple of positions are essential for competent AIP production (Figure 1.6a) (55, 58).

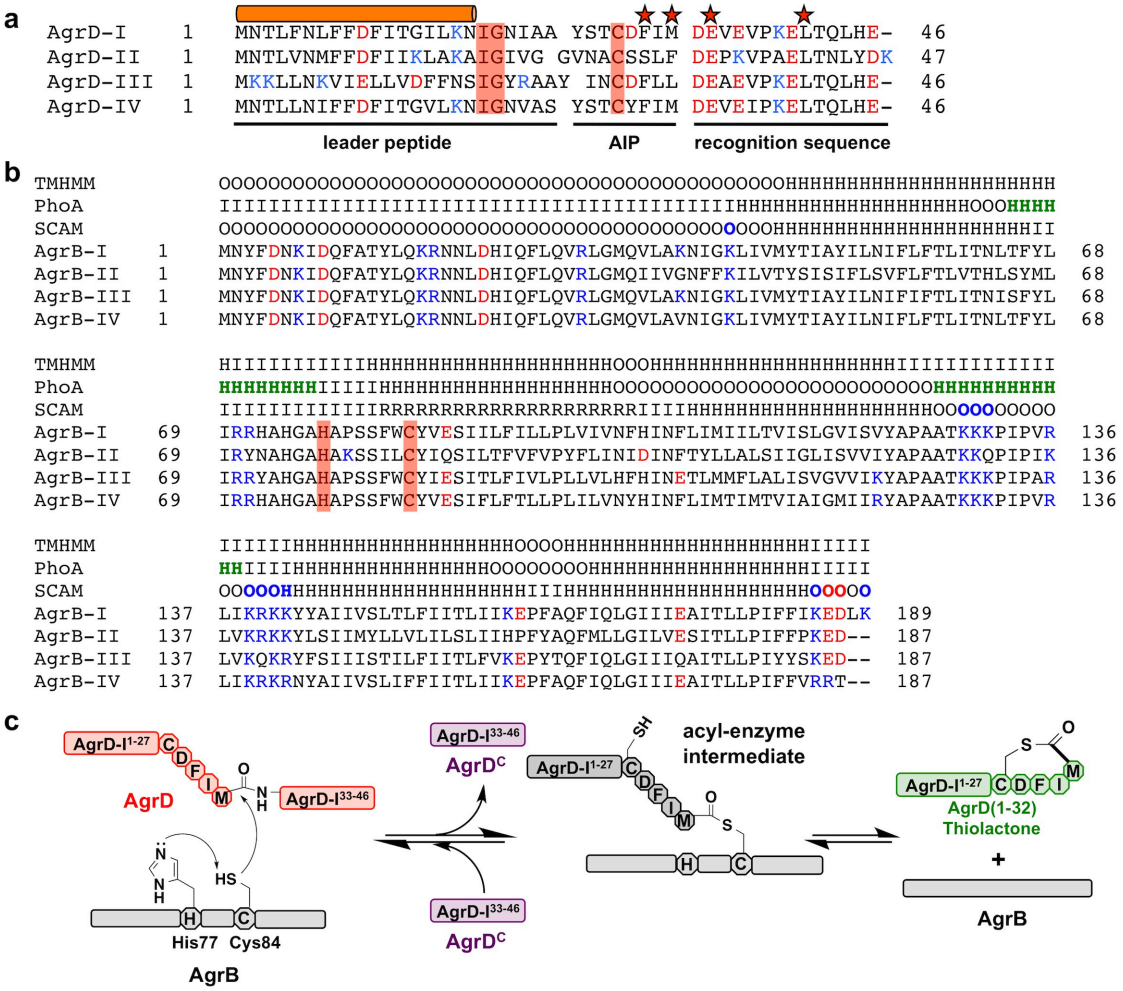


Figure 1.6 Formation of the AIP thiolactone

(a and **b)** Sequence alignment of the four **(a)** AgrD and **(b)** AgrB variants from *S. aureus*: acidic and basic residues are highlighted in red and blue, respectively. Residues playing explicit roles in AgrD processing are highlighted with red boxes. In **(a)**, the orange cylinder and red stars indicate the amphipathic α-helical sequence and the positions on which alanine mutagenesis abolishes AIP production. In **(b)**, the topology of AgrB predicted *in silico* by TMHMM 2.0 (TMHMM) or determined employing PhoA-fusion (PhoA) or SCAM (SCAM) strategy is displayed above the alignment. I, O, H and R denote an aligned position to reside intracellularly, extracellularly, within a TM helix or within a re-entry loop, respectively. **(c)** The proposed mechanism of the thiolactone formation exemplified by the cleavage of group-I AgrD.

AgrB, the peptidase catalyzing the proteolysis of AgrD, is a multi-pass membrane protein. The topology of AgrB has been experimentally determined twice, each time yielding a model different from the computationally predicted one (Figure 1.6b) (58, 59). Interestingly, while the experimental studies might be expected to be more credible, the topology they determined does not conform to the generally accepted features of α-helical membrane proteins. The six-transmembrane (TM) model generated from a phosphatase-fusion strategy, for instance, contained two 12-aa TM domains that are obviously too short to span a typical bacterial cell membrane when adopting the α-helical conformation (60). The substituted cysteine accessibility method (SCAM), on the other hand, gave rise to a four-TM model in which most basic residues flanking the TM domains localizes to the extracellular side. This configuration contradicts the "positive-inside" rule, which states that membrane-proximal basic residues prefer intracellular localization due to the direction of the physiological electrochemical potential across the cell membrane (61). Due to these controversies, cartoon presentation of AgrB in this dissertation will avoid elaborating on its topology. Ultimately, this controversy may not be addressed without a high-resolution structure of this protein.

The AgrB-catalyzed proteolysis of AgrD has been studied extensively employing both genetic and biochemical approaches. Two residues invariable in all known AgrB homologs, one cysteine and the other histidine, have been identified as the catalytic diad (Figure 1.6b) (54). Interestingly, all three topology models agreed with each other regarding to the intracellular localization of this catalytic diad, which ensures its accessibility to the AgrD substrate. An elegant peptidyl-transfer mechanism has been

proposed for this reaction (55), in which AgrB attacks the scissile bond by its active-site cysteine to form an acyl-enzyme thioester intermediate with the concomintant release of the linear AgrD C-fragment (Figure 1.6c). The intermediate is then resolved by transferring the peptidyl group of the AgrD N-fragment to the side-chain thiol group of its internal cysteine residue (C28). This releases the N-fragment as a thiolactone and recycles the free enzyme. While this mechanism is attractive, direct characterization of the thiolactone motif in the AgrD N-fragment has never been achieved. Despite this ambiguity, two lines of evidence strongly support that AgrB is sufficient for the production of the thiolactone intermediate (55). Firstly, co-expression of AgrB with the AgrD C28A mutant, rather than the wild-type AgrD, resulted in the detection of the acyl-enzyme intermediate on non-reducing SDS-PAGE, suggesting that this cysteine is indeed the acceptor of peptidyl transfer. Secondly, AIP-related activity was found in the growth medium of *Escherichia coli* co-expressing AgrB and AgrD. This observation also argues for the role of AgrB in the cyclization in that the thiolactone macrocycle is required for the autoinduction activity of AIP, while an AgrD cyclase is not likely to exist in the distantly related in *E. coli*.

Besides the catalytic diad, other regions and positions required for this activity have been revealed in both AgrB and AgrD employing mutagenesis strategies (58). Important findings include: (i) the cleavage site on AgrD is defined by residues on the -3, -1 and +2 positions, which are likely involved in enzyme-substrate recognition (Figure 1.6a); (ii) truncating the leader peptide from the N-terminus or the recognition sequence from the C-terminus impedes proteolysis and (iii) the conserved, soluble N-terminus of AgrB (aa

1-42) is required for the proteolytic activity. With respect to group specificity, heterologous AgrB-AgrD pairs involving AgrB-II or AgrD-II are inactive (62). This is not surprising because at the -3 position relative to the cleavage site, AgrD-II has a serine while the rest three have a phenylalanine. Determinants of group specificity in AgrB-II have been determined through sequence shuffling (62). Collectively, these efforts shed light on the AgrB-AgrD recognition despite the absence of a high-resolution structural model.

1.2.3 Translocation and maturation of AIP

After being processed by AgrB, the thiolactone intermediate awaits another proteolysis event to release the freely diffusive AIP pheromone from the membrane-anchoring N-terminal leader peptide. Moreover, an active or facilitated translocation event is also required for the successful secretion of AIP due to its presumed lack of membrane permeability. Intriguingly, the *agr* locus does not encode designated proteins to account for these steps. While the AIP-sensing function of AgrC and AgrA could be confidently predicted from sequence analysis, AgrB, a peptidase from a less studied protein family, had been surmised to also export and/or cleave the thiolactone intermediate (59). While the second cleavage activity turned out to be absent as was shown in biochemical studies (54), the possibility of its involvement in AgrD translocation has not been formally ruled out.

In principle, the translocation could occur either before or after the second proteolysis, albeit the former scenario is favored by a few lines of evidence. In 2007, Kavanaugh and

coworkers showed that the general signal peptidase, SpsB, cleaves a heptapeptide mimicking the leader peptide-AIP junction of AgrD-I at the expected releasing site (57). Because the catalytic domain of SpsB localizes in the extracelluar space, the authors argued in favor of the translocation-first model. It should be noted that this study provided no solid evidence on the cleavage of the native thiolactone intermediate by SpsB, nor did it test the activity of SpsB on the releasing-site sequence from other *S. aureus* subgroups. More recently, large quantities of the AgrD leader peptide have been detected in the *S. aureus* extracellular matrix (63, 64). This result also corroborates the translocation-first model, as the alternative, proteolysis-first model entails separate translocation of both leader and AIP fragments, which would be substantially less economical. The transporter responsible for AIP production, regardless of the substrate being exported, is yet to be identified.

The second proteolysis of AgrD is fascinating in that it affords the tail-length variation among AIPs. Because the C-terminal carboxylate of AIP participates the thiolactone macrocycle, residues N-terminal to the conserved, thiolactone-forming cysteine are collected referred to as the exocyclic tail. The tail length is 2 residues for AIP-III, 3 for AIP-I and IV, and 4 for AIP-II (See Figure 1.4). The sequence contexts of the scissile bonds are so diverged among groups-I, -II and –III that a single protease is not likely to cleave all three with appropriate efficiency and specificity. In line with this notion, SpsB failed to cleave group-II thiolactone intermediate under all conditions tested (Jeffery G. Johnson, unpublished data).

Although the exact identity of the transporter and AIP-releasing protease are unknown, educated predictions can be made on the properties of these proteins. Due to the autoinductory nature and the non-essential role of the *agr* system, they are expected to be either housekeeping proteins (e.g., SpsB) highly efficient in their respective *agr*-related function or *agr*-inducible proteins that carry out specific functions.

1.2.4 AIP-AgrC interaction

In *S. aureus*, AgrC is the only known receptor of the AIP. This 430-aa protein adopts the modular architecture commonly seen for RHKs, consisting of an N-terminal, membrane-integrated sensor module that detects AIP and a C-terminal histidine kinase (HK) module that carries out enzymatic functions (65). Topology analyses of the AgrC sensor module have generated less controversy compared to that of AgrB (65-67). All experimental and computational models agree on the intracellular localization of the HK domain as well as the positioning of the last four TM domains between residue 78 and 204, while the major debate lies on the number of TM domains within the N-terminal 77 residues. An earlier phosphatase-fusion strategy and a set of topology-prediction programs argued for two TM domains in this region (65), whereas a recent SCAM analysis and a different set of *in silico* programs suggested three (66). It is worth mentioning that both experimental studies were performed in *E. coli*, whose inner membrane may not support the exact fold of AgrC in *S. aureus* due to the drastically different lipid composition (68). In this dissertation, AgrC will be drawn following the former, 6-TM model that the Muir Lab has accepted for a long period of time (Figure 1.7a). However, readers should bear in mind that AgrC might adopt a 7-TM topology.

The sensor domain is connected to the HK domain via a short interdomain linker. As in most class-I RHKs, the AgrC HK domain contains two subdomains (Figure 1.7a) (69). The subdomain proximal to the sensor is the *d*imerization and *h*istidine *p*hosphorylation (DHp) subdomain, which folds into an α-helical hairpin and dimerizes through the

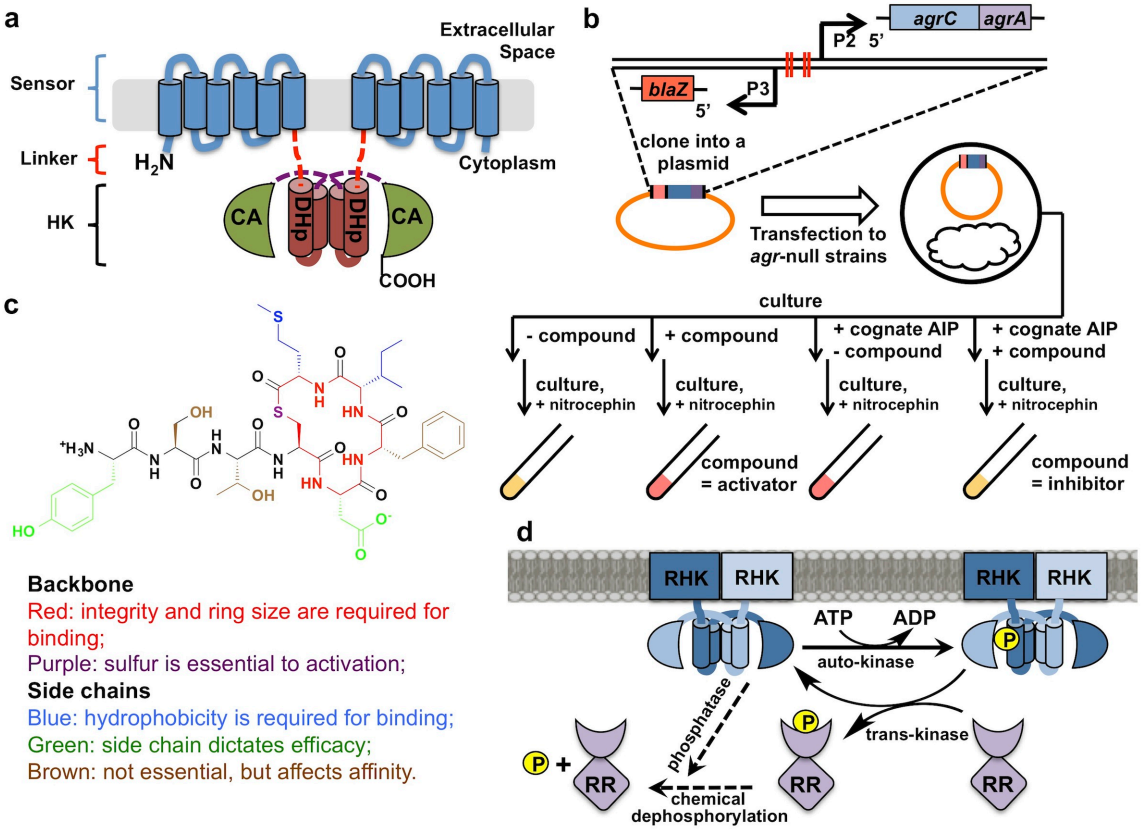


Figure 1.7 The AgrC-AIP interactions.
(a) Domain architecture of AgrC-I. The protein is shown as a homodimer with the integral membrane sensor domain and cytoplasmic DHp and CA subdomains colored in blue, brown and green, respectively. The interdomain linker and the DHp-CA linker region are depicted as red and purple dashed lines. **(b)** The design of reporter-strain assays: diagram shows the strategy to test the activating or inhibitory effect of a certain compound. The reporter protein is a β-lactamase, and nitrocephin is a β-lactam that displays a red color upon its hydrolysis. **(c)** A summary of the consensus conclusions drawn from the SAR studies of *S. aureus* AIPs. **(d)** Diagram showing the two sets of counteracting activities of a typical class-I RHK, with the kinase and phosphatase activities shown in solid and dash arrows, respectively.

formation of a four-helix bundle. Consequently, AgrC, as do most RHKs known to date, forms obligatory dimers. The distal, C-terminal subdomain is the so called *c*atalytic and *A*TP-binding (CA) subdomain, which adopts the common fold of the GHKL (DNA *g*yrase, *H*sp90, histidine *k*inase and Mut*L*) superfamily (70). The CA subdomain binds to the triphosphate nucleotide and catalyzes the phosphorylation at a conserved histidine residue on the DHp subdomain. Auto-phosphorylation of the AgrC homodimer occurs *in trans*, i.e., the CA subdomain of one subunit catalyzes the phosphorylation of the DHp subdomain of the other subunit (Figure 1.7a). This was concluded from the observation that the *agr*$^{-}$ phenotype conferred by mutating the phospho-accepting histidine in *agrC* could be complemented by another copy of *agrC* gene encoding a RHK deficient of ATP-binding (71). The crystal structure of a nucleotide-free AgrC CA domain has been available since 2014 (72).

The AIP binding site has been mapped using domain-swapping approaches. The variable region of the P2 operon includes the coding sequence of the sensor but not HK domain. Swapping the sensor and HK domains among subgroups demonstrated that the AgrC sensor dictates the AIP-sensing specificity (73). Further swapping of the distal (C-terminal 4 TM domains) and the proximal (N-terminal 2 or 3 TM domains) sequences leads to the conclusion that AIP binding, regardless of the efficacy, relies on the distal part of the AgrC sensor (74). Taking advantage of the strong similarity between AgrC-I and AgrC-IV, engineering also revealed that exchanging the second extracellular loop (ECL2) completely switches their group specificity. Since AIP sequence of group-I and –IV differs only at position 4, this residue likely interacts with the ECL2 on AgrC (75).

To pinpoint the structural elements of AIPs required for AgrC binding and/or activation, extensive structure-activity relationship (SAR) studies have been performed. Most of these studies rely on *agr* reporter strains for readout. These *agr*-null strains harbor a plasmid-born *agr* locus devoid of *agrBD*, with the P3 operon fused to a reporter gene, in most cases the β-lactamase (*blaZ*) or green fluorescent protein (*gfp*) gene (Figure 1.7b) (76, 77). Such designs render these strains deficient in AIP production but capable of transducing extracellular AIP signals following the specificity dictated by the *agrAC*. The activity of the reporter protein directly reflects the signaling outcome, i.e., the transcription activity of P3. SAR studies on *S. aureus* AIPs published to date have reached consensus on several important conclusions. First of all, the 5-aa, 16-membered ring structure is of utmost importance for binding (Figure 1.7c) (25, 78). To date, no linear peptide has ever shown any activity on a native AgrC. Alternating the ring size by more than one atom has also never been tolerated, although recently the Muir Lab showed that mutating the Asp5 in AIP-I to homo-β-Asp retains activation despite a loss of potency (79). Notably, changing the thiolactone linkage to a lactam or oxolactone, a modification that maintains the ring size, also abolishes the activation while retains the cross-group inhibition activities (Figure 1.7c) (25, 78). Other modifications to the macrocycle backbone, including methylation of amide bonds and substitution of single residue with the corresponding peptoid mimic, showed heterogeneous effects depending on the subject AIP and the modification site (80-82). Secondly, two positions at the C-terminal end of all known AIPs (not limited to *S. aureus*) are conserved for hydrophobic amino acids (74). Their hydrophobicity is required for the tight binding to AgrC, as

alanine mutation at either position abolishes both cognate-group activation and cross-group inhibition (Figure 1.7c) (25, 78, 83). Last, but not least, the cognate-activation activity of AIP is highly sensitive to structural modification. Modification of the exocyclic tail or mutation of the second ring-participating residue has converted native AIPs into cognate-group inhibitors (Figure 1.7c). Prominent examples include the appendage of a tyrosine residue to the N-terminus of AIP-III, as well as the D5A mutation in AIP-I and D4A in AIP-III (48, 78, 83). In all, these SAR studies provided precious insights on the design principles of AIPs, which have guided the design of many useful reagents including AIPs tagged with affinity epitopes or photocrosslinkers (Aishan Zhao, unpublished work), as well as QS inhibitors of all *S. aureus* subgroups (48, 77).

By sequence homology, AgrC and AgrA forms a *t*wo-*c*omponent signaling *s*ystem (TCS). Such systems exist in all three domains of life and are particularly prevalent in bacteria (84). In a TCS, the RHK component usually processes two counteracting sets of activities (Figure 1.7d): the auto- and trans-kinase activity that installs phosphorylation to its cognate response regulator (RR); and the phosphatase activity that removes the phosphorylation (69). In response to its cognate signal, RHK adjusts the balance between its two sets of activities and thereby modulates the RR phosphorylation level. Signal transduction employing TCSs is common among QS systems, however the effect of AI binding on RHK's kinase-phosphatase balance diverges (26). Importantly, using crude membrane vesicles from *E. coli* over-expressing AgrC, Lina et al. have shown that the presence of the cognate AIP activates the auto-kinase (65). Furthermore, Geisinger et al. showed that the cognate AIP binding at one sensor domain within a dimer affords

qualitatively similar activation to the phosphorylation of both subunits (71). Despite these efforts, a detailed understanding of the biochemistry and the signaling mechanism of the AgrC-AgrA TCS would entail high-resolution structural information as well as quantitative biochemical/biophysical characterization.

1.2.5 AgrA phosphorylation and transcription activation

Acting as the phospho-receiver in the TCS as well as a transcription activator, AgrA consists of two domains, each assuming one of its two functions. The N-terminal *rec*eiver domain is shared across all RR proteins and dimerizes upon phosphorylation at its conserved Asp residue (69). The C-terminal DNA binding domain belongs to the LytTR protein family and binds to the consensus DNA elements located in the P2 and P3 promoter region (85). The crystal structure of the AgrA DNA binding domain in complex with a cognate 16-nt DNA fragment has been solved (Figure 1.8a) (86). In stark contrast to most known transcription factors that bind DNA using α–helices, AgrA LytTR domain is enriched in β-strands and harnesses residues on its inter-strand loops for the interaction with DNA. This unique binding pattern causes bending of the DNA double-helix by 38^{o} upon the interaction (Figure 1.8a). Amino acid residues required for the stability, DNA-binding and/or transcriptional activation have been identified employing structure-guided or alanine-scan mutagenesis approaches (86, 87). Intriguingly, an AgrA mutant, $AgrA^{Y229A}$, is deficient in activating transcription from the P3 promoter even through it binds and bends the DNA element indistinguishably from the wild type. This finding suggested that the surface patch surrounding Y229 might be specialized for transcription regulation through an unknown mechanism.

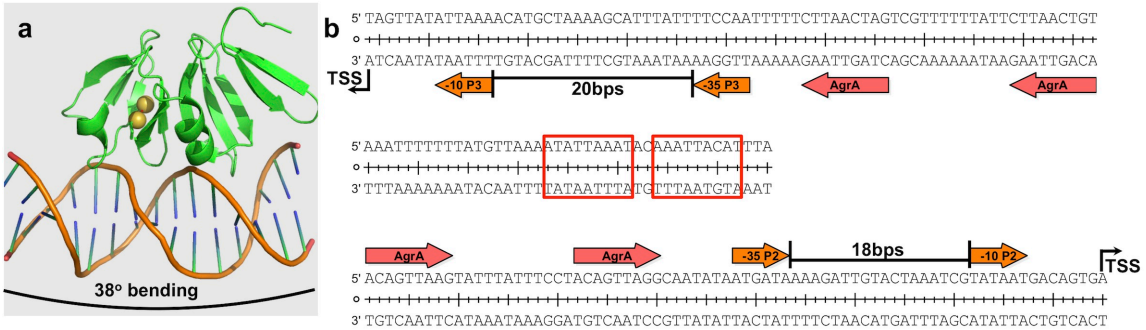


Figure 1.8 The crystal structure of AgrA and the regulatory elements of P2 and P3 transcription.

(a) Ribbon diagram of AgrA in complex with a 16-bp double-strand DNA containing its cognate binding sequence. AgrA is shown in green, with the sulfur atoms in Cys199 and Cys228 capable of disulfide formation under oxidative stress highlighted with yellow spheres. A 38^{o} arc next to the DNA strands illustrates the magnitude of DNA bending as a consequence of AgrA binding. **(b)** Double-strand DNA sequence of the intergenic region within the *agr* locus. The -10 and -35 elements with respect to the transcription-starting site (TSS) are marked with orange arrows and the AgrA-binding elements with red arrows. Red boxes denote the binding elements for SarA.

Autoinduction of *agr* presupposes a different temporal pattern of P2 and P3 transcription. Prior to autoinduction, a tight repression of the P3 operon is preferred to avoid unnecessary RNAIII-mediated mRNA degradation, while a reasonable expression level of P2 genes is required to set the stage of the autoinduction circuit. In agreement with this idea, experiments have confirmed that the baseline transcription level of P2 is higher, while its activation is less dramatic, compared to that of P3 (88). How could the same pool of AgrA possibly exercise differential regulation over two operons? The answer lies within their promoter sequences. Both P2 and P3 promoter regions contain, in the orientation of transcription, two AgrA-binding elements followed by the -35 and -10 boxes required for RNA polymerase (RNAP) recognition (Figure 1.8b). In spite of the similar architecture, AgrA-recognition sequences in P2 provide stronger affinity to AgrA than that in P3 (89). Thus, the promoter occupancy of P2 would be higher than that of P3 given limited availability of phosphorylated AgrA. Moreover, the spacer between the -35

and -10 boxes measures 18 nts in the P2 promoter and 20 nts in P3, both deviating from the optimal, 17-nt length (Figure 1.8b) (88, 90). Strikingly, shortening the spacer of P3 to the optimal length dramatically enhances the baseline transcription activity both *in vitro* and *in vivo* (88, 91). In light of this observation, the DNA-bending effect of AgrA binding as well as the dimerization induced by AgrA phosphorylation, has been surmised to rearrange the -35 and -10 boxes back to the optimal conformation to facilitate the binding of RNAP. Having been supported by *in vitro* transcription data, this theory provides a solid explanation for the more substantial up-regulation of P3-gene production during autoinduction.

The *agr* locus is a hot spot of signal integration due to its central role in direct regulation of dozens of effector genes. AgrA, for instance, can be inactivated by oxidative stress through the formation of a disulfide bond between two cysteine residues at positions 199 and 228 (Figure 1.8a) (92). Apart from AgrA, at least two transcription factors, SarA and SarR, are known to bind directly within the intergenic sequence between P2 and P3 (Figure 1.8b) (93, 94). Production of both SarA and SarR is growth-phase dependent, with their protein levels peaking during late-exponential and stationary phases, respectively (95, 96). While SarA enhances AgrA-dependent P2 transcription, SarR exhibits a repressive effect. Thus, their coordinated expression is thought to play a role in the down-regulation of *agr* activity in the stationary phase. In contrast to these direct *agr*-binding factors, more factors/pathways are known to regulate *agr* effectors in an *agr*-independent manner. At least two TCSs, SaeSR and SrrAB, control the virulence gene expression at the transcriptional level in response to stress conditions (97). Notably, a

global metabolic regulator, CodY, has recently been found to influence *agr* gene expression (98). The *Bacillus* homolog of this protein is known to respond to the secondary messenger, guanosine tetraphosphate (ppGpp) (99). In *S. aureus*, however, the *codY* deletion causes substantial over-expression of both P2 and P3 genes and hence a premature-autoinduction phenotype (100). Unexpectedly, CodY has barely detectable affinity to *agr* promoters, and the connection between CodY and *agr* gene products is still unclear.

1.3 Reagent development for the manipulation of *agr*

S. aureus requires *agr* not for survival but virulence. Interfering with the autoinduction, or QQ, should therefore be effective in combating the infection caused by this bacterium. Meanwhile, the likelihood of resistance development against *agr*-targeting reagents is expected to be lower since they impose milder selection pressure versus classic, bactericidal or bacteristatic antibiotics (101). These potential properties qualify *agr*-targeting reagents as attractive candidates for the treatment of staphylococcal infection. The cross-group inhibitory property of AIPs provides the first view of the effectiveness of QQ (25). AIP inhibitors have been shown to attenuate the spread of *S. aureus* infections in mice, phenocopying the use of agr-null *S. aureus* strains as infective agents (25). Remarkably, administering the antagonistic AIP during the beginning 3 hours of infection is sufficient for its attenuation effect despite the short lifetime of this peptide *in vivo* (30).

Although effective in containing new colonization by *S. aureus*, the effect of QQ reagents on existing infections has not been reported. It is noteworthy that silencing of the *agr* system is known to strengthen *S. aureus* biofilm and potentially benefits the maintenance of chronic infection (102). Conversely, activation of the *agr* system might represent a more attractive strategy in tackling *S. aureus* infection—not only does it disperse the biofilm, the constant induction of the *agr*-regulated genes also takes tolls on the fitness of the bacterium, rendering them more susceptible to classic antibiotics (103). Practical application of this strategy would entail a global activator that simultaneously activates the autoinduction of all four *S. aureus* subgroups. This reagent could potentially be a cocktail of a series of "clean" activators that each activates one or more *agr* variants without substantially affecting the rest. Unfortunately, the outcome of *agr* activators in animal models has not been tested to date, nor has any clean *agr* activator ever been developed. Because of the strong bias against *agr* activators in this field, the rest of this section will be dedicated to the discovery of QQ reagents.

1.3.1 AIP scavenging in host defense

Even in the absence of exogenous drugs, the mammalian host is able to scavenge AIPs and hence block autoinduction (Figure 1.9). Consequently, the lifetime of AIPs *in vivo* is on the order of a couple of hours (104). There are two major mechanisms by which AIPs are scavenged in the host serum. Firstly, AIP-I and AIP-IV contain, at their C-terminal position, a methionine residue essential for their receptor binding. As a consequence, oxidation of the thioether group on the methionine side chain by reactive oxygen species (ROS) into a sulfoxide leads to complete inactivation of these two AIPs (Figure 1.9) (78).

It was found thereafter that this oxidation also occurs in mice, and NADP oxidase-knockout animals are more susceptible to group-I *S. aureus* infection due to their deficiency in ROS production induced by immune response (104). In addition to oxidation, a more general neutralization strategy employing low-density and very-low-density lipoproteins (LDL and VLDL) is capable of removing AIPs free from oxidation-sensitive residues (e.g., AIP-III) (105, 106). Upon sequestration into lipoprotein particles, AIPs are then internalized by macrophages through receptor-mediated endocytosis and subject to lysosomal degradation. Notably, pre-oxidation of LDL by the ROS generated during the immune response significantly enhances its AIP-binding affinity, suggesting a

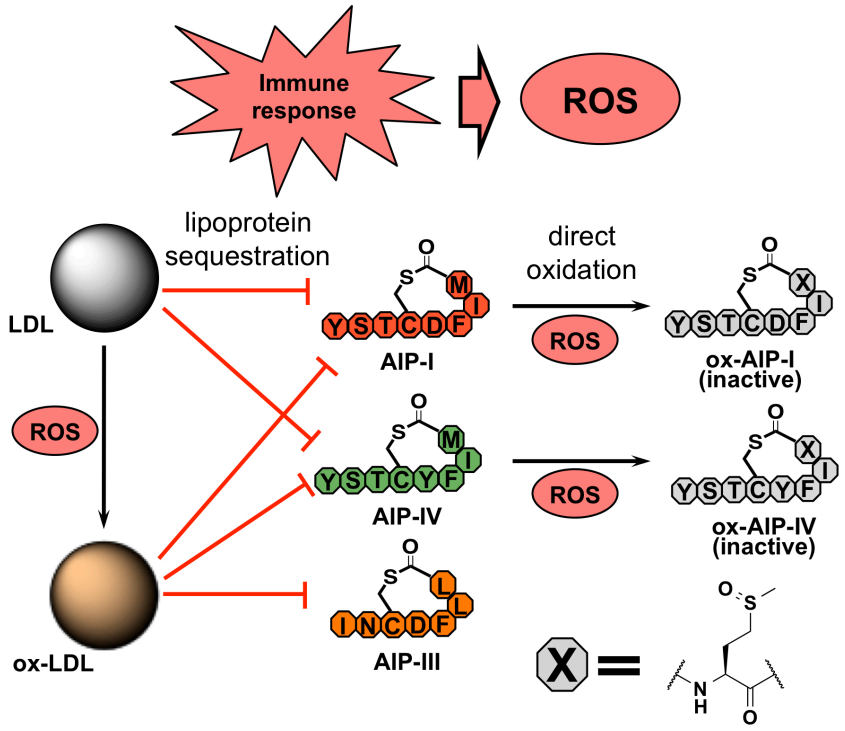


Figure 1.9 Mechanisms of AIP scavenging by mammalian hosts.
Diagram depicts two major mechanisms known to date, namely, lipoprotein sequestration and direct oxidation, that allows mammalian hosts to inactivate *S. aureus* AIPs. Inactivated AIPs upon oxidation are shown in gray, with X denoting the methionine sulfoxide generated by oxidation.

causal relationship between infection and the implication of AIP clearance machineries (Figure 1.9) (106). While this arrangement has its physiological merit, the lack of protection during the beginning few hours following invasion provides the pathogen opportunities to expand and establish its colonization.

1.3.2 AgrC-targeting reagents

In principle, chemicals blocking any step within the autoinduction loop should effectively accomplish quorum quenching. Unfortunately, out of five potential protein targets, i.e., AgrA, AgrB, AgrC, the AIP-releasing protease and the transporter, only the AgrA LytTR domain and the AgrC CA subdomain have had their crystal structures solved, while the last two proteins have not yet been unambiguously identified (72, 86). Consequently, target-specific drug design is predominantly limited to chemical engineering based on the AIP scaffold. Native AIPs provide a rich source of information for the design of QQ reagents due to their cross-group inhibitory properties. As opposed to native AIPs, however, QQ reagents are expected to inhibit the autoinduction of all four groups. Guided by the structure-activity relationship of AIPs (*vide supra*, section 1.2.4), AIP-I, -II and -III have all been successfully modified into global QQs of *S. aureus* (Figure 1.10a) (48, 77, 83). Despite this success, the AIP backbone is peptidic in nature and hence suffers such drawbacks as high immunogenicity and lack of stability *in vivo*. In a prominent attempt to address this problem, modification of single residues in AIP-III through peptoid-mimic substitution or N-methylation has generated a few global QQs (82). Full peptomer mimic possessing global QQ activities, however, has not yet been identified.

Aside from synthetic peptides derived from the native AIP scaffold, reported inhibitors of *S. aureus* autoinduction are predominantly secondary metabolites from other microbes. It should be pointed out that most these compounds have not been shown to directly interact with AgrC, nor have their activities been tested against all four subgroups of *S. aureus*. Nonetheless, some of these inhibitors share astonishing similarity to native AIP despite their disparate origin. For instance, solonamides, a series of tetrapeptides produced by marine bacteria from the genus *Photobacterium*, possess a 16-membered macrocycle containing four one β-hydroxy-acid and four α-amino-acid building blocks (Figure 1.10b) (107). Thus, solonamides and a few other *agr*-repressing natural products

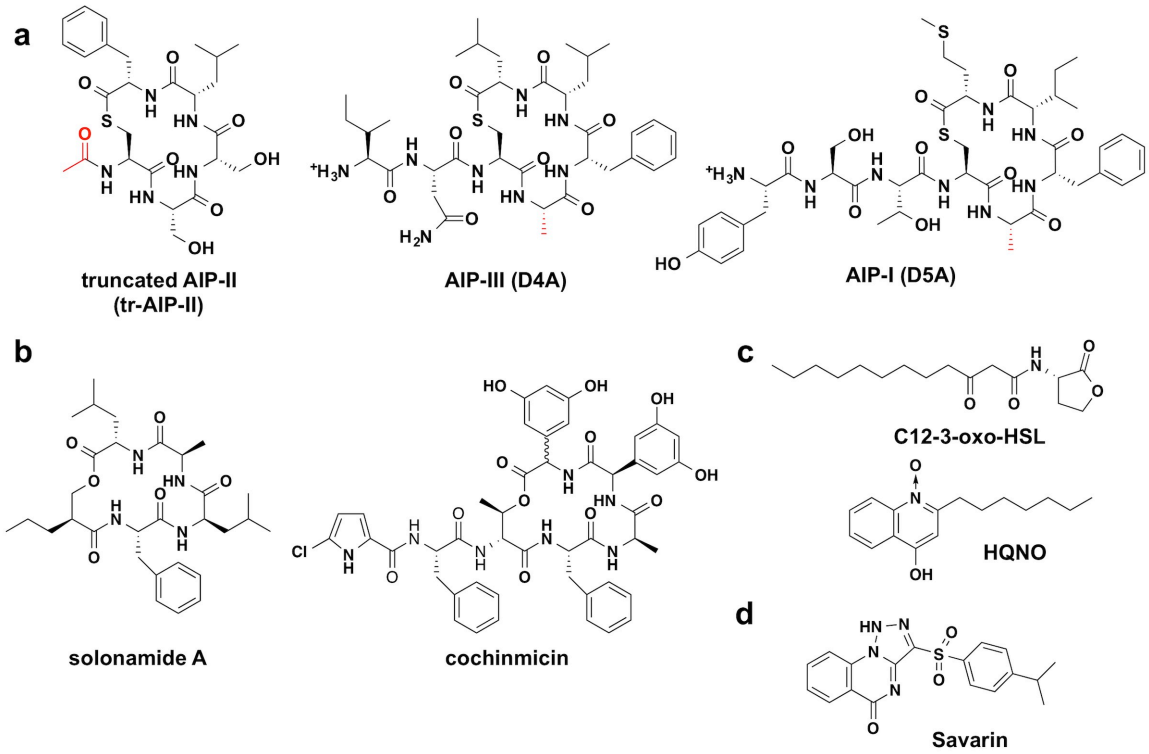


Figure 1.10 Synthetic molecules or natural products that targets *agr*.
(a) Global AgrC inhibitors derived from native AIPs. Groups that differ from the wild-type AIP are highlighted in red. **(b)** Putative AIP-mimic natural products. **(c)** The *Pseudomona* autoinducer 3-oxo-C12-HSL and secondary metabolite HQNO. **(d)** The AgrA-targeting lead compound, savarin.

containing AIP-like macrocycles, including avellanin and cochinmicin (Figure 1.10b) (108, 109), are believed to function through competitive inhibition of AgrC. Two other natural products, 3-oxo-C12-HSL and 4-hydroxy-2-heptylquinoline N-oxide (HQNO) originated from *Pseudomonas aeruginosa*, are capable of quenching *S. aureus* autoinduction with a low-micromolar IC_{50} (Figure 1.10c) (110). Interestingly, both *P. aeruginosa* and *S. aureus* are known to colonize the airways of cystic fibrosis patients, and these two compounds might play a role in the inter-species competition (111). In particular, the HSL compound has bactericidal activity at high-micromolar concentrations and may act as an allosteric inhibitor of AgrC at lower concentrations as was inferred from its interaction with AIPs in cell-based assays (112). Given the similar, amphiphilic structure of these two compounds, it is tempting to surmise that they function through interacting with the cell membrane.

1.3.3 AgrA-targeting reagents

In contrast to AgrC, which has four variants in *S. aureus*, targeting AgrA could be simpler due to its uniform sequence in all four subgroups. On the other hand, because AgrA localizes in the cytoplasm, membrane permeability would be required for effective AgrA inhibition in living organisms. As opposed to the SAR-guided design of AIP analogs into desired AgrC inhibitors, development of AgrA-targeting QQ reagents predominantly relied on the library screening approach employing cell-based readouts. AgrA was identified as the target in the validation process.

According to the crystal structure of AgrA LytTR domain in the absence of bound DNA, binding of a small organic compound to a hydrophobic cleft might disrupt the AgrA-DNA interaction (113). As a proof of principle, Leonard et al. screened over a 500-member library and identified a few AgrA-targeting hits with low-milimolar affinity (113). Screening of a much larger library led to the discovery of S*taphylcoccus* a*ureus* *vir*ulence *in*hibitor (savirin), which optimally blocks *S. aureus* autoinduction at 13.5μM through the same mechanism (Figure 1.10d) (114). This drug robustly inhibits autoinduction phenotypes of *S. aureus* and attenuates the lesion size in the classic murine abscess model. Importantly, resistance did not emerge upon extensive passage of *S. aureus* in the presence of savirin. Because of its optimal physicochemical properties and biological activities, savirin is viewed as a promising lead compound awaiting further SAR studies.

1.3.4 AIP-sequestration reagents

The last class of QQ reagent developed works through sequestration of AIPs from their interaction with AgrC. Employing a stable, oxo-ester variant of AIP-IV as the antigen, a set of AIP-IV specific, mouse monoclonal antibodies was developed (115). The antibodies showed excellent QQ activity *in vitro*, and protected mice from the development of skin-infection symptoms upon inoculation with a group-IV strain. Nonetheless, the prohibitive cost of the development and production of antibody drugs might present a formidable barrier for the application of QQ antibodies in practical anti-staphylococci therapies.

1.4 Summary: specific aims and significance

Elucidation of QS-dependent exo-protein regulation represents a major advance in the understanding of *S. aureus* physiology and pathology. Key to this achievement is the discovery and characterization of the *agr* locus, which can be divided into two phases demarcated by the sequencing of *agr* allelic variants and the isolation of the AIP. Dr. Richard P. Novick, a professor at New York University Medical Center and a close collaborator of the Muir Group, has made leading contributions throughout the first phase, primarily involving the isolation, cloning and sequencing of the *agr* locus. Thereafter, the Muir Group joined the campaign since 1997 and the two groups collectively initiated the second phase featuring the molecular and mechanistic characterization of *agr*. In the following years up until the beginning of my dissertation, the Novick-Muir collaboration has primarily focused on (i) the validation of the *agr* autoinduction circuit as a potential drug target and (ii) the dissection of molecular determinants in AIP and AgrC that define the affinity and signaling efficacy of their interaction. Meanwhile, other research groups around the world have attempted to understand the mechanisms of action for every single *agr* gene product, predominantly employing cell-based genetic/biochemical approaches and qualitative detection methods. Many fundamental biochemical questions, in particular with respect to integral membrane proteins AgrB and AgrC, were therefore left open. During the same period of time, geneticists have outlined the signaling network surrounding the autoinduction circuit, but little effort has gone to identifying missing components of the autoinduction circuit itself, namely the transporter and the AIP-releasing enzyme. In a quest to fill up the knowledge gaps listed above, I decided to investigate individual biochemical events

in the autoinduction circuit primarily through *in vitro* reconstitution using highly purified synthetic or recombinant components. Importantly, this strategy conforms to the pure-chemistry spirit and chemical biology strength of the Muir Group.

The overall aim of this dissertation is to perform comprehensive biophysical, biochemical and bioinformatic characterizations of AgrB and AgrC and thereby provide mechanistic insights to the institution and regulation of their enzymatic activities. Specific aims are addressed in each chapter. A comprehensive database search for homologs of AgrC proteins is included in Chapter 2, accompanied by inferences from the perspective of systematic evolution. To enable quantitative enzymology study on AgrC in various ligand states, the receptor was reconstituted to the nanometer scale lipid-bilayer discs, or nanodiscs, which contain a topologically open membrane. The reconstitution, the recapitulation of ligand responses of AgrC, and the determination of parameters describing its ligand binding and kinase activities are presented in Chapter 3. AgrC is also unique for having a number of activity states depending on the ligands bound. In Chapter 4, a conformation-activity landscape is charted and ligand-induced molecular motions detected for AgrC to shed light on its signaling flexibility. Chapter 5 is dedicated to the reconstitution of AgrB-catalyzed proteolysis of AgrD, focusing on identifying the driving forces that power the formation of the high-energy AgrD thiolactone intermediate without free-energy input from ATP.

QS-mediated behaviors are widely present in pathogenic bacteria. Gram-positive pathogens including *Listeria monocytogenes*, *Enterococcus faecalis* and *Clostridium*

perfringens all possess autoinduction systems sharing striking homology to the *S. aureus agr* system in both protein sequences and regulatory architectures (116-119). This dissertation establishes a solid body of knowledge concerning the biochemical principles of the *agr* system, which may provide guidance for understanding of virulence-protein regulation in many other clinically significant organisms. Moreover, the profound impact of *agr* on protein production and energy expense can be harnessed to modulate the microbe's fitness under certain selective conditions. Although the effectiveness of *agr*-targeting reagents as the standalone drug is still under debate, they still hold considerable potential as components of therapeutic cocktails against *S. aureus* infection. The *in vitro* reconstitution system of AgrB and AgrC may enable target-specific library screening employing the affinity-selection mass spectrometry (ASMS) strategy, which offers much larger throughput compared to cell-based screening assays and should expand the chance of discovering useful reagents in future anti-staphylococcal practices (120).

Chapter Two: Bioinformatic Analysis of the *S. aureus* AgrC

To confer the enzymatic as well as the regulatory functions, most class-I RHKs share five sequence motifs that enable them to bind the nucleotide/metal cofactor and, depending upon the signaling state of the sensor domain(s), adopt the active or inactive conformation through an appropriate DHp-CA interaction. In each motif, one or several position(s) are highly conserved for certain amino acids. Therefore, these motifs are named "homology boxes", denoted by the amino acid on their respective most conserved positions (69). The H-box, for instance, features the phospho-acceptor His residue, and is involved in the kinase-competent DHp-CA interaction as well as the DHp-RR interaction that supports the trans-kinase or phosphatase activity (Figure 2.1a) (121, 122). All other four homology boxes reside in the CA subdomain. Among them, the F-box is unique for its involvement in the kinase-inactive conformation: the conserved Phe associates to a hydrophobic patch on the DHp subdomain and thus holds back the CA from approaching the phospho-acceptor His (Figure 2.1b) (123). The remaining three motifs, namely the N-, G1- and G2- boxes, interact with the divalent metal, the adenine base and the triphosphate anion of the ATP-metal cofactor, respectively (Figure 2.1c-e) (124). Because of their high levels of conservation, database search for RHK sequences depends predominantly upon the detection of these homology boxes.

It has long been noticed that *S. aureus* AgrC protein possesses unusual sequence features in several of its homology boxes. In a seminal study carried out by Grebe and Stock (125), 384 RHK sequences known as of 1999 were classified into a series of subfamilies based on sequence homology in their HK domains. Together with 14 close homologs all

governing peptide-mediated intercellular communication, AgrC falls in the HPK10 subfamily, which was denoted to "show very distinct characteristics that clearly distinguish them from other histidine kinases" (125). These characteristics include: (i) a distinct H-box sequence pattern, F[RK]HDYxN; (ii) the absence of the first Asn residue in the N-box motif, NxxxN; and (iii) the lack of any detectable G1-box motif, DxGxG.

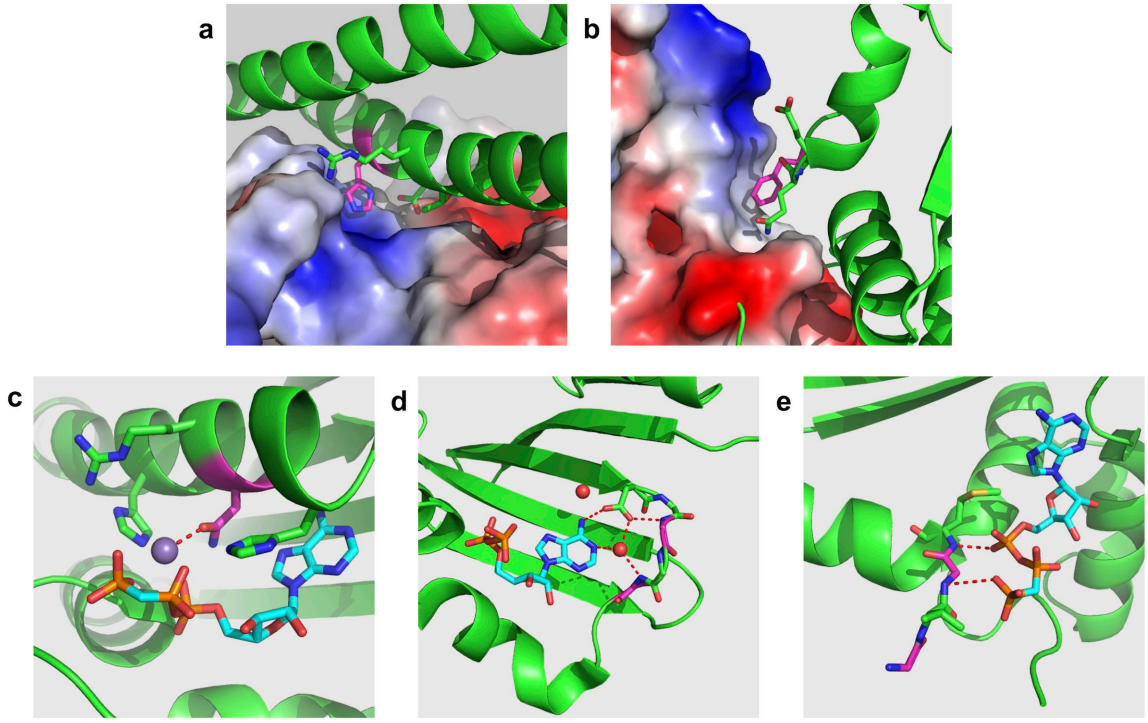


Figure 2.1 Primary functions of HK homology boxes captured in crystal structures. In all panels, the HK domain is shown in green ribbon diagram. Homology-box residues participating interactions and adenosine nucleotide (if present) are highlighted with stick model. Carbon atoms of the namesake residue and the adenosine nucleotide are colored in magenta and cyan, respectively. Red dashes denotes important hydrogen/coordination bonds. **(a)** The H-box of the *E. coli* stress sensor, CpxA docked to the surface of a CA subdomain in an auto-kinase-competent conformation. PDBID: 4biv. **(b)** The F-box of the *Thermotoga maritima* HK853 docked to a hydrophobic patch on the DHp subdomain to enforce the inactive conformation. PDBID: 2c2a. **(c)** The N-box of the *E. coli* chemoreceptor kinase, CheA in complex with ADPCP and Mn^{2+}. The metal cation is shown as a purple sphere. PDBID: 1i5a. **(d)** The G1-box of the *M. maritima* HK853 in complex with ADP. Red spheres are the oxygen atoms of two water molecules. PDBID: 2c2a **(e)** The G2-box of the CheA kinase in complex with ADPCP and Mn^{2+}. PDBID: 1i5a. In **(d)** and **(e)**, structural elements sheltering the homolog boxes are removed for clarity.

Recently, computational analysis identified strong sequence co-variation between the N- and H-box motifs, suggesting that they likely interact at the DHp-CA interface required for auto-phosphorylation (126). While a double-mutant cycle might account for the deviation of AgrC N- and H-boxes from canonical sequences, the AgrC sequence at the G1 region provides no hint on how this RHK harnesses ATP (or any triphosphate nucleotide) for auto-phosphorylation. In this chapter, we first seek to identify, from the genomic databases, more RHK sequences sharing the distinct features of the HPK10 subfamily. Sequence analysis was then performed, leading to a few hypotheses on the biochemical properties of AgrC and the evolution trajectory of the *agr* system.

2.1 Identification of close AgrC homologs from genomic databases

Two rounds of database search were performed (Figure 2.2). In the first round, the intracellular sequence of the group-I *S. aureus* AgrC (residues 205-430) was used as the

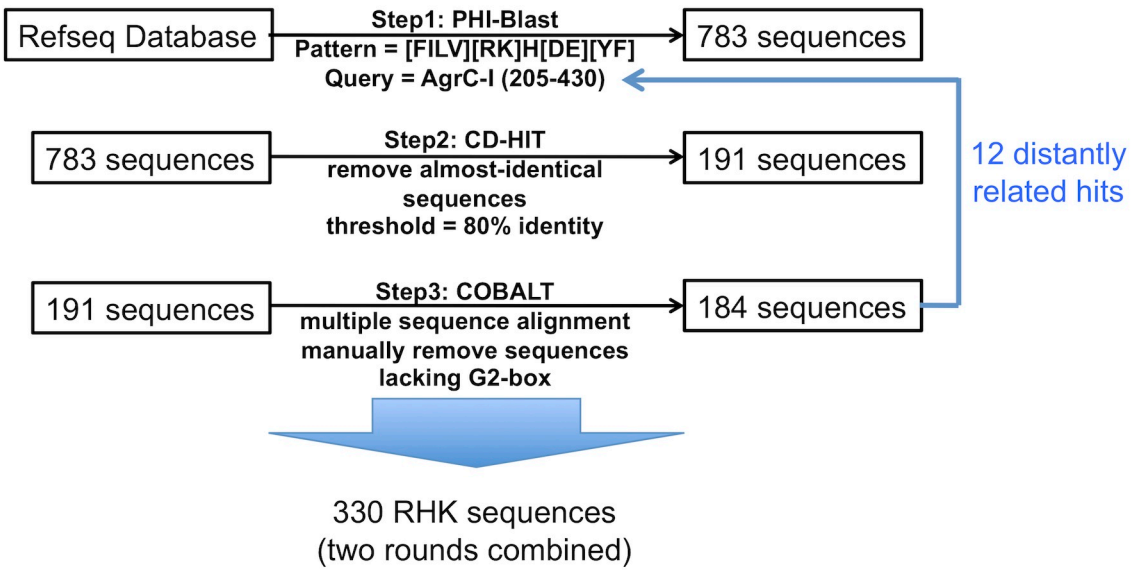


Figure 2.2 Flow chart of the database search for AgrC homologs. See text for details.

query to search the reference-sequence protein database (Refseq_protein) with *P*attern *H*it *I*nitiated BLAST (PHI-BLAST) (127). The initial pattern was [FILV][RK]H[DE][YF]. In order to preserve the initial pattern, no subsequent PSI-BLAST was performed. Hits with E-values below 0.05 were pooled. To reduce the redundancy of the hit pool, the CD-HIT program was used to cluster near-identical sequences with an identity threshold of 80% (128). One representative sequence from each cluster was included in the non-redundant sequence collection, which was aligned using the COBALT program for the detection of the CA domain (129). False (non-RHK) hits lacking a G2-box were removed manually. Out of 183 sequences obtained in the first round, 12 most distantly related to AgrC-I were chosen as queries for a second round of database search with similar strategies (see Appendix Table 1). Hit sequences obtained from both rounds were pooled and, upon removal of redundant and false hits, returned 330 non-redundant RHK sequences (Appendix Table 1).

To understand the G1-box configuration in AgrC, we generated a sequence profile of the 330-sequence collection using HMMER (130), and aligned it to an RHK with a canonical G1-box, the osmotic sensor EnvZ from *E. coli*. Strikingly, aligned to the first position of the EnvZ G1 motif, "DxGxG", is an Asn conserved in 326 out of 330 sequences, to which we refer as the 'G1-box Asn' (Figure 2.3). Notably, the frequency of Asn at this position is 9% in contrast to 85% for Asp in the GHKL super-family to which all CA subdomains belong (70). In contrast, amino acids aligned to either EnvZ G1-box glycines are not conserved among these 330 sequences. This result strongly suggests that AgrC, one of the 326 possessing the G1-box Asn, employs this residue to interact with the

nucleotide cofactor. Furthermore, since we did not assign any sequence pattern when performing the PHI-BLAST, this unique G1-box feature must have co-evolved, directly or indirectly, with the input H-box sequence pattern.

Further analysis revealed another common feature of these 326 sequences: although covering 194 species from 22 taxonomic families, they all come from low-GC gram-positive bacteria in line with 15 founding members of the HPK10 subfamily (Table 2.1). Interestingly, the four sequence hits that did not conform to these two criteria contain both a canonical G1-box and originate from Gram-negative bacteria or archaea. These two features suggested that this 326-sequence collection represents a self-consistent expansion of the original RHK10 subfamily. It is noteworthy that the lack of first Asn in the NxxxN N-box motif [characteristic (ii)] no longer holds true within this expanded RHK10 subfamily.

2.2 Sequence analysis of the HPK10 subfamily

Among the 326 sequences, 318 were predicted to span the membrane at least once and 286 to span five times or more, according to the TMHMM program (Figure 2.4a) (131).

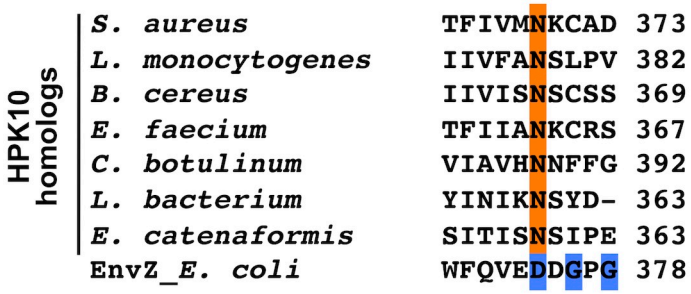


Figure 2.3 Alignment of the *E. coli* EnvZ protein to HPK10 subfamily members at the G1 box. The G1-box Asn and the conserved residues in canonical G1 boxes are highlighted in orange and blue, respectively.

Interestingly, with only one exception, none of these multi-span HPK10 sequences is predicted to contain an extracellular sensor domain, suggesting that most HPK10

Table 2.1 Taxonomic coverage of the HPK10 subfamily

Phylum	Class	Order	Family
Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
	Mollicutes	Acholeplasmatales	Acholeplasmataceae
	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae
	Clostridia	Clostridiales	Lachnospiraceae
			Peptostreptococcaceae
			Eubacteriaceae
			Clostridiaceae
			Ruminococcaceae
			Peptococcaceae
		Thermoanaerobacterales	Thermoanaerobacteraceae
	Bacilli	Bacillales	Listeriaceae
			Sporolactobacillaceae
			Paenibacillaceae
			Bacillaceae
			Staphylococcaceae
			Planococcaceae
		Lactobacillales	Aerococcaceae
			Carnobacteriaceae
			Enterococcaceae
			Streptococcaceae
			Lactobacillaceae
			Leuconostocaceae

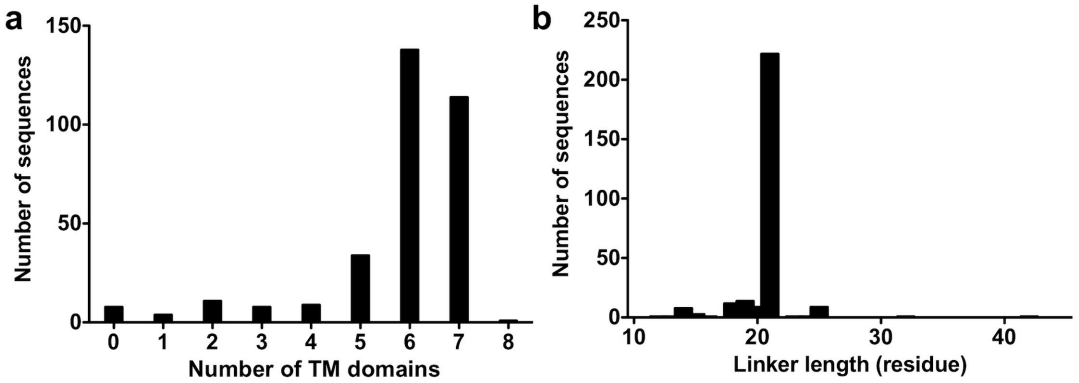


Figure 2.4 Statistics of the expanded HPK10 subfamily.
(a) Number of TM domains predicted for the 326-sequence collection by the software package, TMHMM 2.0. **(b)** Predicted length of the interdomain linker (in residues) in the multiple sequence alignment of the 285 sequences.

subfamily members likely sense extracellular or intra-membrane signals with membrane-embedded sensors (132). To better delineate the boundaries of the AgrC-I intracellular modules, we generate a new hidden Markov model (HMM) using the alignment of these 286 multi-span sequences, and harnessed it to analyze a sub-alignment of 18 sequences, each representing a taxonomic family (Figure 2.5a). These analyses revealed a less conserved, short linker between the well-aligned last TMH and the DHp subdomain in all sequences (Figures 2.4b and 2.5b). Specifically in AgrC-I, the conserved DHp subdomain that aligns to the HMM with high confidence begins at residue 222 rather than at residue 200 immediately after the last TMH (Figure 2.5b). We therefore refer to AgrC-I$^{1\text{-}200}$ as the sensor domain and AgrC-I$^{201\text{-}221}$ as the interdomain linker region.

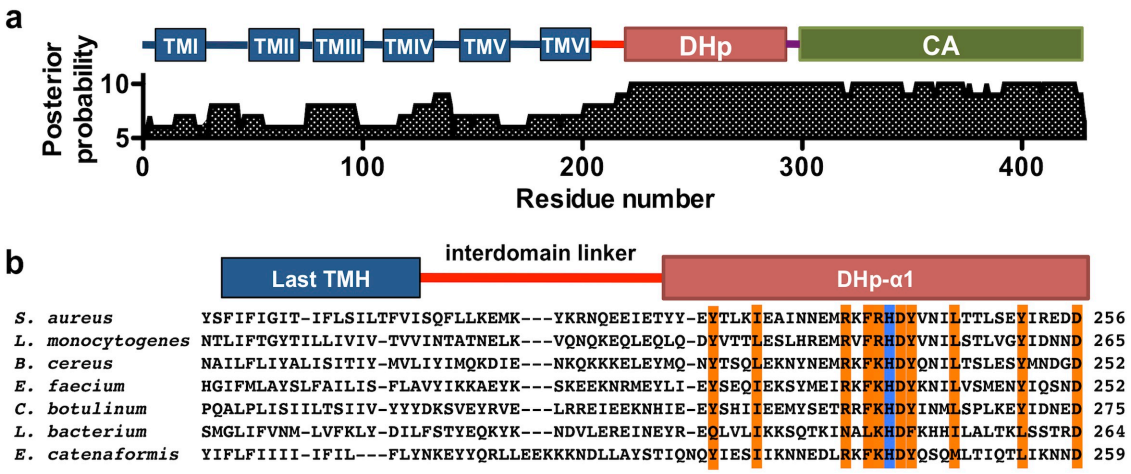


Figure 2.5 Identification of a less conserved interdomain linker within AgrC
(a) Posterior probabilities of the 18-sequence sub-alignment according to the profile HMM: positions in the HMM are displayed with numbering corresponding to residues in AgrC-I; predicted domains of AgrC-I are shown above. **(b)** HPK10 subfamily alignment from the last TMH to the first DHp helix: highly conserved residues and the phospho-acceptor histidine are highlighted in orange and blue, respectively.

2.3 Summary and discussions

The taxonomic restriction of the HPK10 subfamily revealed in this chapter provides intriguing insights on the evolution trajectory of the AgrCA TCS as well as the *agr* QS circuit. First of all, that all subfamily members come from the phylum *Firmicutes* strongly supports the theory that the ancestral HK domain carrying all HPK10-specific features emerged after the branching of this phylum from the tree of life. A subphylum-specific emergence is also possible, as the distribution of HPK10 members across the phylum is highly uneven, with more than 95% sequence found in orders *Bacillales, Lactobacillales* and *Clostridales*.

Despite the rapid expansion of the sequence database, our knowledge on the function of HPK10 subfamily members has not progressed as much. Founding members of the subfamily were discovered about two decades ago for their involvement in the autoinduction circuits encoded by homologs of the *S. aureus agr* locus and the *Streptococcus pneumoniae com* locus—and this remains the only known function of this RHK subfamily to date (125). In our database search, a number of HPK10 genes were found without an *agrB* or a *comAB* homolog (see below) in their genomic context, and their functions are so far intractable. The AI of the prototypical streptococcal *com* system is a linear peptide named the CSP (for competence-stimulating peptide), which is sensed by the ComDE TCS, a close homolog of AgrCA (Figure 2.6) (133). Notwithstanding the homology of the AI-sensing module, the ComAB machinery that both processes and secretes the CSP is completely unrelated to AgrB—rather it shares close homology with the plasmid-borne *E. coli* hemolysin transporter, HlyBD (Figure 2.6) (134, 135).

Interestingly, the AgrBD homologs were also found to exist in a broader spectrum of *Firmicutes* than the intact *agr* system (136). These findings raised the possibility that both systems emerged through the co-regulation of pre-existing AI production and sensing genes, which were subsequently grouped into a single operon. Given the prevalence of *agr* homologs in *Bacillales* and *Clostridales* and *com* homologs in *Lactobacillales*, these two systems likely emerged independently and gained their autoinductory function thereafter through convergent evolution. A comparative analysis of the phylogeny of the HPK10 subfamily to that of AgrB and ComAB would test this hypothesis more rigorously.

The bioinformatic analysis described in this chapter was directed to address two questions, namely, what aspects of AgrC biochemical properties we should focus in the subsequent *in vitro* reconstitution study and how to interpret the results from an evolutionary perspective. While the classic theme of phospho-relay has been well

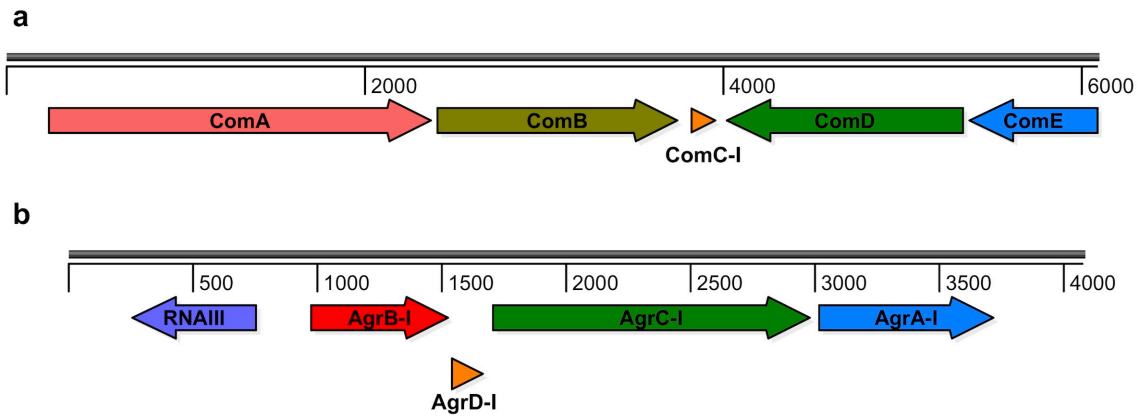


Figure 2.6 Comparison between the prototypic *com* and *agr* loci
Gene-products encoded by the **(a)** *com* and **(b)** *agr* loci are shown in arrows indicating their direction of transcription. Coding region of the HPK10-subfamily RHK and the LytTR-family RR are colored in green and blue, respectively.

established for TCSs, the sequence peculiarity in the H- and G1-boxes of the HPK10 subfamily foreshadows some intriguing subfamily-specific behaviors related to these two homology boxes. Specifically, the unique H-box sequence may enable a distinct regulatory pattern of the AgrC auto-kinase, trans-kinase and phosphatase activities following signal input. We will therefore investigate all three above activities of AgrC under resting, activated or inhibited ligand states. Furthermore, as the only conserved residue that putatively interacts with the nucleobase, the G1-box Asn must either harness a different triphosphate nucleotide as the phosphoryl donor, or bind to the ATP adenine at an altered affinity. Consequently, the nucleotide specificity and the affinity to the favorite nucleotide will be another focus of our biochemical studies. Data that distinguish AgrC from the majority of RHKs are expected, and will be interpreted within the boundary of the HPK10 subfamily and the extracellular peptide sensing in *Firmicutes*.

Chapter Three: Reconstitution of Biochemical Events Related to AgrC

Quantitative information on how AgrC works as a kinase and how its activity is regulated by native AIP ligands is critical to the understanding of the regulatory architecture of *S. aureus* autoinduction. Unfortunately, such information is in a short supply, however, primarily due to the limitations of cell-based reporter assays. The complicated whole-cell system occludes specific detection of individual proteins, while the reporter-protein readout cannot faithfully reflect the rate of a single biochemical event involved in two-component signaling cascade. To circumvent these drawbacks and provide quantitative insights into the activity and regulation of this pivotal receptor histidine kinase (RHK), an *in vitro* reconstitution system involving highly purified components must be implemented. In this chapter, we describe the reconstitution of AgrC homologs to an artificial membrane system that restores their expected biochemical activity and ligand response. During the course of our reconstitution attempts, a polyclonal, pan-specific antibody for phosphohistidine (pHis) detection was developed in the Muir Lab, enabling highly efficient, quantitative detection of AgrC phosphorylation (137). Taking advantage of this new reagent as well as our reconstitution system, we determined physicochemical parameters for the interaction between AgrC-I and native AIPs, as well as kinetic parameters of the RHK's auto-kinase and phospho-relay activities in the *apo*- and activator-bound states.

3.1 Functional reconstitution of AgrC to nanodiscs

We elected to use nanometer-scale lipid-bilayer discs (nanodiscs) as the model membrane for the incorporation of AgrC (Figure 3.1a) (138). Nanodiscs are a supramolecular mimic

of high-density lipoproteins consisting of a discoidal lipid bilayer surrounded and stabilized by two copies of a membrane scaffold protein (MSP). Importantly, the topologically open nanodiscs, as opposed to membrane vesicles, enables AgrC, when incorporated, to simultaneously access its AIP ligands and enzymatic substrates in the bulk solution because both sets of reagents are membrane-impermeable.

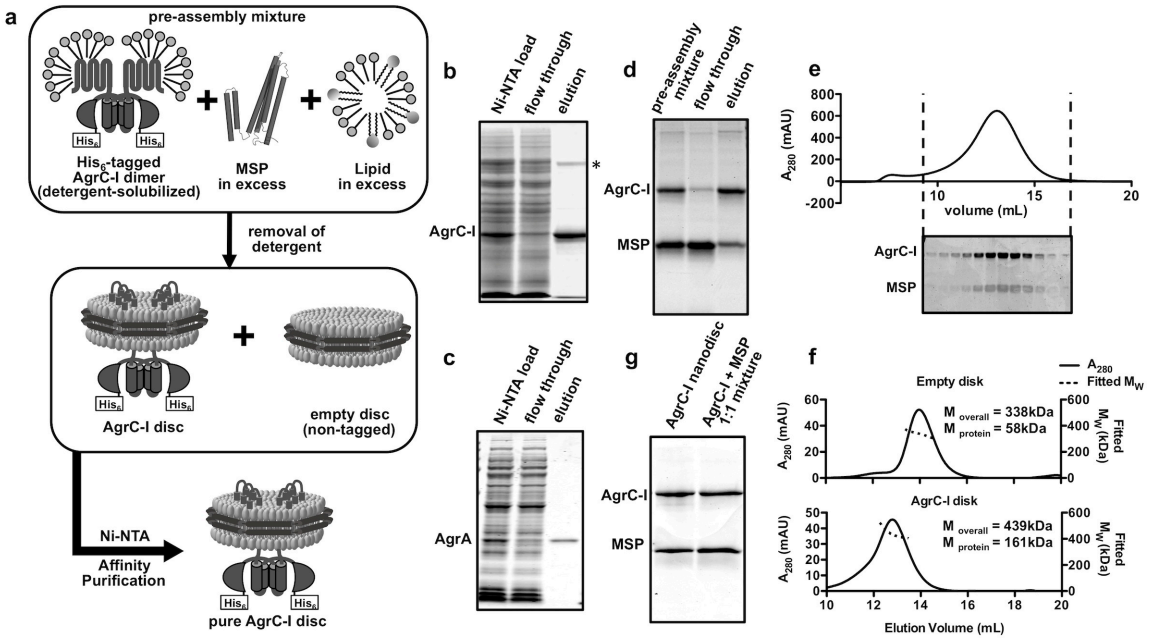


Figure 3.1 Reconstitution of AgrC-I to nanodiscs composed of DMPC and DMPG
(a) Schematic showing the workflow used for reconstituting AgrC-I dimers into nanodiscs. **(b** and **c)** Purification of **(b)** AgrC-I from *E. coli* membrane extract and **(c)** AgrA from *E. coli* total soluble lysate as analyzed by SDS-PAGE with CBB-stain. Bands corresponding to the target proteins are indicated. Asterisk indicates a 100-kDa protein co-purified with AgrC-I. **(d)** Purification of reconstituted AgrC-I discs by Ni-NTA affinity chromatography as analyzed by SDS-PAGE with Coomassie-brilliant blue (CBB) staining; bands corresponding to AgrC-I and the scaffold protein (MSP) are indicated. **(e)** AgrC-I discs analyzed by size exclusion chromatography (SEC). Fractions collected between the dashed lines were further analyzed by SDS-PAGE with CBB staining (bottom). **(f)** SEC-MALS analysis of empty and purified AgrC-I discs. $M_{overall}$ corresponds to the average molecular weight (M_W) of the peak calculated from the UV, light-scattering, and differential refractive-index readouts. $M_{protein}$ corresponds to the measured M_W of the proteins in each case; AgrC-I = 51 kDa, MSP = 29 kDa. **(g)** Stoichiometry of AgrC-I discs as analyzed by SDS-PAGE: AgrC-I discs (left lane) and an equimolar mixture of purified AgrC-I and MSP (right lane) were resolved and visualized with CBB staining.

Among four AgrC homologs in *S. aureus*, the group-I variant was the first to be successfully over-expressed in *E. coli* and hence employed in nanodisc-reconstitution trials. Both AgrC-I and AgrA were purified as His_6-tagged recombinant proteins (Figure 3.1b and c). The detergent-solubilized AgrC-I did not have any auto-kinase activity regardless of the presence of AIP (data not shown). We then set out to prepare homogeneous nanodiscs containing one copy of the AgrC-I dimer per disc. To this end, purified AgrC-I was added to mixed micelles containing excess MSP and lipids [1,2-dimyristoyl-sn-3-glycero-phosphocholine (DMPC) and 1,2-dimyristoyl-sn-3-glycero-phosphochglycerol (DMPG), molar ratio 1:3, see below], lowering the chance of incorporating more than one copy of AgrC-I dimer in each nanodisc (Figure 3.1a) (139). Detergent was then removed to drive self-assembly, and nanodiscs containing AgrC-I (AgrC-I discs) were separated from empty discs using Ni-NTA affinity purification (Figure 3.1a and d). Purified AgrC-I discs gave a single, symmetric peak on size-exclusion chromatography (SEC) from which the peak fractions were pooled for further studies (Figure 3.1e). The product has a narrow molecular weight distribution analyzed by size-exclusion chromatography in tandem with multi-angle light scattering (SEC-MALS, Figure 3.1f) (140). The expected stoichiometry of one copy of AgrC-I dimer per disc was also confirmed in the SEC-MALS experiment, as well as using SDS-PAGE employing ratiometric standards (Figure 3.1g).

In preliminary trials of AgrC-I reconstitution, a series of phosphatidylcholines (PC) were tested for their ability to mediate nanodisc assembly (139). DMPC was selected because of the relative ease in obtaining a monodisperse population of AgrC-I discs (data not

shown). Notwithstanding this favorable behavior, AgrC-I embedded in DMPC nanodiscs, although possessing a basal level of auto-kinase activity, failed to activate upon binding to AIP-I, its cognate activator peptide (Figure 3.2a, lanes 1 and 2). Because the major phospholipid in *S. aureus* cell membrane is the anionic phosphatidylglycerol (PG), we adjusted the nanodisc lipid to a mixture of DMPC and DMPG at a molar ratio of 1:3 and successfully restored the AIP-I dependent auto-kinase activation of AgrC-I (Figure 3.2a, lanes 3 and 4) (68). Surprisingly, co-migration assays showed that AgrC-I embedded in nanodiscs with either lipid composition binds to the activator AIP with equal capacity,

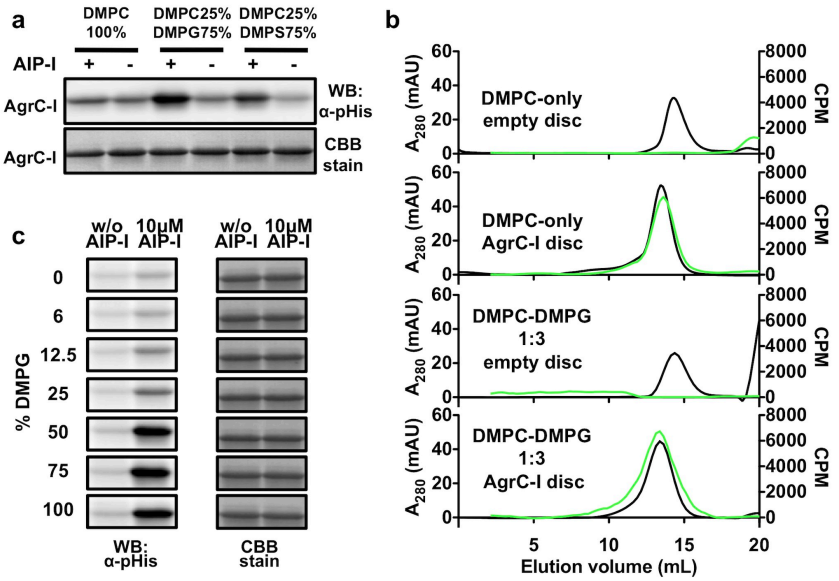


Figure 3.2 Dependence of AgrC-I activation on lipid composition
(a) AgrC-I auto-kinase activation: AgrC-I discs reconstituted with different lipid mixtures were phosphorylated with 1 mM ATP at 37 ^{o}C for 3 minutes. The cognate AIP-I was included as indicated. pHis levels in AgrC-I were analyzed by anti-pHis western blotting (upper). The blot was CBB stained thereafter as a loading control (lower). **(b)** Co-migration assay of a fluorescein-labeled AIP-I, FAM-AIP-I (refer to Figure 3.7a) with selected nanodisc samples: AgrC-I or empty discs assembled with DMPC only or a DMPC-DMPG mixture were incubated with excess FAM-AIP-I and analyzed by SEC. UV and fluorescence detection of the chromatograms are overlaid. **(c)** AgrC-I discs reconstituted with DMPC and DMPG at a series of ratios were tested for auto-kinase activation as in **(a)**.

suggesting that the zwitterionic DMPC membrane disrupts the translation of signal binding into enzymatic response (Figure 3.2b). Lipid environments enriched for anionic DMPS largely restored the responsiveness of the receptor to AIP (Figure 3.2a, lanes 5 and 6). To further dissect the role of zwitterionic and anionic lipids in AgrC-I activation, we reconstituted the receptor into nanodiscs consisting of DMPC and/or DMPG at seven different ratios. The magnitude of AIP-I-dependent activation increased gradually as the DMPG content increased, reaching a maximal at a DMPG content of 75% (Figure 3.2c). These data suggested that a reasonably high negative-charge density, rather than the presence or absence of any specific lipid per se, is required for the full activation of the AgrC-I auto-kinase. Thus, all AgrC-I discs in this dissertation, unless noted otherwise, were assembled from this activation-optimal lipid mixture containing DMPC and DMPG at a molar ratio of 1:3.

Subsequent success in the expression and purification of AgrC-II and AgrC-III allowed us to reconstitute these two receptors for comparative studies versus AgrC-I (Figure 3.3a). Unexpectedly, neither AgrC-II nor AgrC-III was activated by the cognate AIP when embedded in nanodiscs composed of the PG-rich, 1,2-dimyristoyl (DM) phospholipid mixture (Figure 3.3b, lanes 7 and 8; 19 and 20). In search of a membrane environment that bolsters their activation, we further varied the fatty-acyl groups of the lipids participating the reconstitution. Gratifyingly, a mixture of 1-palmitoyl-2-oleyl-sn-glycero-phosphocholine (POPC) and 1-palmitoyl-2-oleyl-sn-glycero-phosphoglycerol (POPG) at 1:3 molar ratio supports robust activation to both receptors (Figure 3.3b, lanes 11 and 12; 23 and 24). As a result, auto-kinase assays involving the comparison among

three AgrC variants were performed on nanodiscs assembled with this lipid mixture. It is noteworthy, however, that the AgrC nanodiscs prepared using 1-palmitoyl-2-oleyl (PO) phospholipids did not exhibit solution behaviors as homogenous as those made of DM phospholipids. The heterogeneity of the former was reflected by the SEC elution profile, in which a shoulder eluted earlier than the major peak indicated the presence of abnormally large particles (Figure 3.3c). As a result, only nanodiscs from the major-peak

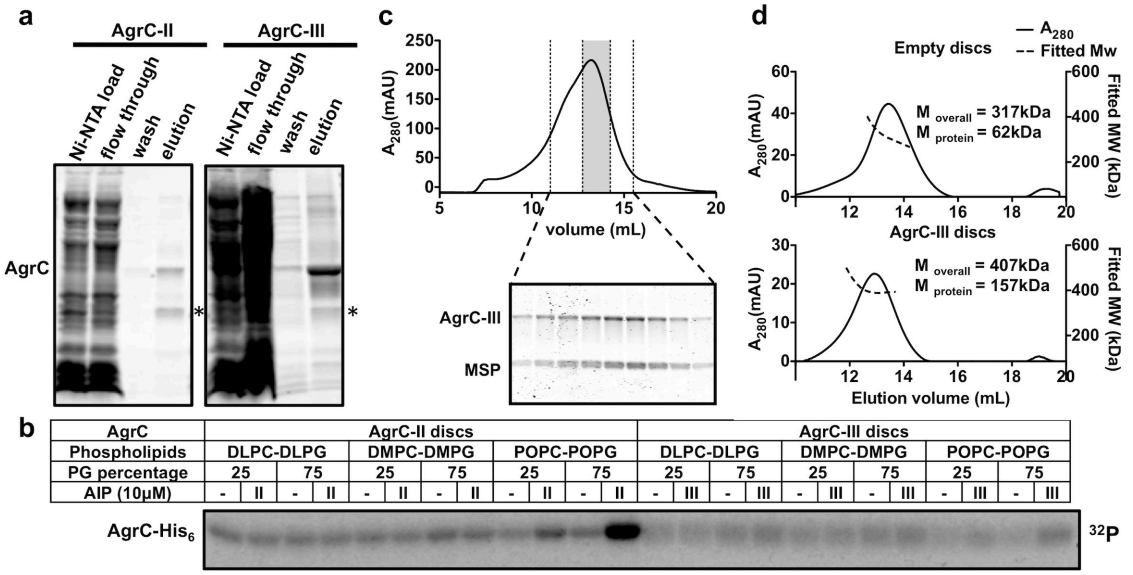


Figure 3.3 Nanodisc reconstitution of AgrC-II and -III using POPC and POPG
(a) Purification of AgrC-II and AgrC-III from *E. coli* membrane extracts as analyzed by SDS-PAGE with CBB-stain. The position of target-protein bands is indicated by "AgrC". Bands indicated by asterisks likely correspond the soluble HK domain generated from the proteolysis of the full-length protein by host proteases. **(b)** Lipid-composition screen for AgrC-II and AgrC-III activation: 3 μM AgrC dimer, 12 μM MSP and appropriate amount of the indicated phospholipid stock were mixed and subject to detergent removal. The post-assembly mixtures, without further purification, were treated with 20 μM [γ-^{32}P] ATP, 5 mM $MgCl_2$ and the activator AIP (if indicated) at 37 ^{o}C for 40 minutes. pHis levels were analyzed by autoradiography. **(c)** SEC traces of post-assembly mixtures exemplified by AgrC-III discs. Fractions collected between the wider dashed lines were further analyzed by SDS-PAGE with CBB staining (bottom). Shadow indicates the segment of elution from which AgrC-III discs with expected sizes were pooled for further use. **(d)** SEC-MALS analysis of empty and purified AgrC-III discs. As in Figure 3.1f, $M_{overall}$ and $M_{protein}$ of the peak were shown in each case; AgrC-III = 52 kDa, MSP = 29 kDa.

fractions (Figure 3.3c, shadowed area) were collected for biochemical studies. The expected, one-dimer-per-disc stoichiometry of this nanodisc sample was confirmed employing SEC-MALS (Figure 3.3d).

3.2 Recapitulation of AIP-dependent regulation of AgrC kinase activity

Both AIP and AgrC from *S. aureus* subgroup-I have been extensively investigated for their structure-activity relationships (48, 74, 75, 78). These studies provided access to potent synthetic AIP agonists and antagonists of this system as well as a body of genetic data on the cognate RHK, AgrC-I, including the availability of constitutive mutants of the receptor. Next, we attempted to recapitulate major findings from these studies employing our *in vitro* system. Auto-kinase assays revealed a baseline level of activity in AgrC-I dimers embedded in nanodiscs composed of DMPC-DMPG mixture (Figure 3.4a, lane 5). Native AIPs from all four *S. aureus agr* variants also behaved in line with what has been observed previously in cell-based studies: AIP-I, the cognate AIP, strongly stimulated the auto-kinase activity of AgrC-I; AIP-II suppressed AgrC-I activity below the basal level and so acted as an inverse agonist, AIP-III behaved as a neutral antagonist and so did not effect the basal activity, and AIP-IV was an agonistic ligand with similar efficacy to AIP-I (Figure 3.4a, lanes 1-4 and Figure 3.4b) (48). Interestingly, a truncated AIP-I analog, tr-AIP-I, behaved as a partial agonist of AgrC-I when added at saturating concentrations (Figure 3.4a, lane 6, Figure 3.4b and c) (48). Importantly, AgrC-I discs were pHis-free prior to treatment with ATP, and the phosphorylation of AgrC-I survived treatment with a few common buffer molecules at high concentrations, but was labile to hydroxylamine treatment, a signature of pHis-containing proteins (Figure 3.4d and e) (141). The

phosphorylation level also decreased upon incubation with ADP, which is indicative of the reverse auto-kinase reaction seen in other systems (Figure 3.4d) (142). As expected, auto-kinase activity was abolished when we reconstituted AgrC-I carrying inactivating

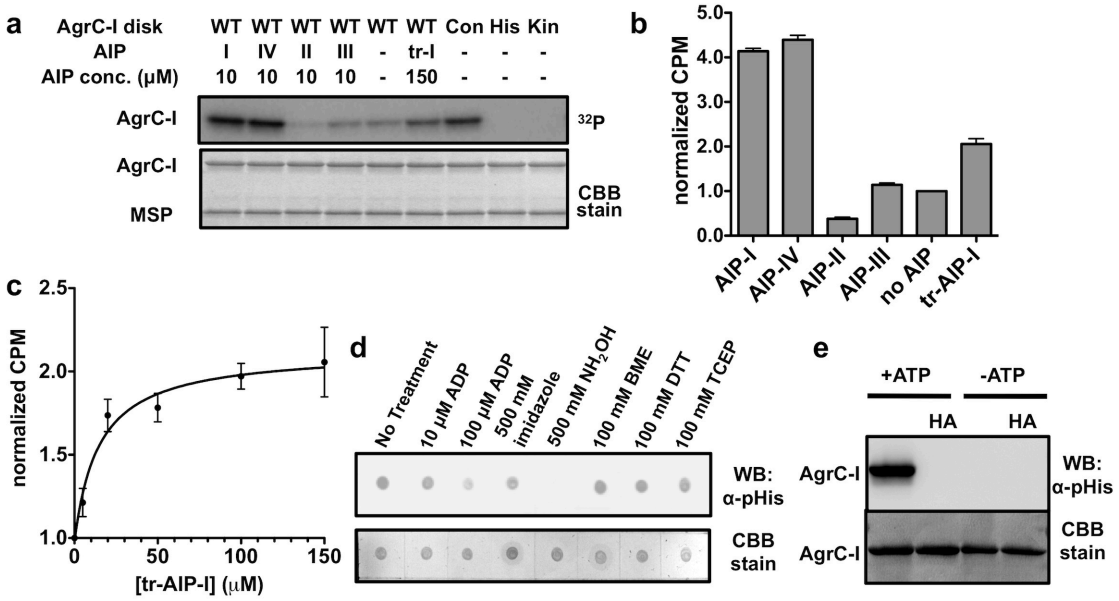


Figure 3.4 Auto-kinase activity and AIP-dependent regulation of AgrC-I
(a and **b)** Auto-kinase assays with AgrC-I variants or AIP analogs: 1.4 μM nanodiscs containing AgrC-I wild-type (WT), constitutive mutant R238H (Con), DHp-inactive mutant H239Q (His) or CA-inactive mutant G394A/G396A (Kin) were phosphorylated with 20 μM [γ-^{32}P] ATP at 37 ^{o}C for 40 minutes. WT-AgrC-I was treated with indicated AIP peptides; truncated AIP-I is referred to as tr-I or tr-AIP-I. pHis levels were analyzed by autoradiography in **(a)** or quantified using scintillation counting in **(b)**. Bar-graph shows CPM values normalized to the AIP-free reaction. Error bars = SD (n = 3). **(c)** Saturation of tr-AIP-I binding to AgrC-I. 1.4μM AgrC-I discs were treated with tr-AIP-I at indicated concentrations and then incubated with 20 μM [γ-^{32}P] ATP at 37 ^{o}C for 40 minutes. Plot shows pHis levels normalized to the peptide-free reaction versus tr-AIP-I concentration, fit to a saturation-binding model. Error bars = S.D. (n = 3). **(d)** Stability test of phosphorylation on AgrC-I. Pre-phosphorylated AgrC-I discs (refer to Figure 3.10c) were treated with indicated reagent at 37 ^{o}C for 20 min in a buffer containing 50 mM Tris, pH = 8.0, 100 mM NaCl, 5 mM $MgCl_2$. pHis levels of AgrC-I were analyzed by anti-pHis immuno-blotting (upper). The blot was CBB-stained thereafter as a loading control (lower). **(e)** Hydroxylamine (HA) treatment of phosphorylated AgrC-I. AgrC-I discs were phosphorylated with ATP (lanes 1 and 2) as described in Figure 3A or mock-treated (lanes 3 and 4) and then incubated with either buffer (lanes 1 and 3) or 500 mM HA (lanes 2 and 4) at 37^{o}C for 5 min. pHis levels of AgrC-I were then analyzed by western blotting.

mutations in either the CA or DHp subdomains (Figure 3.4a, lanes 8 and 9) (71). Conversely, incorporation of a constitutively active AgrC-I mutant into nanodiscs led to significantly elevated levels of baseline activity (Figure 3.4a, lane 7) (143), providing further evidence that this *in vitro* reconstitution system replicates the cellular behavior of this RHK.

Quorum sensing is mutually inhibitory among *S. aureus* subgroups-I, -II and –III. Because the effects of AIP-II and AIP-III on AgrC-I auto-kinase activity are qualitatively different, it would be interesting to also distinguish inverse agonism versus neutral antagonism on kinase activities conferred by other non-cognate AIP-AgrC pairs. To this end, auto-kinase assays were performed on three AgrC homologs incorporated to POPC-POPG nanodiscs (Figure 3.5a and b). In this membrane environment, AgrC-I exhibited similar responses to AIP-I and –II, although a slight but significant activation was observed in the presence of AIP-III. The efficacy of non-cognate AIP binding is reciprocal between every two groups: AIP-I confers weak activation to AgrC-III, while all combinations involving AIP or AgrC from the more distantly related group-II causes repression of the kinase activity. Intriguingly, with the presence of excess cognate AIPs, AgrC-II and AgrC-III exhibited 70% and 40% the activity of their group-I homolog (Figure 3.5a and b). The activity ranking of these AgrC homologs in their activated states is consistent with the order of autoinduction timing of their respective subgroups: the most active AgrC-I mediates the earliest induction in group-I cells, while group-III cells harboring the least active AgrC-III induce at highest cell densities (50). Thus, the uneven

kinase activity in AgrC homologs may act as the intrinsic timer for *S. aureus* autoinduction.

3.3 The AgrC-I-AgrA phospho-relay and AgrA dephosphorylation

Phospho-relay from AgrC-I to AgrA was reconstituted by adding the purified full-length RR to AgrC-I discs that had been pre-phosphorylated with [γ-^{32}P]-ATP and then exchanged into an ATP-free buffer. Time courses revealed efficient phosphoryl group transfer between the RHK and RR, with a $t_{1/2}$ of 66 seconds (Figure 3.6a, lanes 2-8 and Figure 3.6b). The inclusion of AIP-1 in the mix had little effect on the rate of transfer

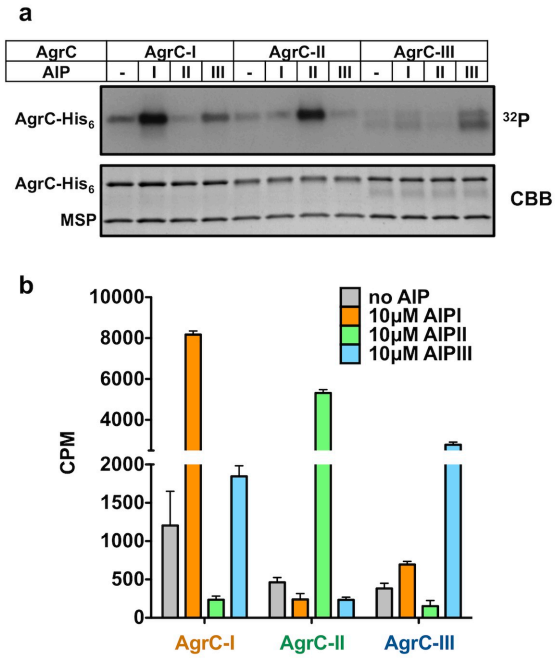


Figure 3.5 Comparison of the auto-kinase activity and regulation among *S. aureus* subgroups-I, -II and -III
Auto-kinase assays of three AgrC homologs: 1.4 μM POPC-POPG nanodiscs containing wild-type AgrC-I, AgrC-II or AgrC-III were pre-incubated with the indicated AIP peptide or vehicle and then treated with 20 μM [γ-^{32}P] ATP and 5 mM $MgCl_2$ at 37 ^{o}C for 40 minutes. pHis levels were analyzed by autoradiography in **(a)** or quantified using scintillation counting in **(b)**. In **(b)**, bar graph shows CPM values grouped according to AgrC homologs, with bars colored for different ligand states. Error bar = S.D. (n = 3).

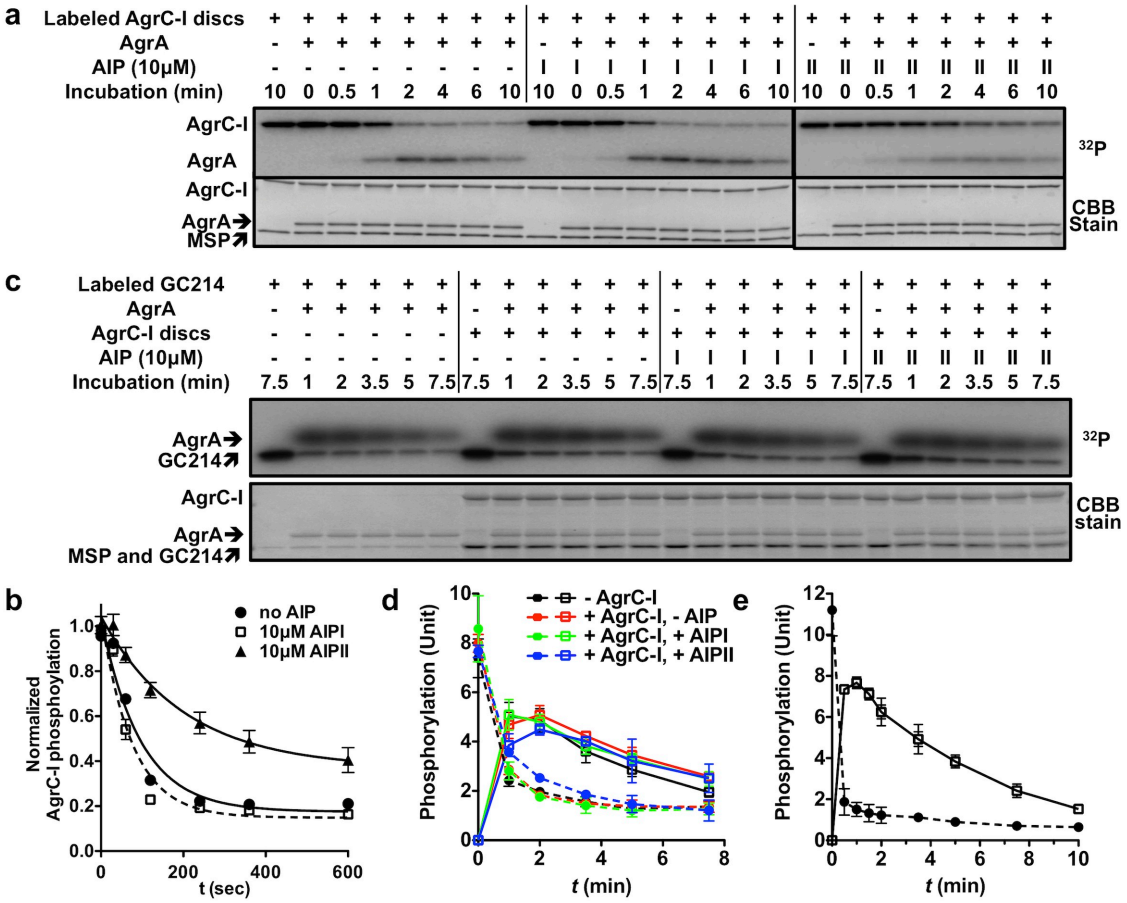


Figure 3.6 Phospho-relay and phosphatase assays of AgrC-I and AgrA
(a) Phospho-relay from AgrC-I to AgrA: AgrA was incubated with AgrC-I discs labeled with [γ-^{32}P]-ATP and indicated AIPs. Aliquots removed at various time-points were analyzed by SDS-PAGE followed by autoradiography. **(b)** Quantification of the autoradiograms in panel **(a)** showing decay of pHis levels in AgrC-I (normalized to t = 0) as a function of time. Error bars = range (n = 2). **(c and d)** Dephosphorylation of AgrA: AgrA (2 μM) was phosphorylated by incubation with the constitutively active chimera protein GC214 (0.1 μM dimer – see text for details) labeled with a [^{32}P]-phosphoryl group. This reaction was then supplemented with 2 μM AgrC-I discs and 10 μM AIP-I or AIP-II as indicated. Progress of the reaction was **(c)** analyzed by autoradiography and **(d)** plotted showing time courses of phosphorylation levels of GC214 (closed circles and dashed lines) and AgrA (open squares and solid lines). In **(d)**, time courses are colored according to the presence of AgrC-I and/or indicated AIP. Error bars = S.D. (n = 2). **(e)** Kinetics of AgrA chemical dephosphorylation: the AgrC-I-free reaction in **(c)** was repeated and samples were withdrawn at more time points for analysis. Phosphorylation levels of GC214 and AgrA-I were plotted as mean ± S.D. (n = 4). Data symbols and connecting lines are same as in **(d)**.

(Figure 3.6a, lanes 10-16 and Figure 3.6b, $t_{1/2}$ = 52 sec), indicating that the agonist AIP acts solely at the level of AgrC phosphorylation. Surprisingly, we observed that the presence of the inhibitor AIP-II slowed down the rate of transfer (Figure 3.6a, lanes 18-24 and Figure 3.6b, $t_{1/2}$ = 140 sec). Thus, this inverse-agonistic AIP can inhibit the TCS both at the level of HK auto-phosphorylation and phospho-relay to the RR. Whether similar inhibition of phospho-relay occurs in other inverse-agonistic AIP-AgrC interactions is currently under investigation.

Next we asked if AgrC possesses any phosphatase activity against phosphorylated AgrA, an activity common among RHKs (69). Since AgrA forms precipitate quickly following acetyl-phosphate treatment (data not shown), we employed an enzymatic approach to phosphorylate AgrA *in situ* using a soluble, highly active AgrC variant (GC214, see Chapter 4) as the phospho-donor. We then monitored the decay of [^{32}P]-phosphorylated AgrA in the presence or absence of AgrC-I disks. Surprisingly, we found no acceleration of AgrA dephosphorylation in the presence of AgrC-I disks at 20-fold excess to the GC214 phospho-donor (Figure 3.6c, compare lanes 1-6 to 7-12, and Figure 3.6d). Inclusion of either AIP-I or AIP-II along with AgrC-I discs also had no effect (Figure 3.6c, lanes 13-18, 19-24 and Figure 3.6d). Consequently, the decrease in AgrA phosphorylation levels is due to the self-catalyzed dephosphorylation of the RR, with a $t_{1/2}$ of 3.9 minutes (Figures 3.6c and e). We therefore conclude that AIP signals do not regulate *agr* signaling at the level of RR dephosphorylation.

3.4 Physicochemical parameters of the AIP-I-AgrC-I interaction

All *S. aureus* AIPs feature a 5-residue macrocycle closed by a thiolactone bond formed by condensation of the sulfhydryl group of a conserved cysteine and the peptide C-terminus (Figure 3.7a). It has been speculated that the intrinsic chemical reactivity of this thiolactone could lead to a covalent interaction, either irreversible or transient, with AgrC

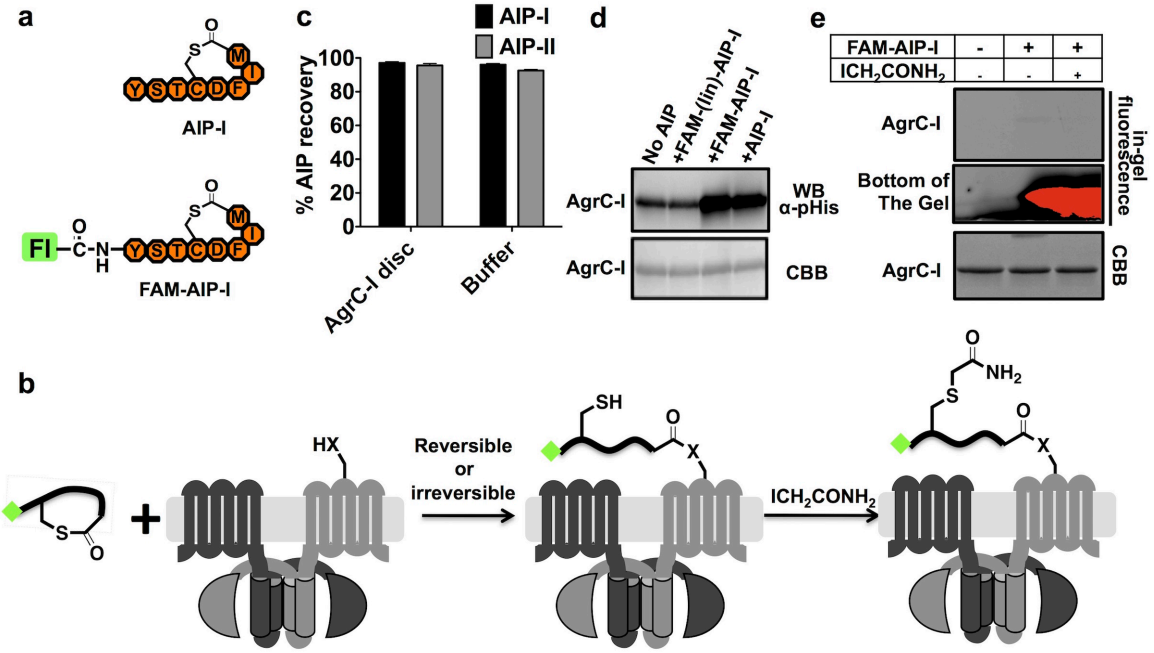


Figure 3.7 Ring-opening of the thiolactone is involved in AIP-AgrC interactions
(a) Schematic presentation of AIP-I and FAM-AIP-I. **(b)** Diagram showing the AgrC-mediated ring-opening hypothesis (see text for details) and the rationale to capture the covalent adduct employing iodoacetamide. **(c)** Stability test of AIPs in the presence of AgrC-I discs: AIP-I or AIP-II (2 nmol) were incubated in the presence or absence of AgrC-I discs (0.2 nmol) at 37°C. Percentage of intact AIP that remained after each treatment was quantified using HPLC and plotted as mean ± SD (n = 3). **(d)** AgrC-I discs were phosphorylated as described in Figure 3.2a with the indicated peptides (10 μM) for 3 minutes. The reactions were then analyzed by western blotting. **(e)** Attempt to capture the putative covalent adduct between FAM-AIP-I and AgrC-I: AgrC-I discs (2 μM) were incubated with vehicle (DMSO) or FAM-AIP-I (5 μM) at 37°C for 30min followed by treatment with vehicle (buffer) or 50 mM iodoacetamide at 37°C for 30min. The reaction mixture was analyzed by non-reducing SDS-PAGE and visualized by fluorescence (top and middle panels). The gel was then stained with CBB for loading control (bottom panel).

(see Figure 3.7b) (25). This was readily tested using our *in vitro* system. The free levels of intact AIP-I or AIP-II did not change after treatment with AgrC-I nanodiscs, indicating that receptor engagement does not lead to an irreversible adduct or to release of hydrolyzed peptide (Figure 3.7c). Conceivably, the covalent adduct could be transient and chemically reversible. To explore this, we exploited a fluorescein-labeled AIP-I analog (FAM-AIP-I) that has AgrC-I agonist activity indistinguishable from that of the native AIP-I (Figure 3.7a and d) (48). Incubation of FAM-AIP-I and AgrC-I nanodiscs in the presence of excess iodoacetamide, which would alkylate the newly generated cysteine within the linearized AIP and hence trap the putative adduct, failed to produce any fluorescently labeled AgrC-I as indicated by non-reducing SDS-PAGE (Figure 3.7b and e). We therefore conclude that the AIP-AgrC complex does not rely on the chemical reactivity of the thiolactone bond.

We next turned to the stoichiometry of the AIP-AgrC binding interaction. For this we titrated AIP into AgrC-I nanodiscs fixed at a concentration of 1.25 μM, about two orders of magnitude higher than the EC_{50} value for AIP-I (25, 48). Assuming that the EC_{50} value is a reasonable estimate of the K_D value (vide infra), the AIP-I added to the system should bind to AgrC-I in a linear fashion until saturation of binding sites is achieved. The auto-kinase activity of AgrC-I was used as a measure of the level of AIP-I binding. As predicted, the initial auto-kinase velocity of the system first increased linearly and then reached a plateau after the AIP-I concentration reached ~2 equivalents with respect to AgrC-I nanodiscs (Figure 3.8a). These observations strongly suggest that each AgrC-I dimer possesses two binding sites for AIP-I, and more intriguingly, that AIP-I binding at

one site causes the same level of auto-kinase activation regardless of the occupancy of the other site – *i.e.*, there is no evidence of cooperatives binding between the two subunits.

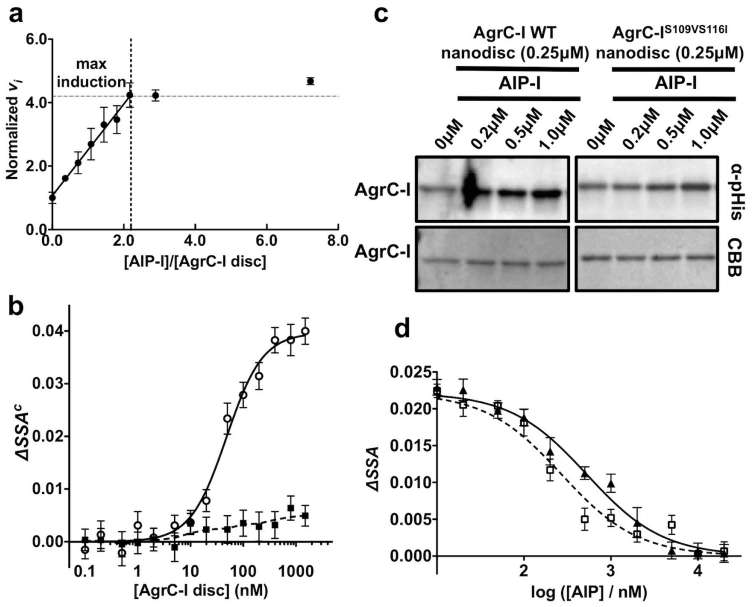


Figure 3.8 The stoichiometry and dissociation constants of the interaction between AgrC-I and AIPs

(a) Stoichiometry of the AgrC-I/AIP-I interaction. AgrC-I discs were incubated with 20 μM [γ-^{32}P]-ATP along with AIP-I at different concentrations and in each case the initial auto-kinase velocity (v_i) was determined by monitoring build-up of radio-labeled AgrC-I over time using autoradiography. Plot shows v_i (normalized to the AIP-free reaction) versus [AIP-I]/[AgrC-I disc] ratio. Error bars = SD (n = 3). **(b)** Equilibrium binding of FAM-AIP-I to AgrC-I discs. Corrected *s*teady-*s*tate *a*nisotropy change (ΔSSA^c, see Section 7.4.3) of FAM-AIP-I fluorescence was plotted as a function of added wild-type (open circle and solid line) or mutant (S109VS116I, closed circles and dashed line) AgrC-I discs and fit to a Hill equation. One representative titration of 4 is shown. Error bars (technical) = SEM (n = 6). **(c)** Auto-kinase activation of AgrC-I^{S109VS116I} dimers: 0.25 μM nanodiscs containing wild-type AgrC-I dimers (left four lanes) or AgrC-I^{S109VS116I} mutant dimers (right four lanes) were incubated with 1 mM ATP, 5 mM $MgCl_2$ and AIP-I at indicated concentrations at 37^{o}C for 3min. The reactions were then analyzed by anti-pHis western-blotting (top) and Coomassie blue staining (bottom). **(d)** Equilibrium binding of native AIPs to AgrC-I discs based on competitive displacement of FAM-AIP-I from AgrC-I. ΔSSA of FAM-AIP-I is plotted as a function of added AIP-I (open squares and the dashed curve) or AIP-II (closed triangles and the solid curve). The data are fit to a competitive binding model (see Section 7.4.3). One representative titration of 3 is shown. Error bars (technical) = SEM (n = 6).

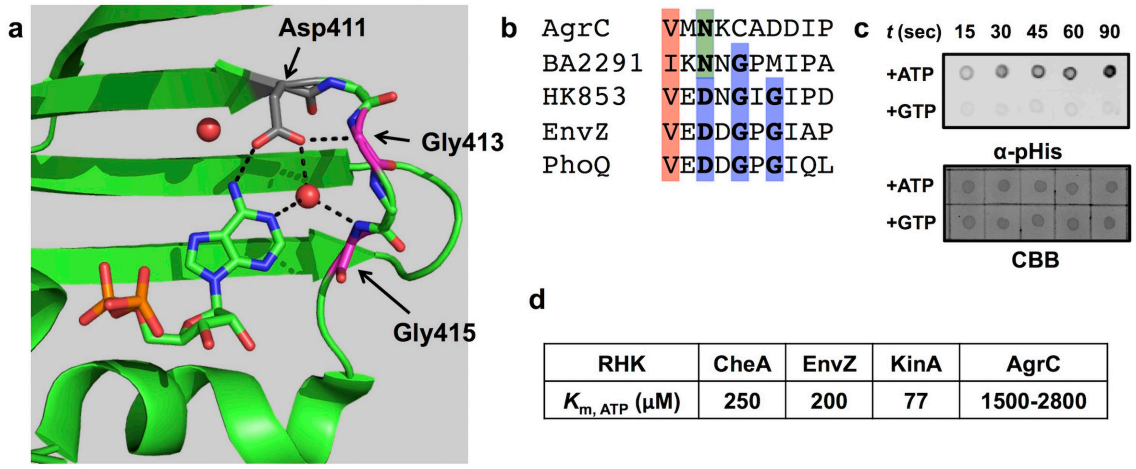
FAM-AIP-I also provided us with an opportunity to measure the equilibrium binding affinity of the AIP-AgrC interaction using a fluorescence-based assay. The fluorescence anisotropy of FAM-AIP-I was monitored as a function of added AgrC-I nanodiscs and the data fit to a transformed Hill equation (Figure 3.8b). The binding isotherm yielded a K_D for FAM-AIP-I binding to AgrC-I of 122 ± 26 nM (n = 4), with a Hill coefficient of 0.97 ± 0.20. Importantly, we observed no change in anisotropy when AIP-I was titrated with nanodiscs containing an AgrC-I^{S109VS116I} mutant that is defective in AIP-I-dependent activation (Figure 3.8b and c) (75). The lack of cooperation in the interaction between AIP-I and AgrC-I correlates well with the independent, but additive, role of the two binding sites in the auto-kinase activation of AgrC-I. The K_D for native AIP-I and AIP-II were 63 ± 13nM (n = 3) and 160 ± 20nM (n = 3), respectively, obtained from competitive titrations of a pre-formed complex between FAM-AIP-I and AgrC-I (Figure 3.8d). The K_D for AIP-I was 2.2-fold greater than its EC_{50} (28nM) obtained from reporter-gene assays, likely due to amplification effects in AgrC-mediated two-component signaling, which have been seen in other types of signal transduction pathways (48).

3.5 Nucleotide specificity and kinetics of AgrC-I auto-phosphorylation

Atomic-resolution structures of CA subdomains in complex with adenosine nucleotides reveal a role for the conserved ‘G1 box’ residues in nucleotide binding: the amine group of the adenine ring hydrogen bonds to the Asp residue, while the negative charge on the Asp is stabilized by interactions with the backbone amides of the two glycine residues (Figure 3.9a) (123). Sequence analysis described in Chapter 2 suggested that AgrC-I lacks these key residues and, therefore, the interaction between CA subdomain and ATP

might be altered (Figure 3.9b). Notably, the Asn residue aligned to the position of the conserved Asp is reminiscent of a GTP-dependent RHK, BA2291 from *Bacillus anthracis* (144). This notwithstanding, AgrC-I apparently favors ATP over GTP, as no substantial phosphorylation was observed upon incubation with the latter (Figure 3.9c).

To further investigate the influence of this non-canonical G1 motif on nucleotide binding, we determined the apparent K_m values for the cofactor in the presence or absence of AIP-I using dot-blot-based kinetic assays (Figure 3.10a). The first-order kinetic constant (k_1) for AgrC-I auto-phosphorylation was determined at a series of ATP concentrations and fit



RHK	CheA	EnvZ	KinA	AgrC
$K_{m,\ ATP}$ (μM)	250	200	77	1500-2800

Figure 3-9 Nucleotide specificity of AgrC-I
(a) Diagram depicting the adenine binding by conserved G1-box residues in the crystal structure of HK853 from *Thermotoga maritime* (PDB ID: 2c2a). The CA subdomain of HK853 is shown as ribbon diagram in green, with ADP and the conserved Asp411, Gly413 and Gly415 of HK853 highlighted by stick models. Two water molecules are shown in red spheres, and hydrogen bonds in black dashed lines. An α-helix stacking against the ADP nucleotide is omitted for clarity. **(b)** Sequence alignment showing the G1 box in *S. aureus* AgrC and *B. anthracis* BA2291 versus three RHKs carrying canonical G1-box sequences. Conserved Asp and Gly in canonical G1-boxes, Asn residues aligned to the conserved Asp and a hydrophobic position well-aligned in all sequences are highlighted in blue, green and red, respectively. **(c)** Nucleotide specificity of AgrC-I: AgrC-I discs were incubated with 1 mM ATP or GTP, 5 mM $MgCl_2$ and excess AIP-I peptide at 37°C. Aliquots were removed from each reaction at indicated time points, spotted in a row on the blotting membrane and analyzed for pHis levels through immuno-blotting. **(d)** Comparison of the K_m for ATP among RHKs. All shown RHKs except AgrC possess canonical G1-box sequences.

to the Michaelis-Menten kinetic model (Figure 3.10b). As expected, the presence of AIP-I stimulated receptor auto-phosphorylation, reflected in an ~8-fold increase in the

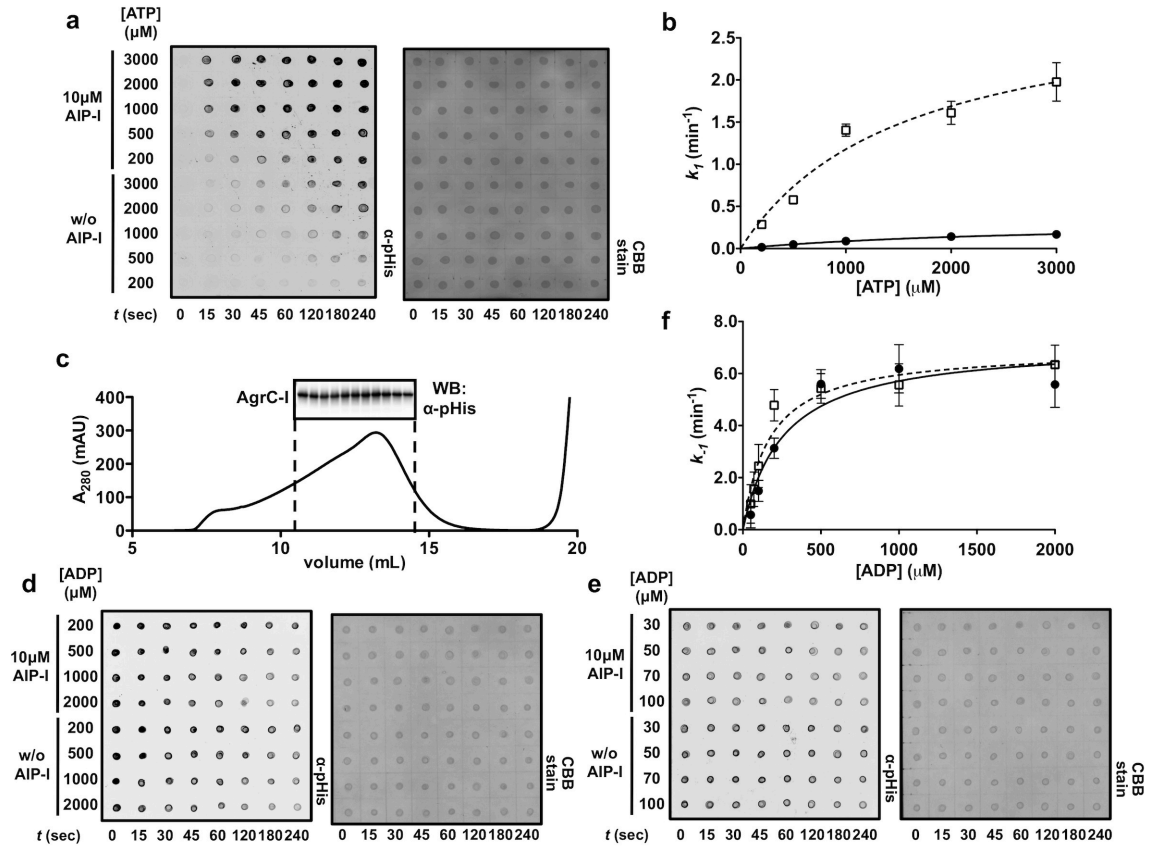


Figure 3.10 Kinetics of the forward and reverse auto-kinase reactions in AgrC-I
(a and **b)** Kinetics of the auto-kinase reaction: AgrC-I discs were incubated at 37 ^{o}C with ATP (at indicated concentration) and 5 mM $MgCl_2$ in the presence or absence of 10 μM AIP-I peptide. Aliquots were removed from each reaction at indicated time points and spotted in a row on the blotting membrane. pHis levels of AgrC-I in each reaction were **(a)** analyzed by anti-pHis immuno-blotting and **(b)** converted into first-order kinetic constants and plotted as a function of ATP concentration for reactions in the presence (open square, dashed curve) or absence (closed circles, solid curve) of 10 μM AIP-I. Error bar = SD (n = 3). In **(b)**, one representative membrane of three is shown. **(c)** Pre-phosphorylated AgrC-I discs analyzed by size-exclusion chromatography. Fractions collected between the dashed lines were further analyzed by western blotting with anti-pHis antibody (inset). **(d-f)** Kinetics of the reverse reaction: pre-phosphorylated AgrC-I discs were incubated at 37 ^{o}C with ADP (at indicated concentration) and 5 mM $MgCl_2$ in the presence or absence of 10 μM AIP-I peptide. Reactions were analyzed as in **(a)** and Michaelis-Menten plotting was performed as in **(b)**.

Table 3.1 Kinetic parameters of AgrC-I forward and reverse auto-kinase activities

Reactions	AIP-I (μM)	K_M (mM)	k_{max}(min^{-1})[a]
Forward	10	1.5 ± 0.5	3.0 ± 0.5
Forward	0	2.8 ± 0.4	0.33 ± 0.03
Reverse	10	0.17 ± 0.06	6.9 ± 0.6
Reverse	0	0.26 ± 0.07	7.2 ± 0.5

a: k_{max}, the maximal first-order kinetic constant.

maximum first-order rate constant, k_{max}, compared to the *apo* receptor (Table 3.1). The apparent K_m for ATP was 2.8 mM for *apo*-AgrC-I and 1.5 mM for AIP-I-bound AgrC-I, both at least an order of magnitude larger than a few well-characterized RHKs with normal G1 boxes (Figure 3.9d) (142, 145, 146). Therefore, sequence variation in the G1-box weakened the affinity between ATP and AgrC-I presumably due in part to the substitution of a neutral Asn residue for a stronger hydrogen-bond acceptor, Asp. Intriguingly, mutation of the G1-Asn back to Asp in AgrC-I abolished the auto-kinase activity (data not shown), suggesting that the AgrC ATP-binding pocket has co-evolved with this conserved Asn and is no longer able to adapt to its replacement by Asp.

A similar kinetic characterization was performed for the reverse auto-kinase reaction, in which phosphorylated receptor transfers the phosphoryl group back to ADP. AgrC-I discs were first phosphorylated in the absence of AIP and then separated from nucleotides using size-exclusion chromatography (Figure 3.10c). Using this protocol, 52.5% of the histidine sites were phosphorylated in the product. Kinetic constants for the reverse reaction were determined at a series of ADP concentrations in the presence or absence of AIP-I (Figures 3.10d and 3.10e). The reverse reaction was significantly more efficient: the K_m for ADP was lower while the k_{max} was higher than the forward reaction (Figure 3.10f and Table 3.1). This observation is consistent with the more favorable free energy

of hydrolysis of phosphohistidine compared to ATP (-13 kcal/mol vs. -7.3 kcal/mol) (141). More surprisingly, and in stark contrast to the forward reaction, the presence of AIP-I had no significant effect on the efficiency of the reverse reaction. Thus, AIP-I binding to AgrC-I shifts the equilibrium in the favor of the forward reaction.

3.6 Summary and discussions

In this Chapter, we have presented a comprehensive biochemical characterization of the AgrCA TCS. We found that AgrC-I dimers reconstituted into lipid nanodiscs possess a substantial level of basal activity. The cognate autoinducer non-cooperatively binds to AgrC-I in a 2:2 stoichiometry with nanomolar affinity, with individual binding events adding equally to stimulate the auto-kinase activity of the RHK. The pHis moiety on auto-phosphorylated AgrC-I is stable under physiological conditions, but rapidly transfers the phosphoryl group to AgrA. Remarkably, AgrC homologs from three *S. aureus* subgroups exhibit uneven auto-kinase activities in the activator-bound state. Meanwhile, AgrC lacks any detectable phosphatase activity, leaving self-catalyzed AgrA dephosphorylation the only reaction that counteracts the phospho-relay cascade. These biochemical properties are likely to play essential roles in controlling the onset of virulence in *S. aureus*. At low cell density, for instance, the basal auto-kinase activity of AgrC maintains a low level of phosphorylated AgrA that, due to its differential affinity to the regulatory elements in the P2 and P3 promoter regions, enables baseline transcription of the P2 operon rather than P3 (88). AIP, produced as a consequence, would accumulate to levels that significantly activate the AgrCA TCS during log-phase bacterial growth. The strong affinity of AIP for AgrC, the lack of phosphatase activity in AgrC, and the

positive feedback of the autoinduction loop collectively leads to a burst of RNAIII production and thereby an abrupt switch into a high-density, virulent mode of life. Since AgrA, and hence the kinetics of AgrA chemical dephosphorylation is conserved, the kinase activity conferred by the activated RHK is therefore most likely responsible for the difference in autoinduction timing among *S. aureus* subgroups (50). Finally, the $t_{1/2}$ of phosphorylated AgrA is a mere 3.9 minutes at 37^{o}C, shorter than most characterized transcription-factor RRs including the *E. coli* osmolarity RR, OmpR ($t_{1/2}$ ~ 70min), the *E. coli* cation-homeostasis RR, PhoB ($t_{1/2}$ ~ 15min) and the *B. subtilis* sporulation regulator Spo0A ($t_{1/2}$ ~ 7 min) (147). This enables *agr*-activated cells to rapidly turn off virulon expression and revert to vegetative growth if separated from the quorum.

AgrC-I is also distinguished from most other characterized RHKs by its dramatically lowered affinity for adenosine nucleotides, likely due to substitutions in the G1-box. We reason that this atypically low affinity might allow AgrC to sense cellular ATP levels, which vary significantly between optimal and stress conditions. Indeed, down regulation of accessory gene expression in times of nutrient or oxygen deprivation had been reported (29). Since the cellular ATP concentration data is not available for *S. aureus*, the values for *E. coli* must be used as a reference (148). With abundant carbon source, *E. coli* contains approximately 3.5 mM cellular ATP, which decreases to sub-milimolar levels at stationary phase, or after switching to a weaker electron acceptor for respiration. Assuming similar variations in *S. aureus*, a drastic decrease in the auto-kinase activity would be expected for AgrC (apparent K_m for AgrC-I ~2 mM) under nutrient-limiting or anaerobic conditions. Thus, we propose that inactivation of the virulon under stress

conditions is at least partly a consequence of the acute sensitivity of AgrC to falling ATP levels, ultimately leading to a down-regulation of the *agr* response. This sensitivity may extend to processes mediated by other receptors of the HPK10 subfamily given the conserved nature of the altered G1-box.

During the course of our reconstitution studies, a remarkable dependence on anionic bilayer environment was found for AgrC-I activation. The exact basis of this lipid-sensitivity in AgrC-I activation remains unclear. We did, however, rule out the possibility of global misfolding for AgrC-I embedded in 100% DMPC nanodiscs in that it forms dimers that bind to a fluorescent AIP-I analog, albeit without any resulting auto-kinase stimulation (Figure 3.2a and c). Given the highly basic amino acid sequence of the AgrC-I sensor domain (theoretical pI = 9.3), we speculate that anionic lipids enable competent conformational coupling between the AIP-binding site and AgrC-I^{HK} by neutralizing excessive positive charges in the sensor domain. Intriguingly, the strongly anionic cell membrane found in *S. aureus* is shared among other species from the low-GC Gram-positive phylum (*Firmicutes*), and distinguishes them from those representing major Gram-negative phyla (149). In addition, our bioinformatic analysis revealed a striking restriction of the HPK10 subfamily, to which AgrC belongs, to the phylum Firmicutes. All characterized RHKs from this subfamily are receptors of peptidic pheromones, and, perhaps not coincidentally, peptide-mediated QS is not known in any Gram negative species. A plausible explanation for this taxonomic boundary is that HPK10 subfamily receptors are all evolved in, and presumably well adapted to, the anionic lipid environment. Such adaptation, if generally present, would create a considerable barrier to

horizontal gene transfer of HPK10 homologs into Gram-negative species, which have PE-rich inner membranes. Biochemical studies into additional HPK10 homologs, perhaps using the nanodisc technology employed herein, might lend further support to the idea that peptide-responsive RHKs are restricted to Gram-negative bacteria because of the lipid composition of their membranes.

Chapter Four: Mechanisms by which AgrC-I Responds to Signals

Quorum-sensing machineries in bacterial pathogens have been long regarded as potential therapeutic targets and, consequently, previous drug-development campaigns provide a broad spectrum of small-molecule ligands for several QS-mediating RHKs including AgrC-I (27). Among all these RHKs, *S. aureus* AgrC-I stands out for its versatility in ligand response (48). In line with the observations made from reporter strains, using the *in vitro* reconstitution system described in Chapter 3, distinct levels of kinase activity have been observed for AgrC-I in at least four ligand states (Figure 3.4a), contradicting the common on/off switch view for RHKs (150). To understand this signaling flexibility of AgrC, two pieces of information are needed, namely (i) the conformational change of the AgrC-I HK domain induced by AIP binding and (ii) the response of the AgrC-I HK activities to diverse conformational changes. Importantly, ligand-induced molecular motions have never been detected for any membrane-bound RHKs prior to our study. In this chapter, we first systematically perturb the conformation of the AgrC-I HK domain, employing a protein-chimera strategy, in order to chart a conformation-activity landscape of this module. The nanodisc reconstitution system was then harnessed for the detection of molecular motions triggered by AIP binding.

Constitutive mutants of AgrC-I have been identified previously employing a library-screening approach (143). These mutants, although all capable of activating P3-driven gene expression in reporter strains in the absence of AIP-I, vary dramatically with their signaling behavior. Some mutants, for instance, do not respond to any AIP added, while others are often revertible by AIP-II but not AIP-III. Interestingly, mutants in the first

category all bear mutations in the HK domain, while those in the second may carry mutations in either the sensor or the HK domain. In this chapter, we sought to understand (i) how a constitutive mutation in the HK domain, R238H, alters the conformation-activity landscape and (ii) why AIP-II is able to revert constitutive activity of a subset of mutants. Besides explaining some interesting *in vivo* phenotypes, the diverse *in vitro* behaviors of these mutants revealed in this work expand our view on the evolvability of the signal-response properties of AgrC-I.

4.1 Perturbing the conformation of the AgrC-I HK domain

Structural and functional studies on isolated HK domains have given rise to models of receptor activation involving rotation of helices that comprise the DHp subdomain (Figure 4.1a) (123, 150). These helical movements are thought to modify packing of the DHp leading to a repositioning of the CA subdomain such that phosphorylation of the His is, or is not, permitted on the basis of spatial proximity (123). This "helix-rotation" model entails a relatively rigid linkage between the DHp subdomain and the preceding sensor or transmission domains to enable effective propagation of conformational changes generated by the sensor. Interestingly, in crystal structures obtained for dimeric RHKs, the linkers between the DHp subdomain and the preceding module predominantly form a pair of α-helices, extending the first α-helix of each DHp subdomain (DHp-α1) (123, 151-153). The possibility of the AgrC-I interdomain linker (residues 201-221) being helical is particularly attractive because it would allow the formation of a contiguous helix spanning the last TM domain, the linker and the DHp-α1, providing a ready conduit for the conformational change in the sensor being transmitted to the HK domain (Figure

4.1b). This idea is supported by or analysis of the amino-acid composition of the interdomain linker sequence across the HPK10 subfamily (Figure 4.1c). Helical-breaking residues, namely glycine and proline, are strongly underrepresented in this region compared to their average frequency in proteins (154), suggesting a strong propensity of

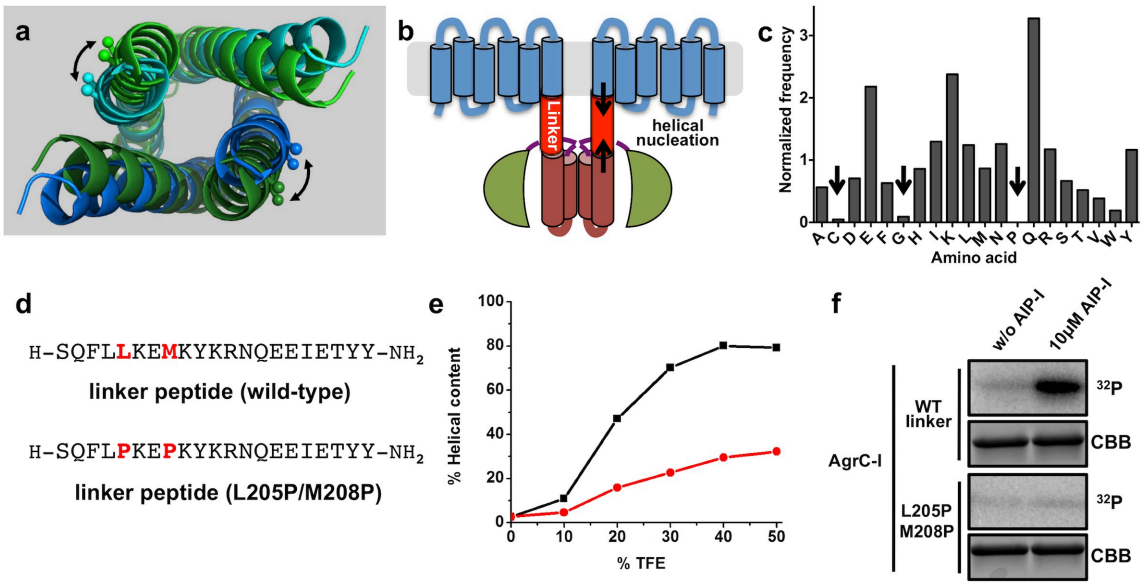


Figure 4.1 An α-helical linker connecting the AgrC-I sensor and HK domains.
(a) Structural alignment of the DHp subdomain in the *T. maritima* RHK, HK853, captured at the resting (green) and phosphatase-active (blue) states. Cβ atoms on residues 249 and 253 in the DHp-α1 are highlighted with spheres to help visualize the structural rearrangement of the four-helix bundle. **(b)** Cartoon of full-length AgrC-I showing α-helical interdomain linkers (in red). Helical nucleation that would stabilize the linker helix is highlighted with arrows. **(c)** Amino acid frequency in the interdomain linkers within HPK10 sequences. Frequency of each amino acid is normalized to its overall composition in proteins. Under-represented amino acids are indicated with arrows. **(d)** Sequence of the wild-type linker peptide (upper) and the L205PM208P mutant peptide (lower). Mutated residues are highlighted in red. **(e)** Helical content of linker peptides: plot shows calculated helical content as a function of % TFE for the wild-type (black) and L205P/M208P mutant (red) peptides. CD spectroscopy was performed at 20°C. The 15-μM peptide samples were prepared in buffers containing 10 mM phosphate, pH = 7.0 and 50 mM NaF and the indicated percentage of TFE (v/v). **(f)** Auto-kinase activation of AgrC-I$^{L205P/M208P}$ mutant: nanodiscs containing wild-type (WT) or mutant AgrC-I dimers were phosphorylated with [γ-^{32}P]-ATP and analyzed by autoradiography. CBB indicates the Coomassie-stained loading control.

α-helix formation in this region (Figure 4.1c). Furthermore, we synthesized two peptides corresponding to the AgrC-I linker, one with the wild-type sequence and the other with two residues, Leu 205 and Met 208, mutated to Pro (Figure 4.1d). Analysis of these peptides by circular dichroism (CD) spectroscopy revealed, in the presence of 20% (v/v) 2,2,2-trifluroethanol (TFE), high helical propensity in the wild-type peptide, which is greatly attenuated in the case of the mutant (Figure 4.1e). No helicity was found in the complete absence of TFE, suggesting that packing between two (or more) linkers is not strong enough to stabilize the helical conformation as in classical coiled coil peptides. Notably, this finding does not conflict with the linkers being helical in full-length AgrC-I in which its helicity could be further promoted by the presence of flanking helical elements through helical propagation (155). The L205P/M208P mutations that diminished helicity of the linker peptide abolished the AIP-I-dependent kinase activation in full-length AgrC-I, further underscoring the importance of the helical continuity in this region (Figure 4.1f).

Given the dimeric configuration of the AgrC-I HK domain and the helical propensity of the interdomain linker, we hypothesized that fusing the linker to a GCN4 peptide, which adopts a dimeric, parallel coiled-coil conformation, would lend control to the rotational conformation of the interdomain linker pair (Figure 4.2a). The GCN4 dimer is extremely stable (K_D = 0.28 nM) (156), which we expected to provide the rigid body needed to impose twisting force on the appended linker region, presumably driving the conformation away from the resting state. We prepared 15 chimera proteins, each comprising an N-terminal GCN4 coiled-coil sequence and a C-terminal AgrC-I HK

domain separated by an intervening interdomain linker segment of variable length. These protein chimeras were termed GC205 to GC219 after *G*CN4-Agr*C*-I and the position of the first AgrC-I residue fused to GCN4, to which we refer as the junction residue. Since the C-terminal residue of the GCN4 portion registers at the 'd' position in the classic

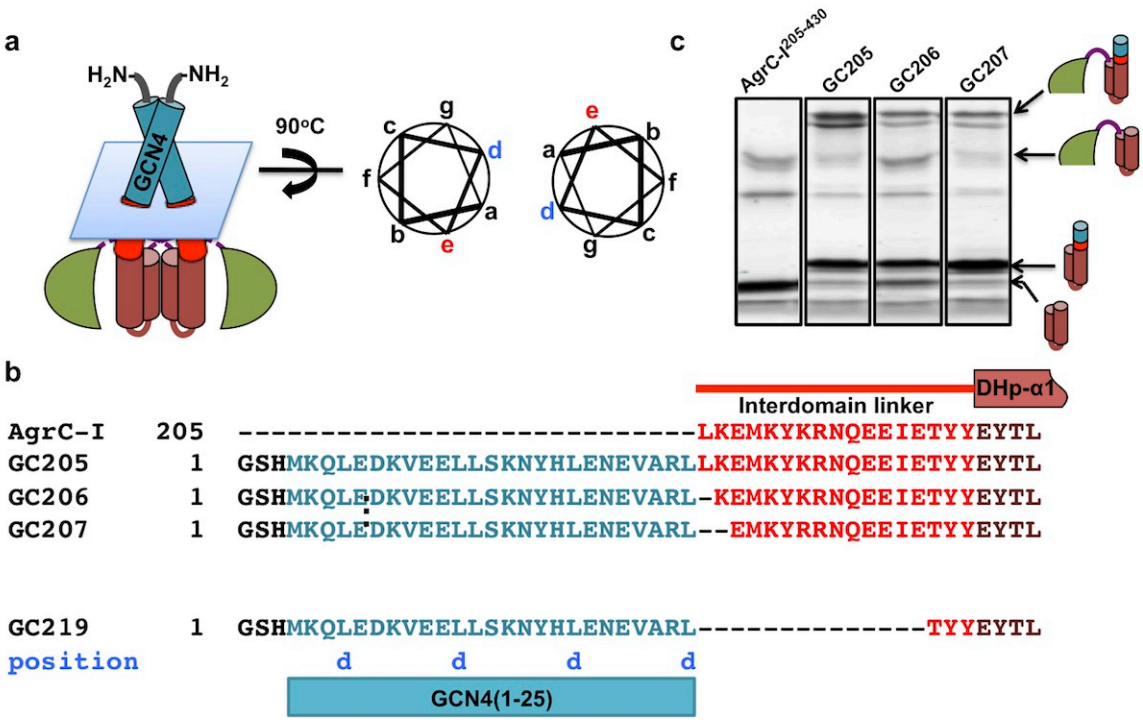


Figure 4.2 Design principles of the GCN4-AgrC-I chimera series
(a) Cartoon of a GCN4-AgrC-I chimera with the interface between the GCN4 (in cyan) and the interdomain linker (in red) indicated by a blue plane. The DHp and CA subdomains are colored brown and green, respectively. Right panel, coiled-coil helical wheel representation of the interface highlighting the register of the last residue of GCN4 ('d', in blue) and the junction residue ('e', in red). **(b)** Design of GCN4-AgrC-I chimera proteins. The 25-aa GCN4 coiled-coil was sequentially fused to residues in the interdomain linker. GCN4 residues registering at coiled-coil position 'd' are noted underneath the sequence in blue. **(c)** Limited trypsinolysis of AgrC-$I^{205-430}$ and three chimera proteins: substrate proteins (4 mg/mL) were incubated with trypsin (60 μg/mL) in ice for 20min. The reactions were stopped using protease inhibitors and resolved on SDS-PAGE. Tryptic fragments either including the peptide bond next to Arg212 or generated from the cleavage at this site are identified employing HPLC-MS (now shown). Corresponding gel bands of these fragments are recognized based on their molecular weight and abundance, and highlighted with cartoons and arrows.

coiled-coil helical wheel, we inferred that the junction residue in each chimera should be moved to the 'e' position (Figure 4-2b). Because the resting register of the linker residues in full-length AgrC-I follows the α-helical periodicity, the magnitude that these chimeras twist the linkers away from their resting conformation should exhibit the same periodicity with respect to the linker length between the GCN4 coiled coil and the AgrC-I HK domain.

Structural integrity of the interdomain-linker in these chimera proteins was preliminarily investigated employing limited trypsinolysis. The peptide bond next to Arg212 in the linker is highly susceptible to trypsin cleavage in an intracellular AgrC-I construct, AgrC-$I^{205\text{-}430}$ (Figure 4-2c). This result is reminiscent of the lack of helicity in the standalone linker peptide, as was revealed in the CD-spectrometry experiment in the absence of TFE. This peptide bond gains resistance against trypsin cleavage, however, in GC205, GC206 and GC207 (Figure 4-2c), suggesting a more defined linker structure in these chimera proteins. It is therefore highly probable that the interdomain linker adopts stable helical conformations only when sandwiched by two other helical elements: this requirement is fulfilled in both full-length AgrC and GCN4-AgrC-I chimeras, but not in AgrC-$I^{205\text{-}430}$.

4.2 Conformation-activity landscape of the AgrC-I HK domain and the molecular motion induced in full-length AgrC-I by AIP-I or AIP-II binding

All 15 GCN4-AgrC-I chimeras were expressed in *E. coli* and purified for auto-kinase assays. Strikingly, the phosphorylation levels of these GCN4-AgrC-I fusion proteins

fluctuated periodically with the length of the linker (Figure 4.3a). The periodicity, obtained from fitting the data to a sine wave function, was 3.6, constituting one turn of an α-helix (Figure 4.3b). Because the magnitude of twisting applied to the linkers is also expected to follow the same periodicity, this result directly correlates the auto-kinase

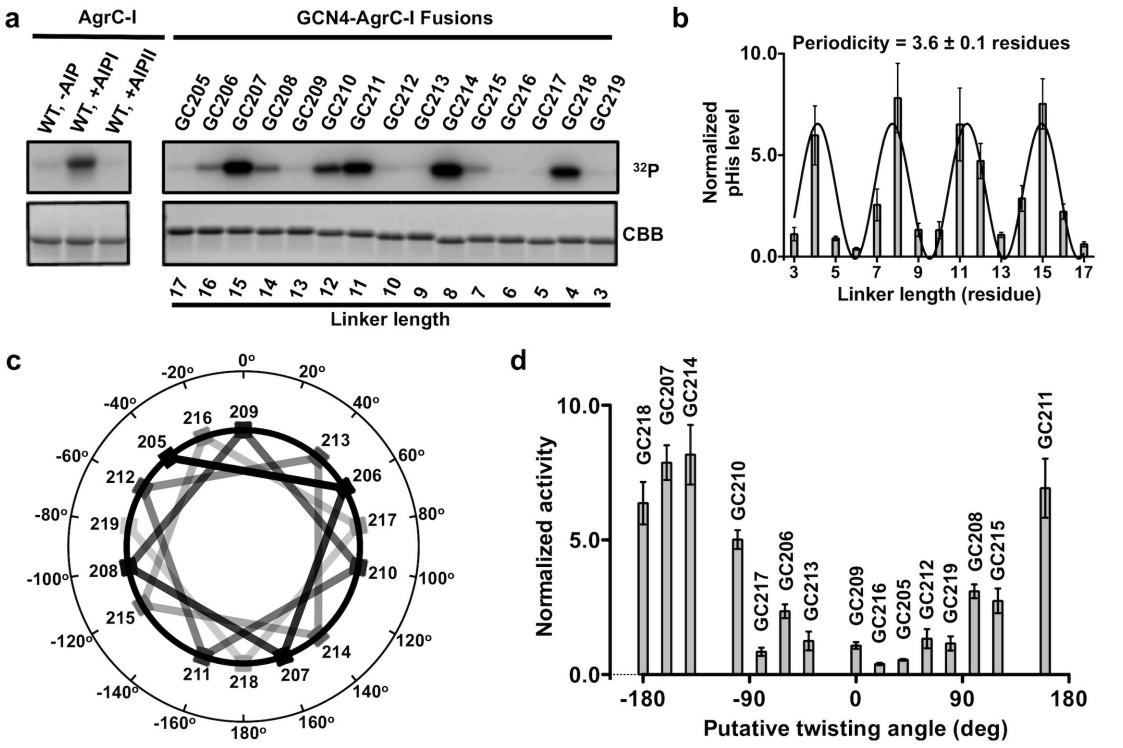


Figure 4.3 Twisting of the interdomain linkers exercises rheostat-like control over the auto-kinase activity of the AgrC-I HK domain.
(a) Auto-kinase activities of GCN4-AgrC-I chimeras: dimeric chimeras and AgrC-I discs (with or without added AIP as indicated) were treated with [γ-^{32}P]-ATP and analyzed by autoradiography as in Figure 4.1f. Linker length (in residues) is indicated underneath the loading control. **(b)** Quantification of the data in panel **(a)**. Plot shows normalized phosphorylation levels of the chimeras (normalized to that of the apo-AgrC-I discs) as a function of the linker length. Curve is a sinusoidal function fit to the data. Error bars = SD (n = 3). **(c)** Classical helical wheel showing the angle (θ) by which each position on the linker helix deviates from position 209. Since the linker conformation in GC209 resembles that of AgrC-I in the resting state (refer to Figure 4.4d and e), the putative twisting angle imposed to the linker helices in each GCN4-AgrC-I chimera should be the inverse number of θ for its junction residue position. **(d)** The conformation-activity landscape of the AgrC-I HK domain: Plot shows activity of all 15 GCN4-AgrC-I chimeras versus the putative twisting angle. Error bar = S.D. (n = 3).

activity to the twisting movement of the linkers. Consequently, a conformation-activity landscape is obtained from plotting the auto-kinase activity of each chimera with the putative twisting angle of the linkers away from their resting conformation represented by that in GC209 (Figure 4.3c and d, see below for the reason why GC209 is the reference). This landscape is remarkably smooth, allowing the rheostat-like control of the kinase activity by gradually twisting the linker helices. Chimeras with activity equivalent to that of AgrC-I in all three ligand states were found (Figure 4.3a), suggesting that the ligand-dependent differential activities of the full-length receptor might also be achieved through adjusting the linker conformation.

A Cys crosslinking strategy was used to study the conformational changes in the linker region of full-length AgrC-I that accompany AIP binding. Based on the helical twisting model that emerged from the GCN4 chimera data, we hypothesized that, depending on the location of the Cys and the ligand state of the receptor, inter-subunit disulfide bonds would be more or less likely to form under oxidizing conditions. The absence of native Cys residues in the linker region facilitated this approach (Figure 4.1d). Five AgrC-I mutants with Cys replacements at residues 205-209 were expressed, purified and incorporated into nanodiscs. The mutants behaved similarly to wild-type AgrC-I in terms of their auto-kinase response to agonist and antagonist AIPs (Figure 4.4a). Non-reducing SDS-PAGE analysis of the mutants following oxidation with glutathione revealed a marked difference in the amount of covalent dimer formed depending on the mutant and the presence or absence of AIP (Figure 4.4b). The yield of covalent dimer from each sample was normalized for loading, and grouped according to three activity states (*apo*,

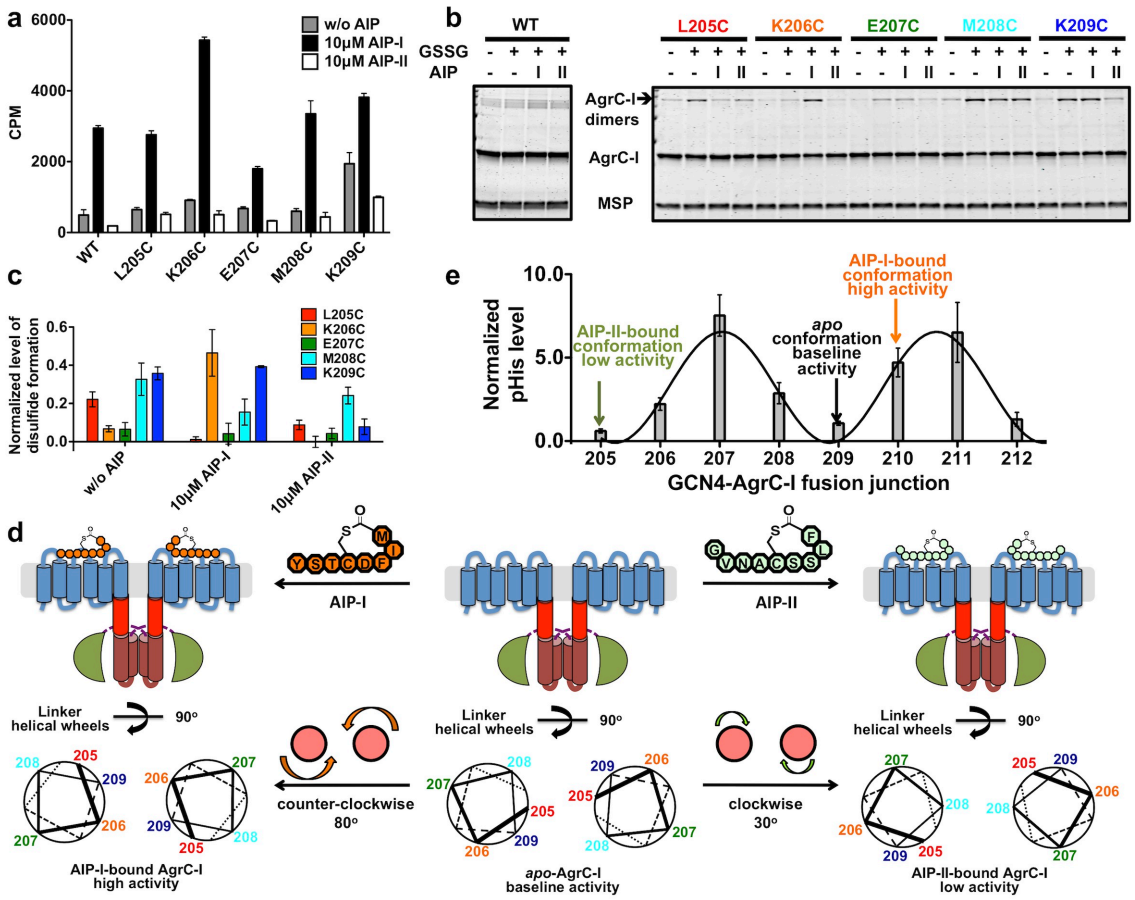


Figure 4.4. AIP binding induces rotational movement of the AgrC-I linker helices
(a) Auto-kinase activities of AgrC-I single-point Cys mutants. Nanodiscs containing wild-type or Cys-mutant AgrC-I dimers were incubated with 20 μM [γ-^{32}P]-ATP, 5 mM $MgCl_2$ and excess AIP peptide (as indicated) at 37°C for 40min. ^{32}P-labeled AgrC-I in each reaction was quantified through scintillation counting. Error bar = SD (n = 3). **(b and c)** inter-subunit crosslinking of AgrC-I single-point Cys mutants: nanodiscs incorporated with wild-type or mutant AgrC-I dimers were mock-treated or incubated with 10 mM oxidized glutathione (GSSG) along with indicated AIPs. **(b)** Formation of covalent dimers as analyzed by SDS-PAGE with Coomassie stain. MSP, membrane scaffold protein. **(c)** Quantification of the crosslinking results. Plot shows levels of disulfide-linked dimer formation normalized to loading, grouped according to the three ligand states. Error bars = SD (n = 3). **(d)** Rotation of the interdomain linker helices in AgrC-I induced by AIP binding. Top: Cartoons of AgrC-I dimer in the apo- (middle), AIP-I-bound (left) and AIP-II-bound (right) states. Bottom: The corresponding conformation of the linker helices is depicted with helical wheels showing the positions 205-209 according to the Cys crosslinking results. Ligand binding and the consequent rotation of linker helices are highlighted between panels. **(e)** Auto-kinase activities of chimera proteins in which interdomain linker conformation mimics that of AgrC-I in three ligand states: plot shows normalized phosphorylation levels of GC205 through GC212 as a function of junction residue positions (with AgrC-I numbering). Linker-conformation surrogate of each AgrC-I activity state is marked with an arrow.

AIP-I-bound or AIP-II-bound) as in Figure 4.4c. This revealed a periodicity within each state that, analogous to the GCN4 chimera data, could be interpreted by assigning parallel, juxtaposed helical structure to the linkers (Figure 4.4d). When analyzed in this way, the linker helices within the *apo*-AgrC-I dimer are oriented such that positions 205, 208 and 209 are juxtaposed and able to form crosslinks when Cys residues are placed at these sites. Binding of the AIP-I activator leads to a ~80° counter-clockwise rotation of the two linker helices in the dimer such that residues 206 and 209 are now in optimal position to form crosslinks. Conversely, binding of the AIP-II inhibitor leads to a ~30° clockwise rotation of the helices from the resting state resulting in residue 208 being optimal for crosslinking.

This interpretation of the crosslinking data was used to further analyze the differential activity associated with GCN4-chimeras. As noted above, the junction residue in each chimera can be assigned to the 'e' position of the coiled-coil helical wheel. Based on this logic, we predict that GC209, GC210 and GC205 should be conformational surrogates of the *apo*-, AIP-I- and AIP-II-bound forms of AgrC-I, respectively (Figure 4.4d, refer to Figure 4.3b for the register of position 'e'). Remarkably, these chimeras accurately phenocopy the three activity states of AgrC-I: compared to GC209, GC210 had much higher auto-kinase activity while GC205 was not as active (Figure 4.4e). We therefore conclude that ligand binding regulates the auto-kinase activity of AgrC-I by inducing symmetric rotation in the helical TMH-DHp linker region.

4.3 A constitutive mutation, R238H, modifies the conformation-activity landscape.

AgrC-I^{R238H} is the most prominent and robust constitutive mutant among all such mutants identified for AgrC-I. It conveys full activation to the P3-driven gene expression in the absence of any ligand and, therefore, has been employed in a few studies as a surrogate of wild-type AgrC-I in the AIP-I-bound state. The R238H mutation, however, occurs within the DHp subdomain, suggesting a modulation of the input-response property of the HK domain rather than the conformational states of the sensor. Perhaps as a consequence, the activity of P3-driven reporter genes in *S. aureus* expressing this mutant is unaffected by the presence of AIP-I or AIP-II. Being aware of the non-linearity of the reporter-gene readout, we decided to further investigate the effect of both AIPs on the kinase activity of nanodisc-embedded AgrC-I^{R238H}.

AgrC-I^{R238H} was expressed and purified from *E. coli* and reconstituted to nanodiscs in which the lipid bilayer consists of POPC and POPG at a molar ratio of 1:3. Wild-type AgrC-I discs of the same lipid composition were used for back-to-back comparison in the auto-kinase assay. AgrC-I^{R238H} had a strong baseline activity, which elevated only moderately when the RHK binds to the activator AIP-I (Figures 4.5a and b). By contrast, the wild-type RHK, although exhibiting a much lower baseline activity in the *apo* state, was turned on by AIP-I-binding so potently that its activity surpassed that of the mutant in the activated state (Figures 4.5a and b). This mutation also diminished the inhibitory effect of AIP-II on the auto-kinase activity (Figures 4.5a and b). Consistent with the unchanged sensor domain in this mutant, binding affinities of AIP peptides are largely unaffected. The fluorophore-tagged AIP, for instance, binds to AgrC-I^{R238H} discs with

slightly stronger affinity than to the wild-type (51 vs 122 nM), as was determined in the titration experiment employing a fluorescent-anisotropy readout (Figure 4.5c). The presence of excess AIP-I or AIP-II effectively displaces FAM-AIP-I from its binding site on the mutant (Figure 4.5d). We surmise, based on these observations, that this mutation renders the AgrC-I HK domain less sensitive to the conformational input from the sensor domains, rather than elevating its kinase activity at all conformational states.

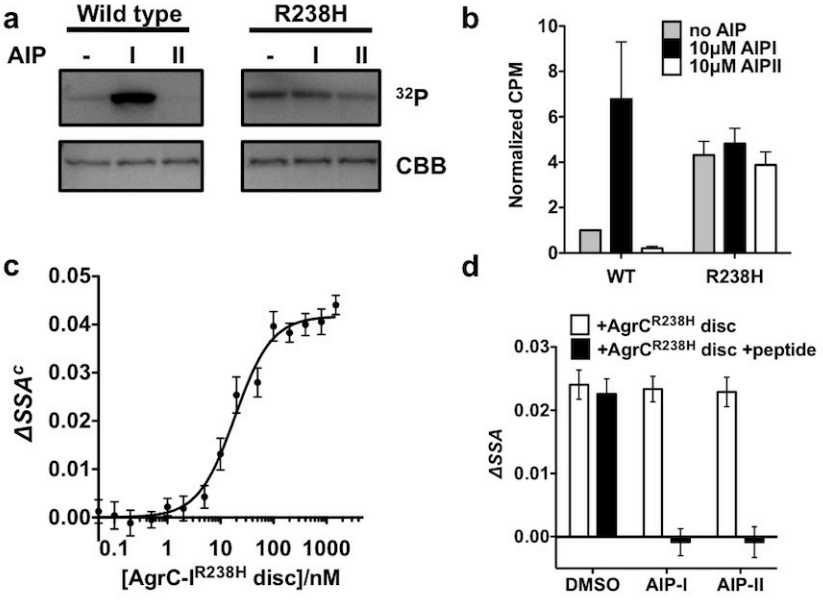


Figure 4.5 Auto-kinase activity and ligand responses of AgrC-I^{R238H}
(a and **b)** Auto-kinase assays comparing the R238H mutant to the wild type: the nanodisc-incorporated receptors were pre-incubated with indicated AIP-peptides or the vehicle (DMSO), and then phosphorylated with 20 μM [γ-^{32}P] ATP at 37 ^{o}C for 40 minutes. pHis levels were analyzed by autoradiography in **(a)** or quantified using scintillation counting in **(b)**. In **(b)**, the count per minute (CPM) values were normalized to that of the wild-type AgrC-I in the absence of AIP. Error bars = S.D. (n = 3) **(c)** Equilibrium binding of FAM-AIP-I to AgrC-I^{R238H} discs. As in Figure 3.8b, corrected *s*teady-*s*tate *a*nisotropy change (*ΔSSA*^{c}, see Section 7.4.3) of FAM-AIP-I fluorescence was plotted as a function of added AgrC-I^{R238H} discs and fit to a Hill equation. One representative titration of 3 is shown. Error bars (technical) = SEM (n = 6). **(d)** Dissociation of FAM-AIP-I from AgrC-I^{R238H} discs as a result of competition with native AIPs: bar graph shows the *ΔSSA* of 20 nM FAM-AIP-I upon forming complex with 150 nM AgrC-I^{R238H} discs alone (open bars) or in the presence of 10 μM indicated AIP competitors (closed bars). Errors bars = S.D. (n = 3).

To further substantiate this idea, we employed the chimera protein strategy introduced in section **4.2** to chart the conformation-activity landscape for this mutant HK domain. Eleven GCN4-AgrC-I^{R238H} chimeras were prepared and their auto-kinase activities were compared to the corresponding GCN4-AgrC-I constructs (Figure 4.6a-d). Remarkably, activities of the mutant chimera series also changed periodically with the length of the interdomain linker (Figure 4.6a and b). Fitting the activity-linker length relationship to a sine wave function revealed that the fluctuation amplitude in the mutant chimera series is less than 30% of that observed in the wild-type series (compare Figure 4.6b to d), confirming the notion that the mutant HK domain is less responsive to conformational perturbations. Other parameters of the two waves are strikingly similar: both followed the α-helical "wavelength" of 3.6 residues, while no phase shift was observed between them. The "baseline" of the two waves, defined as the mean of the maxima and minima of the fluctuation, is also comparable between the two series. Previous experiments on the GCN4-AgrC-I chimera series identified GC205, GC209 and GC210 as the conformational surrogate of the AIP-II-bound, *apo* and AIP-I-bound AgrC-I and phenocopy their activity levels (Figure 4.6e). These relationships hold true in the mutant chimera series, as GC205^{R238H}, GC209^{R238H} and GC210^{R238H} accurately recapitulated the strong activity of AgrC-I^{R238H} in the *apo* state and the weak effect that AIP-I and AIP-II have on this system (Figure 4.6e). Plotting the auto-kinase activity of each GCN4-AgrC-I^{R238H} chimera as in Figure 4.3d revealed the conformation-activity landscape for the mutant HK domain (Figure 4.6f). Notably, the activity minimum of this landscape is much higher than that of the wild-type series, and this level of kinase activity probably affords full activation of the P3 promoter in the reporter-strain settings.

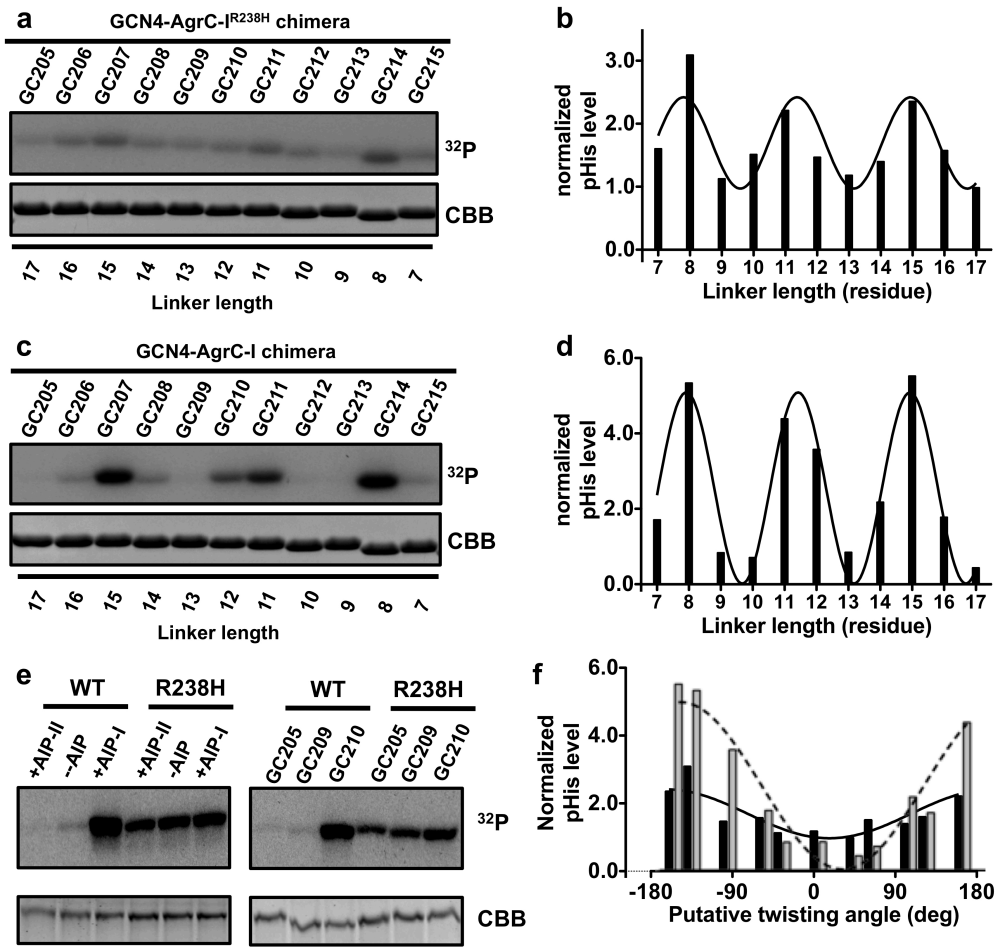


Figure 4.6 The R238H mutation renders the AgrC-I HK domain less responsive to the twisting movement of linker helices.

(a and **b)** Auto-kinase assay of 11 GCN4-AgrC-I chimeras bearing the mutant AgrC-I HK domain. An autoradiogram with the corresponding Coomassie-stained loading control **(a)** and an activity-linker length plot **(b)** are shown. In **(b)**, curve shows the sinusoidal function fit to the data. **(c** and **d)** Auto-kinase assay data and the quantification result for the chimera series carrying the wild-type AgrC-I HK domain. Note that the panels in **(a)** and **(c)** were scanned from the same film (for autoradiograms) or stained in the same gel (for loading controls). In **(b)** and **(d)**, intensity of each band on the autoradiogram was normalized to that of GC209. **(e)** Auto-kinase activity of the linker-conformation surrogates (refer to Figure 4.4 d and e) versus the full-length receptors. **(f)** Comparison of the conformation-activity landscape between the wild-type and mutant AgrC-I HK domains. Putative twisting angles are calculated according to Figures 4.4e. Normalized auto-kinase activity of the chimera series bearing wild-type (gray bars) and R238H mutant (black bars) HK domains are plotted. Fitting to a sinusoidal function returns smoothened landscapes for the wild-type (dashed curve) and R238H (solid curve) HK domain. For clarity, the wild-type data (gray bars) were shifted to the right by 10^{o} along the horizontal axis. In panels **(a)**, **(c)** and **(e)**, auto-phosphorylation was performed employing [γ-^{32}P] ATP and pHis levels analyzed by autoradiography.

We next asked whether an intracellular construct, AgrC-I$^{205-430}$ bearing the R238H mutation would convey the high level of auto-kinase activity expected for the full-length mutant. To our surprise, this construct gains very little activity upon the R238H mutation (Figure 4.7a). Suspecting that the sensor domain might be necessary for the auto-kinase activation by this mutation, we truncated AgrC-I to leave the C-terminal four, two or one TM domain(s). All truncated AgrC-I mutants, after nanodisc reconstitution, exhibited activities as strong as the full-length mutant (Figure 4.7a), suggesting the intact sensor domains and the inter-subunit interaction mediated thereby are not required for the gain of activity conferred by this mutation. Apart from bridging specific interactions, other possible roles of TM domains in organizing the nanodisc-incorporated histidine kinase include (i) anchoring the soluble portion of AgrC-I in proximity to the lipid bilayer and (ii) propagating helicity to the interdomain linker to which it directly connects. The later role is analogous to the GCN4-dependent stabilization of the linker helices in the chimera proteins (see Figure 4.4e). To dissect between these two roles, we installed a pair of phospholipid anchors to a dimeric AgrC-I soluble-domain (AgrC-I$^{208-430}$) at the N-terminus of both subunits (Figure 4.7b and c). Reconstitution of this protein-lipid conjugate to nanodiscs successfully attached the soluble AgrC-I construct to the lipid bilayer (Figure 4.7d). Unexpectedly, this attachment *via* lipid moieties, as opposed to TM-domain anchors, was detrimental to the auto-kinase activity of both the wild-type and mutant AgrC-I$^{208-430}$ (Figure 4.7e). On the other hand, kinase activation by the R238H mutation is tolerant to the substitution of the TM sequence in the single-TM AgrC-I construct with an unrelated TM domain from an *E. coli* protein (LepB) or the insertion of a two-aa spacer to this LepB-AgrC-I chimera between the TM domain and

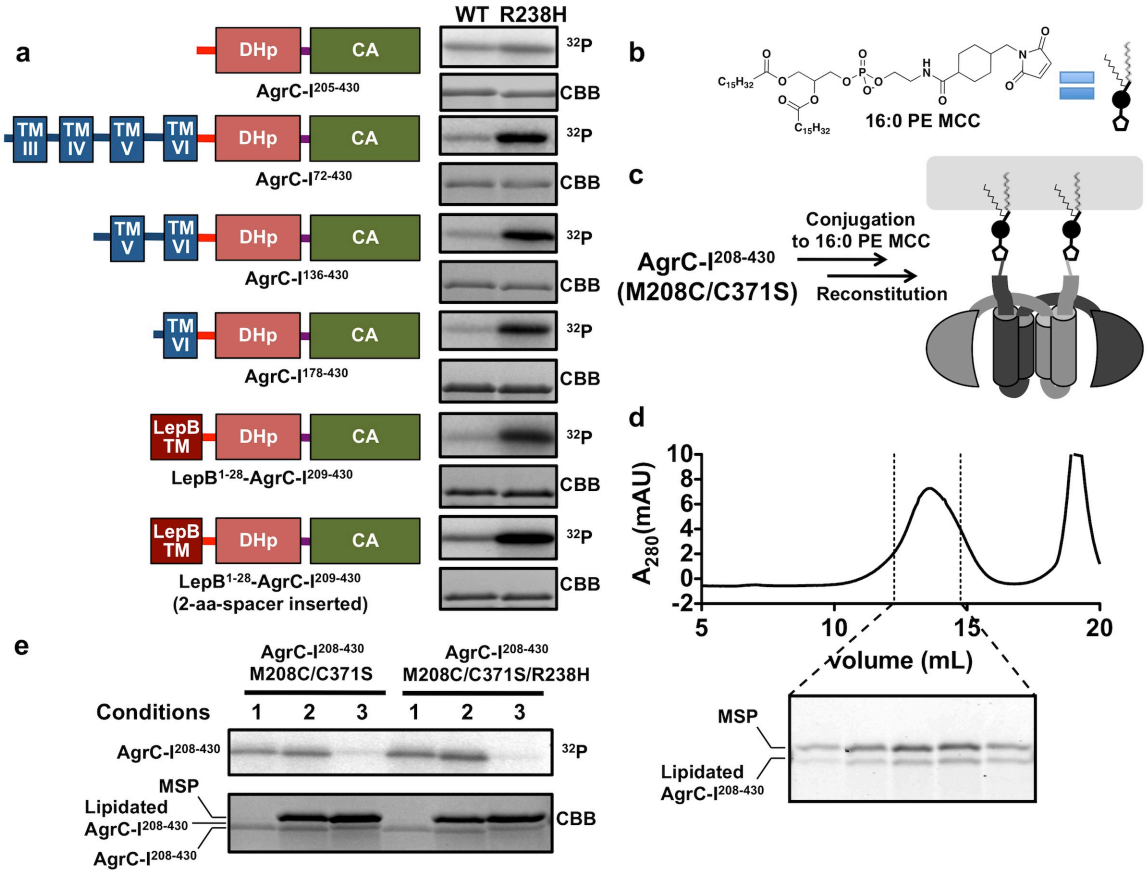


Figure 4.7 Auto-kinase activation by the R238H mutation requires the attachment of the AgrC-I construct to nanodiscs by a TM domain

(a) The effect of the R238H mutation on the auto-kinase activity of various truncated or chimera AgrC-I constructs. All constructs except AgrC-I$^{205-430}$ were first reconstituted to nanodiscs prior to activity analysis. A domain diagram is displayed for each construct next to the corresponding autoradiogram panel, with AgrC-I sensor domain shown in blue, the *E. coli* LepB TM domain in ocher, the interdomain in red, the DHp subdomain in salmon, and the CA subdomain in green. **(b)** The chemical structure and a cartoon presentation of the phospholipid-maleimide derivative, 16:0 PE MCC. **(c)** Diagram showing the putative configuration of the lipidated AgrC-I$^{208-430}$ construct upon nanodisc reconstitution. Intersection of the nanodisc membrane is shown in gray. **(d)** SEC analysis of nanodiscs carrying the lipidated AgrC-I$^{208-430}$. Fractions between the dotted lines were analyzed by SDS-PAGE with Coomassie staining. MSP, membrane scaffold protein. **(e)** Auto-kinase analysis of AgrC-I$^{208-430}$ constructs with or without the R238H mutation. Reaction contains the unmodified AgrC-I$^{208-430}$ (condition 1), the unmodified AgrC-I$^{208-430}$ with empty nanodiscs added *in trans* (condition 2), or the nanodisc-reconstituted, lipidated AgrC-I$^{208-430}$ (condition 3). In panels **(a)** and **(e)**, Auto-phosphorylation was performed employing [γ-^{32}P] ATP and pHis levels analyzed by autoradiography.

the interdomain linker (Figure 4.7a). Therefore, the helicity, and perhaps the orientation of the TM-domain anchors, rather than the HK domain's proximity to the lipid bilayer, plays a crucial role in supporting the strong kinase activity of the mutant HK domain.

4.4 Reconstitution of the AIP-II-revertible constitutive mutants

By contrast to what was observed for AgrC-I^{R238H}, the activation of P3-driven gene transcription caused by a subset of AgrC-I constitutive mutants is revertible by their binding to AIP-II—the peptide was therefore identified as an inverse agonist of AgrC-I (143). However, it is unclear how the baseline auto-kinase activity of AgrC-I and its response to AIP binding change in these mutants. Constitutive point mutations causing AIP-II revertible phenotypes were identified in all modules of AgrC-I. We picked four representative mutants bearing mutation in the sensor (AgrC-I^{S183F}), the interdomain linker (AgrC-I^{L205R}), the DHp subdomain (AgrC-I^{R238K}) or the CA subdomain (AgrC-I^{E306Q}) for auto-kinase assays employing the nanodisc-reconstitution system.

The above four mutants were purified from recombinant expression in *E. coli* and incorporated to nanodiscs (Figure 4.8a). The phospholipid stock used for nanodisc assembly consisted of POPC and POPG at a molar ratio of 1:3. Auto-phosphorylation of the wild-type and mutant AgrC-I discs with [γ-^{32}P]-ATP and $MgCl_2$ revealed mild elevation of the baseline kinase activity in all four mutants (Figure 4.8b). In contrast to the R238H mutant, however, AgrC-I^{R238K} and AgrC-I^{E306Q} carrying HK-domain mutations maintained the sensitivity to AIP-I: their activities in the AIP-I-bound state were indistinguishable from the wild type. AgrC-I^{S183F} and AgrC-I^{L205R} bearing

mutations preceding the HK domain, on the other hand, were only moderately activated by AIP-I. Surprisingly, all four mutants exhibited reduced sensitivity to the inverse agonist, AIP-II. As a consequence, AgrC-I^{L205R} and AgrC-I^{R238K} exhibited, when bound to AIP-II, substantially stronger auto-kinase activities compared to the baseline activity of

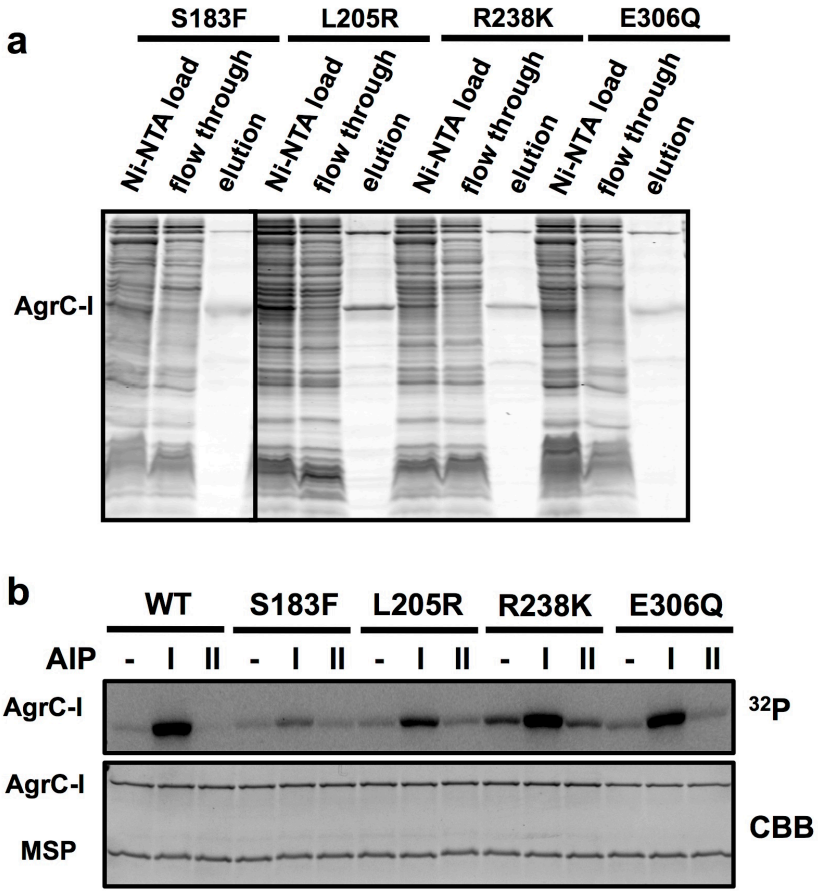


Figure 4.8 Ligand responses of the AIP-II-revertible constitutive mutants
(a) Affinity purification of AgrC-I mutants: fractions from the Ni-NTA purification of indicated AgrC-I mutants were resolved on SDS-PAGE and proteins visualized by Coomassie staining. The expected position for full-length AgrC-I is indicated next to the gel. **(c)** Auto-kinase analysis of AIP-II-revertible AgrC-I mutants: wild-type or mutant AgrC-I discs (0.7 μM) were incubated with 20 μM [γ-^{32}P] ATP, 5 mM $MgCl_2$ and 10 μM AIP if indicated, and the pHis levels were analyzed with autoradiography.

the wild type (Figure 4.8b, compare lanes 9 and 12 to lane 1). It is therefore unlikely that AIP-II reverts the constitutive phenotype of these mutants solely by suppressing their auto-kinase activities. We speculate that the newly discovered activity of AIP-II, i.e., retarding the AgrC-I-AgrA phospho-relay, might play the critical role in reverting the constitutive phenotype of these mutants. This possibility will be investigated in the near future.

4.5 Summary and discussions

4.5.1 Auto-kinase regulation of AgrC-I

In this chapter, we have combined the strength of the nanodisc-reconstitution system and an artificial protein-chimera approach to investigate the signaling mechanism of AgrC-I. Our studies indicate that the interdomain linker assumes a helical conformation whose structural properties are intimately tied to the activity state of the receptor. Cys crosslinking studies in the context of full-length AgrC-I revealed distinct crosslinking patterns for the resting, activated and inhibited states, indicating a substantial reorganization of the structure in the linker in response to binding of both activator and inhibitor AIPs. Interestingly, the first of these patterns (resting state) is not a simple weighted average of the latter two, which is inconsistent with a model wherein the interdomain linkers are in equilibrium between ‘on’ and ‘off’ conformations with ligand binding altering the equilibrium position of the system (Figure 4.9). Rather, this crosslinking data is more consistent with discrete linker conformations for AgrC in all three states.

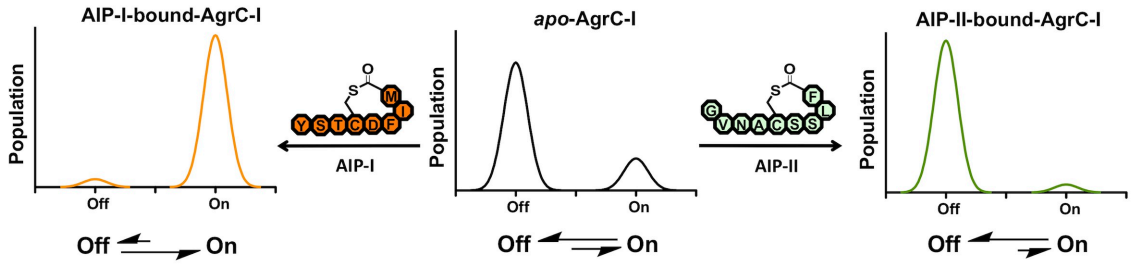


Figure 4.9 An alternative model of the linker conformation in three ligand states
In this model, the linkers are in an equilibrium between an 'on' state associated with high auto-kinase activity and an 'off' state with low activity. Binding of activator or inhibitor AIPs shifts the equilibrium position. See text for details.

Our analysis of a series of GCN4-AgrC-I chimeras provided further insight on the signal-response behavior of AgrC-I^{HK}. Assuming that the interdomain linker adopts a parallel helical conformation, that the GCN4 coiled coil maintains itself as a rigid body in all chimeras and that α-helicity continues at the GCN4-AgrC-I junction, then the activity of AgrC-I HK domain is seen to change gradually as a function of the twisting angle away from the resting state (Figure 4.3d). Intriguingly, the auto-kinase activity of full-length AgrC-I in the activator-bound state is significant weaker than the maximum activity of AgrC-I HK domain exhibited in chimeras GC207 and GC214 (Figure 4.4a), suggesting that the "working range" of the native RHK might be tailored by natural selection to meet the demand of optimal autoinduction timing.

Our chimera approach has some interesting parallels with the work of Moffat and coworkers on FixL from *Bradyrhizobium japonicum* (157). Similar to AgrC, the linker region in FixL connecting the sensor and HK domains is proposed to be helical. Chimeras between the FixL HK domain and a LOV light sensor domain from a heterologous protein (YtvA from *B. subtilis*) are active only when the FixL residue at the junction occupies two periodically equivalent positions ('*i*' and '*i* + 4') in each haptad

repeat of the putative linker coiled-coil. Maintaining the appropriate periodicity of the helical linker is therefore thought to be critical to the construction of functional RHKs via shuffling between sensor and HK domains, including such events as they occur during evolution (158). In stark contrast to the YtvA-FixL chimeras, our GCN4-AgrC-I constructs exhibited many intermediate activities between the two extremes: the system is remarkably plastic. Shuffling between the AgrC-I^{HK} (or other HK domains with similar behaviors) and heterologous sensory modules could therefore result in new RHKs exhibiting variable degrees of basal kinase activity and possibly dynamic range, depending on the fusion junction and the structural rearrangement of the sensor between signaling states. Conceivably, such functional plasticity would be beneficial during RHK evolution by increasing the chance of generating competent chimera receptors that could serve as starting points for further selection for an optimal working range for the new sensory function.

Following the publication of our work discussed above (159), three papers came out each describing a crystal structure in which HK domain dimers were captured in the catalytic-active state (122, 153, 160). As opposed to the symmetric, inactive HKs (Figure 4.10a), the active HK dimers are strikingly asymmetric: while one CA subdomain binds to the phospho-acceptor His in a catalysis-competent conformation, the other still adopts the inactive conformation as if in an inactive dimer (Figure 4.10b). Key to this asymmetry is the extra bending of the DHp-α1 colored in yellow in Figures 4.10a and b (Figure 4.10a, red arrow), which disrupts a hydrophobic patch that would otherwise sequester the catalysis-competent CA subdomain (Figure 4.10a and b, compare the electrostatic surface

in the dashed circle). In compliance with these structural evidences, helical "tilting" or "scissoring" of the linkers preceding the DHp subdomains has been observed in a few full-length RHKs as a consequence of signaling input (152, 161, 162). Rotational movements, on the other hand, have been proposed in fewer models and mostly as a

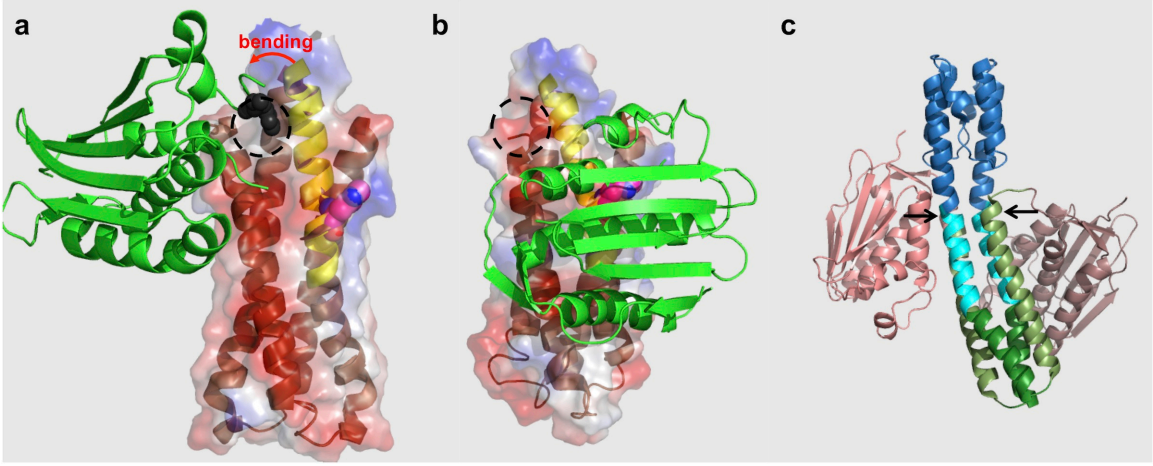


Figure 4.10 New structural evidence on the molecular movement related to RHK activation
(a and **b)** Comparison between crystal structures of two HK-domain dimers, one captured in the inactive state (PDBID: 2c2a, panel **a**) and the other in the active state (PDBID: 4i5s, panel **b**). In each dimer, one inactive CA subdomain is hidden for clarity. The remaining CA subdomain is colored green, adopting the inactive conformation in **(a)** and the kinase-competent conformation in **(b)**. The DHp-α1 helix that experiences bending movement is colored in yellow and the rest of the DHp four-helix bundle is shown in ocher. The phospho-acceptor His that would react with the displayed CA is highlighted in magenta. A transparent electrostatic surface is also displayed for the DHp subdomain. In **(a)**, the F-box Phe residue (black spheres) binds to a hydrophobic patch (dashed circle) in the DHp four-helix bundle and holds the CA from approaching the phospho-acceptor His. A red arrow indicates the bending motion of the yellow helical segment that would activate the displayed CA subdomain. In **(b)**, the dashed circle indicates the disruption of the hydrophobic patch as a consequence of the bending motion, and the CA subdomain relocated to bind the His in a catalysis-competent conformation. **(c)** The intracellular domains of the *E. coli* stress sensor CpxA (PDBID: 4biv). The highly symmetric HAMP domain is colored in blue, and the helical elements in the DHp subdomain are colored sequentially in cyan, green and light green. Arrows highlight the differential helical bending between the HAMP and DHp in two subunits that breaks the symmetry of the dimer.

component of the overall linker movement in conjunction with tilting or scissoring (152, 162). Notably, however, the local, symmetric helical rotation found in the AgrC-I interdomain linkers in our study needs not contradict with the asymmetry in the HK module necessary for kinase activation. This notion is supported by the perfect symmetry of HAMP transmission modules in an overall asymmetric structure of the intracellular portion of the *E. coli* stress sensor RHK, CpxA in its active conformation (Figure 4.10c) (122). Specifically, the interdomain linkers connecting the HAMP and DHp modules are bent at different magnitudes to adapt to the conformational mismatch between their flanking modules (Figure 4.10c, black arrows). In our inter-subunit crosslinking analysis of AgrC-I, the linker segment (residues 205-209) in which we performed Cys substitution is very close to the sensor domain and perhaps upstream to the putative kink site. Assuming this arrangement to be true, this segment of the linker pair may retain a two-fold symmetry in full-length AgrC-I at different ligand states.

The success of the protein-chimera strategy also strictly relies on the capability of α-helices to bend and/or unwind in order to relieve conformational stress. According to the designing principle, the junction residues would be rotated from its resting register by about 180 degree in some chimeras. This movement, if faithfully propagated to the DHp subdomain, would displace the hydrophobic residues out of the four-helix bundle core and thereby dismantle the HK dimer. The interdomain linkers in these chimeras, therefore, must modulate the severe rotational input into a combination of milder rotation and tilting output discernible for the HK domain. Conceivably, this modulation is not possible when the linker is too short and perhaps as a consequence, almost no

intermediate level of auto-kinase activities was observed in chimeras with linker length shorter than 7 residues (Figure 4.3a). It is currently unclear how this modulation enables the continuous correlation between the rotational conformation input and the auto-kinase activity. A multi-facet strategy involving high-resolution structural characterization and single-molecule fluorescence measurement is proposed in section 6.1.3 to address this fascinating puzzle.

4.5.2 Trans-kinase regulation of AgrC-I

Upon auto-phosphorylation, the RHK component relies on its trans-kinase activity to relay the phosphoryl group to its cognate RR. Despite the essential role of this activity in the TCS signaling, its regulation has been overlooked in most biochemical studies of other TCS systems. Two reasons might account for the lack of quantitative kinetic studies on phospho-relay reactions: (1) that these reactions are too fast to be accurately quantified ($t_{1/2}$ on the order of a few seconds, see Figure 3.6e); and (ii) that the phosphatase activity of the RHK, which is also variable among signaling states, interferes with the phospho-relay kinetics. In our *in vitro* reconstitution studies on the wild-type AgrC-I, AIP-II was found to significantly inhibit the phospho-relay from the receptor to its cognate RR, AgrA (Figure 3.6a and b). In fact, the cognate AIP-I slightly accelerates this process (Figure 3.6a, compare lanes 11 and 12 to 3 and 4). Key to these findings is the complete lack of phosphatase activity in AgrC-I (Figure 3.6c and d), a rare property among RHKs that may facilitate further investigations into the conformation-activity relationship of the AgrC-I HK domain with respect to the trans-kinase activity. To realize quantitative measurement of the trans-kinase activity in the GCN4-AgrC-I chimera

series, an assay based on a three-component phospho-equilibration system is designed in section 6.1.2.

4.5.3 Mechanisms of constitutive activity

Novel TCSs emerge in two ways: lineage-specific expansion (LSE), in which the HK domain of the ancestral RHK recombines with a new set of sensor domain(s) to carry out a new sensory function in the same organism, or horizontal gene transfer (HGT), in which the ancestral TCS is introduced to, and performs its original function in, a distantly related organism (163). In both scenarios, however, successful establishment of a sensory pathway entails adaptation of the nascent TCS to the signaling demand of the host organism, which would depend primarily on natural selection on beneficial mutations that occur randomly (158). Central to this adaptation is the adjustment of the threshold level of the signal as well the sensitivity of the response to signals above the threshold. At the molecular level, these properties of a TCS depend in part on the conformation-activity landscape of the RHK component, which dictates its resting and activated kinase activities. Therefore, how this landscape changes following point mutations would be indicative to the evolutionary trajectory of RHKs. In this Chapter, we examined the conformation-activity landscapes of an AgrC-I HK domain bearing a point mutation, R238H. We found that the auto-kinase activity of this mutant is less sensitive to rotational motions compared to the wild type: that it fluctuates so mildly around a relatively high level, enables the corresponding full-length RHK, regardless of the ligand state, to activate P3-driven reporter genes. This result exemplifies the profound impact that a point mutation might have on the conformation-activity landscape of an HK

domain, and should inspire a more comprehensive effort to understand the "evolvability" of this landscape, perhaps employing a saturation mutagenesis approach (164).

In contrast to the HK-domain constitutive mutations (R238H, R238K and E306Q), those occurring in or proximal to the sensor domain (S183F and L205R) apparently do not alter the conformation-activity landscapes of the HK domain and, therefore, the mechanism of action for these mutations is less tractable. In the previous report in which these mutants were identified, the authors suggested that both mutants tend to shift the last TM domain towards the cytoplasm and thereby mimic the activated conformation even in the absence of AIP-I (143). Such "piston-like" movements have been ruled out for the mechanism of type-I RHK activation (165, 166). Nonetheless, the multi-span sensor domain in AgrC-I may allow this TM domain to, after the piston-like shift, rotate or tilt in order to optimize its packing with the rest of the sensor domain and thereby confer auto-kinase activation. This model needs to be tested experimentally in the future, primarily through assessing the solvent accessibility of certain membrane-proximal residues and mapping the mutation-induced conformational change of the interdomain linkers.

Chapter Five: The Key Driving Forces That Power AIP Thiolactone Formation

As was described in section 1.2.3, the AIP contains a thiolactone ring formed by condensation of the C-terminal carboxyl group and the sulfhydryl group of an internal cysteine. This macrocyclic structure is absolutely necessary for binding of the peptide to its receptor, AgrC. AIP biosynthesis is of particular interest to us because the high-energy thiolactone linkage forms directly from the AgrB-catalyzed proteolysis of AgrD without free-energy input from ATP hydrolysis (Figure 5.1a) (55). Notably, the reaction occurs at the inner leaflet of the cytosolic membrane and the N-cleavage fragment, herein referred to as AgrD(1-32)-thiolactone, is then translocated to the extracellular leaflet of the cell membrane (Figure 5.1a) (56). It is unclear how the bacteria overcome the thermodynamic

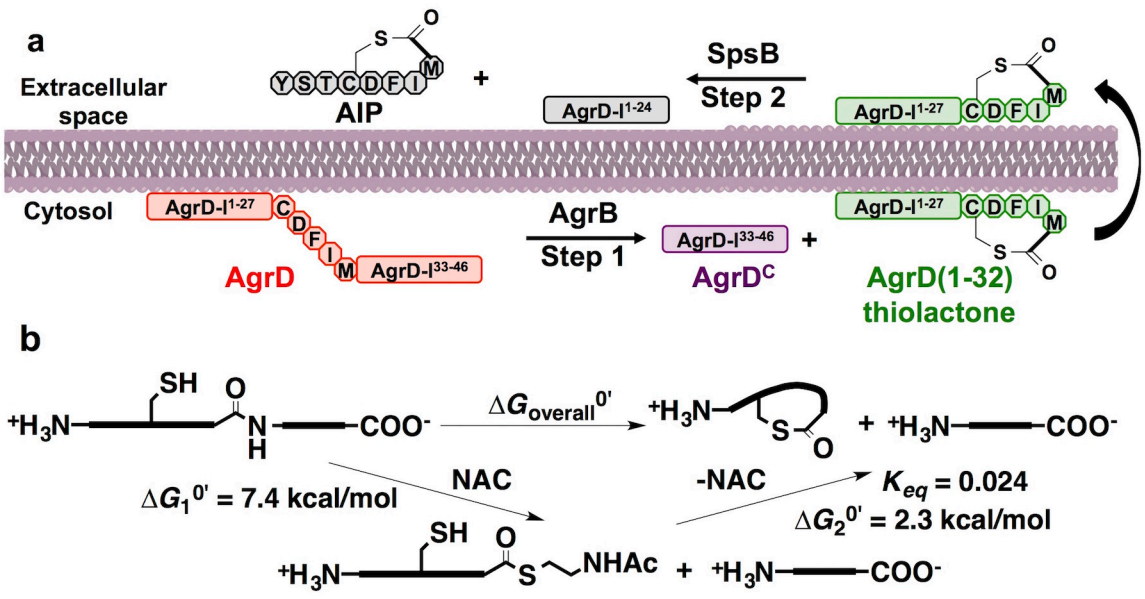


Figure 5.1 The AIP biosynthesis pathway and the model for free-energy change estimation for the thiolactone formation.
(a) The AIP production pathway in *S. aureus* cells as exemplified by processing of AgrD-I. **(b)** A two-step model that recapitulates the thiolactone formation from a linear peptide. Estimated values of free-energy change and/or equilibrium constant are shown for each step. See text for details.

challenge to maintain, on the cytoplasmic side of the cell membrane, an abundant pool of the AgrD(1-32)-thiolactone to support its rapid translocation required for punctual *agr* autoinduction. In this chapter, we both estimated the ΔG for the thiolactone-formation reaction based on existing data and determined it experimentally. To account for the discrepancy between the estimation and the experimental results, the stability of the thiolactone macrocycle was also probed under various conditions employing a ring-opening equilibrium approach. Our data show that efficient thiolactone production is driven by two unanticipated features of the system; (i) membrane-association of the thiolactone-containing intermediate which stabilizes the macrocycle and (ii) rapid degradation of the C-terminal proteolysis fragment, $AgrD^C$, which affects the reaction equilibrium position.

5.1 Estimation of the free energy cost associated with thiolactone formation

To get a better picture for the thermodynamic challenge associated with conversion of AgrD into a thiolactone, we first estimated the ΔG for the reaction based on existing data on analogous model reactions. A two-step process would recapitulate such a reaction: the scissile peptide bond first undergoes thiolysis with a coenzyme A mimic, N-acetylcysteamine (NAC), to give an N-terminal thioester fragment and an unmodified C-fragment (Figure 5.1b) and the N-terminal fragment subsequently cyclizes into a thiolactone to release NAC (Figure 5.1b). In the following discussion, ΔG^{o}' is defined as the free energy change of a chemical process at 37^{o}C, 100kPa and pH = 7.0 with all reactants involved present at standard activities.

The first step can be accomplished through first the hydrolysis of the scissile peptide bond and the subsequent thioesterification of the N-fragment by NAC. For the hydrolysis process we referred to the hydrolysis of a Gly-Gly bond, $\Delta G^{0'}$ = 2.2kcal/mol (167). For the thioesterification process, the $\Delta G^{0'}$ for the condensation between acetic acid and coenzyme A (CoA), i.e., 7.5 kcal/mol (168), needs to be adjusted for the difference in acidity between the C-terminal carboxylic group of a generic peptide (pK_a = 3.3) and acetic acid (pK_a = 4.8) (169). pK_a value of acids have been found to linearly correlate with the pK_{eq} (negative logarithm of the equilibrium constant) of the hydrolysis of their methyl esters into their deprotonated anions and methanol (170). The slope is 1.02. Assuming this linear relationship holds for thioesters, thioesterification of a C-terminal carboxylate on a peptide should be 2.1 kcal/mol less favorable than that of acetate (169). Therefore, the estimated overall $\Delta G_1^{0'}$ for the thiolysis of a generic peptide bond should be:

$\Delta G_1^{0'}$ = -2.2 + 7.5 + 2.1 = 7.4 kcal/mol equation 5.1

With respect to the second, ring closure step, the $\Delta G^{0'}$ is strongly dependent on the ring being formed. As a moderate estimation, we assume that the ring formed upon trans-thioesterification to be a typical large ring such that it is neither strongly stabilized, like a five- or six-membered ring, nor destabilized like those in mid-sized rings. We also assume that the trans-thioesterification is isoenthalpic, primarily because the pK_a of the thiol group in a cysteine side chain is close to that in NAC.

We calculate the $\Delta G_2^{0'}$ of the cyclization from an estimated equilibrium constant K_{eq}, which equals the apparent first-order kinetic constant of the forward reaction divided by

the apparent second-order kinetic constant of the reverse (thiolysis) reaction. The denominator should be equivalent to the secondary kinetic constant of a thiol-exchange process involving a linear peptidyl-NAC thioester, in that the rotational movements around the bonds adjacent to the thioester motif are not restrained in the large ring. The equilibrium constant therefore equals to the ratio between the first-order kinetic constant of the intramolecular (cyclization) reaction and the second kinetic constant of an analogous intermolecular (thiol-exchange) reaction between a thioester and a thiol group. This ratio, by definition, represents the effective molarity (EM) of the cyclization. Empirically, the EM of large, unstrained ring formation from bifunctional chain molecules falls predominantly within the range 0.01M - 0.05M independent of the functional groups involved, as was summarized by Illuminati and Mandonlini (171). This range corresponds to a $\Delta G_2^{0'}$ value ranging from 1.8 to 2.8 kcal/mol. We hence used $\Delta G_2^{0'}$ = 2.3 kcal/mol to calculate the $\Delta G^{0'}_{\mathbf{overall}}$ and the $K_{\mathbf{eq,\ overall}}$ for the strain-free thiolactone formation from a generic peptide (equations 5.2 and 5.3).

$\Delta G^{0'}_{\mathbf{overall}} = \Delta G_1^{0'} + \Delta G_2^{0'} = 9.7$ kcal/mol equation 5.2;

$K_{\mathbf{eq,\ overall}} = \exp(-\Delta G^{0'}_{\mathbf{overall}} / RT) = 1.4 \times 10^{-7}$ equation 5.3.

The estimated $K_{\mathbf{eq,\ overall}}$, if applied to the AgrD(1-32)-thiolactone formation from AgrD, would restrict the intracellular abundance of the former to a level incompatible with the AIP production rate required for *agr* autoinduction (see Discussion). Therefore, assuming the thiolactone intermediate to be the direct product of AgrB-mediated proteolysis *in vivo*, some mechanism must exist to render the reaction more permissible.

5.2 Expression and purification of recombinant AgrB and AgrD constructs

We elected to study AgrB and AgrD from *S. aureus* group-I and -II for our reconstitution studies. The current model of AIP biosynthesis has been developed primarily using group-I Agr proteins, which are significantly diverged from their group-II counterparts: AgrB-II and AgrD-II share 61% and 48% sequence identity with AgrB-I and AgrD-I, respectively. Importantly, genetic studies show that heterologous processing of AgrD is forbidden between these two groups (62).

Wild-type AgrD-I and AgrD-II, as well as N-terminally FlagHis_6-tagged versions thereof, were generated using a recombinant fusion protein strategy (Figures 5.2a and 7.1, Appendix Figure 1 and Appendix Table 2). AgrB-I and AgrB-II were over-expressed in *E. coli* as His_6-tagged recombinant proteins and purified (Figure 5.2b). Surprisingly, AgrB-I showed multiple bands when analyzed by SDS-PAGE, while AgrB-II migrated primarily as a single species of the expected molecular weight with a small amount of an apparent dimer. Western blot analysis revealed the presence of the His_6-tag in all of these high molecular weight species, indicating that they relate to AgrB-I and AgrB-II (Figure 5.2b). That these high molecular weight species might result from covalent oligomers, was ruled out by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) (Figure 5.2c).

The unusual electrophoretic behavior of AgrB-I, and to a lesser extent AgrB-II, could be a consequence of heterologous oligomerization induced by SDS micelles. However, native inter-subunit interactions within membrane protein complexes are often resistant to

SDS (172), leading us to wonder whether AgrB might oligomerize under more native conditions. This idea was supported by a crosslinking experiment on AgrB-II solubilized with a zwitterionic detergent, fos-choline-12, in which specific dimer formation was observed upon short glutaraldehyde treatment (Figure 5.3a). Since detergent-solubilized AgrB-II is not active (data not shown), we asked whether the dimerization persists upon incorporation to nanometer-scale lipid bilayer discs, or nanodiscs (see chapter 3) (138). Boldog and coworkers have employed a clever nanodisc based strategy to show that the bacterial chemoreceptor, Tar, forms stable dimers (173). We adapted this strategy to

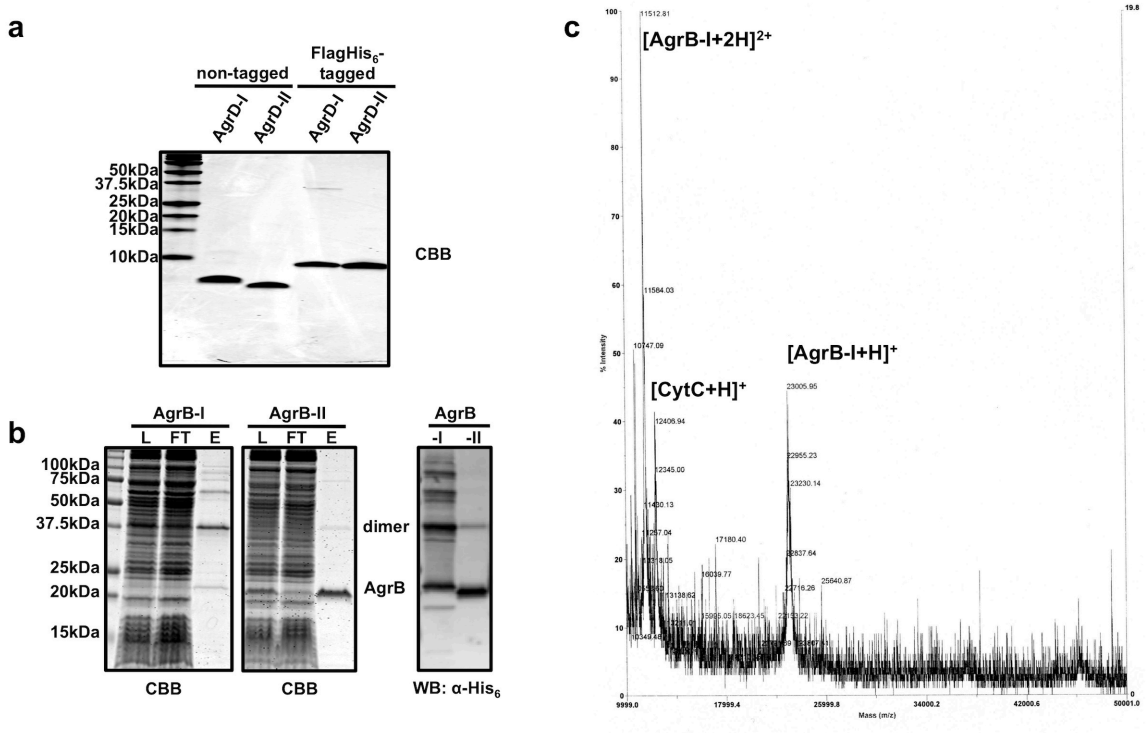


Figure 5.2 Preparation of recombinant AgrB and AgrD
(a and **b)** SDS-PAGE analysis of all four AgrD constructs prepared herein (panel a) and samples from the affinity purification of AgrB proteins (panel b). L: bacterial membrane extract; FT: flow through; E: elution. An anti-His_6 western blot of purified AgrB proteins is shown to the right. **(c)** MALDI mass spectrum of AgrB-I (calculated molecular weight = 22.7 kDa) showing the absence of covalent dimer. Bovine cytochrome C (molecular weight = 12,384 Da, showed up at m/z = 12,407 Da) was used as an internal standard for m/z calibration. The monovalent and divalent ions of the AgrB-I peak at m/z = 23006 Da and 11513 Da, respectively.

study the oligomerization behavior of AgrB in a bilayer (Figure 5.3b). Accordingly, nanodiscs were assembled in the presence of different amounts AgrB-II, such that the amount of the protein varied from being sub-stoichiometric relative to the number of discs to being in large excess. AgrB-II-incorporated nanodiscs from each assembly were isolated through Ni-NTA purification followed by size-exclusion chromatography (SEC), and the averaged protein molecular weight in each particle was analyzed using SEC in tandem with a multi-angle light scattering detector (SEC-MALS, Figure 5.3c). The average AgrB-II (22 kDa) copy number per nanodisc was calculated by factoring in two

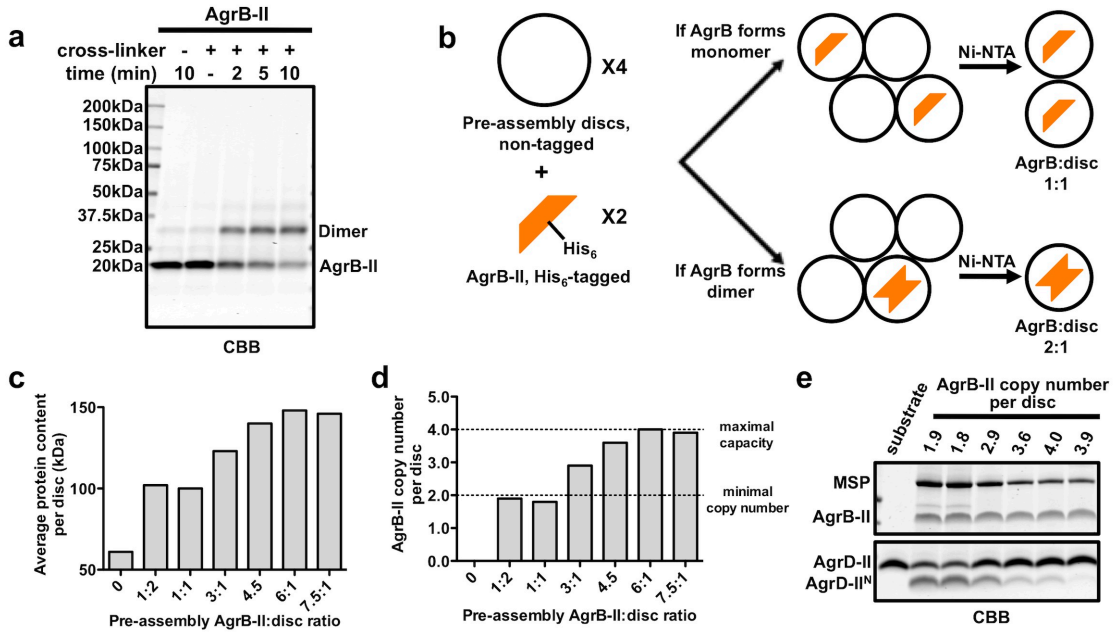


Figure 5.3. Dimerization of AgrB under native conditions.
(a) SDS-PAGE analysis of detergent-solubilized AgrB-II before or after glutaraldehyde treatment. **(b)** Theoretical distribution of AgrB structural units in nanodiscs: diagram showing the scenarios of reconstituting AgrB into excess nanodiscs assuming AgrB preferably forms monomers or dimers. **(c** and **d)** Protein content per disc **(c)** and AgrB copy number per disc **(d)** of affinity-purified AgrB-II nanodiscs as a function of pre-assembly AgrB-to-nanodisc ratio. **(e)** Proteolysis of FlagHis$_6$-AgrD-II after treatment with AgrB-II nanodiscs: reactions were analyzed by SDS-PAGE. Average AgrB-II copy number per disc is indicated. The products, AgrD-II-(1-32) fragments (linear and/or thiolactone) are collectively indicated as AgrD-II^N. MSP, membrane scaffold protein.

copies of MSP (30 kDa each) per particle, a known feature of this system (Figures 5.3c and d) (173). In line with our finding using detergent-solubilized protein, AgrB-II was incorporated as dimers when deposited at a monomer-to-disc ratio of 0.5:1 or 1:1. It is noteworthy that the monomer-to-disc ratio of the purified AgrB-II discs levels out at 4, suggesting that the 12.8 nm-diameter nanodisc is capable of accommodating two AgrB-II dimers. The proteolytic activity of these nanodiscs was tested using the FlagHis_6-tagged AgrD-II substrate (Figure 5.3e). Interestingly, nanodiscs containing a single copy of the AgrB-II dimer were more active than those containing two copies of the dimer (Figure 5.3e, compare lanes 2 and 3 to lanes 6 and 7). This is possibly due to the better availability of membrane surface necessary for substrate docking in nanodiscs containing single copy of AgrB-II dimer (56). The key result, however, is that AgrB appears to form stable dimers in a biologically active state.

5.3 AgrD(1-32)-thiolactone is the major product of the AgrB-catalyzed proteolysis of AgrD

We next set out to characterize the products of AgrB-mediated proteolysis of AgrD. To this end, AgrB and AgrD (from *agr*-I or -II) were co-reconstituted into liposomes composed of POPC and POPG at a ratio of 3:1. This lipid composition was selected based on preliminary proteolysis assays employing AgrB embedded in liposomes comprising POPC-POPG mixtures at different ratios (Figures 5.4a and b). Tris-Tricine SDS-PAGE revealed efficient proteolysis occurring between cognate AgrB-AgrD pairs but not heterologous ones (Figure 5.4c). Further characterization of the proteolysis products of non-tagged AgrD-I and AgrD-II was achieved using reverse-phase (RP)

HPLC and mass-spectrometry (MS) (Figure 5.5a). For each reaction, RP-HPLC/MS identified three products in additional to the starting material, namely; AgrD^{C}, AgrD(1-32)-thiolactone and linear-AgrD(1-32) (Figures 5.5b-d and Appendix Figure 2). Importantly, consistent with the AIP biosynthesis model depicted in Figure 5.1a, the thiolactone is the major AgrD(1-32) fragment in both reactions (Figures 5.5.b and d, green arrows). We therefore refer to the AgrB-catalyzed processing of AgrD as the 'proteolytic cyclization'.

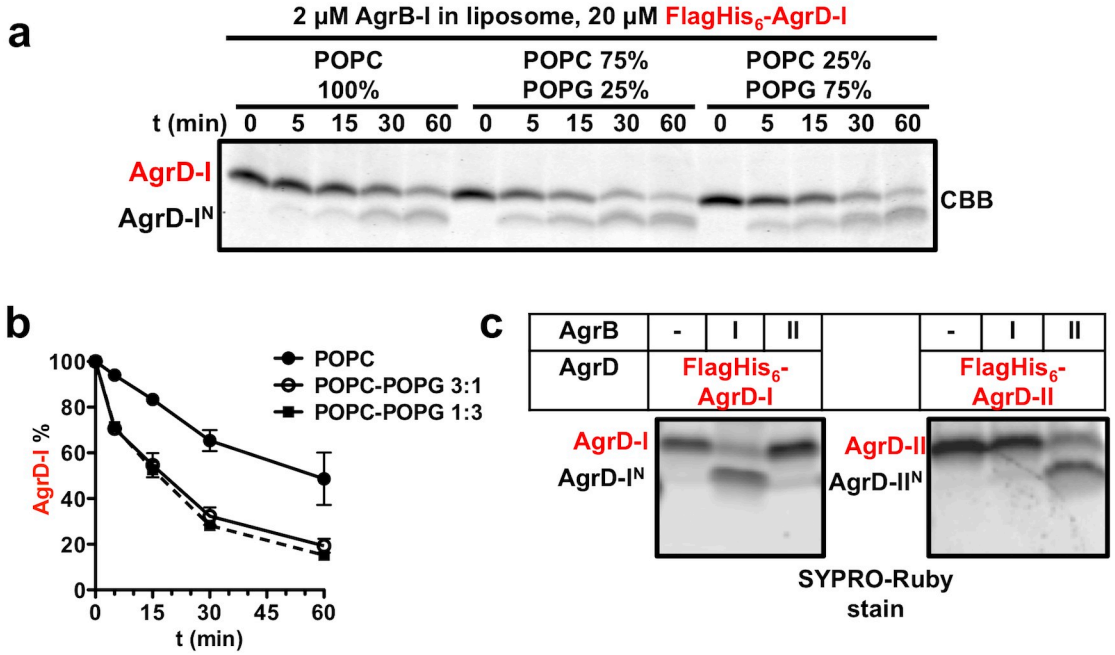


Figure 5.4 Lipid-composition dependence and subgroup specificity of the AgrB-catalyzed proteolysis of AgrD.
(a and **b)** Lipid-composition dependence of AgrB-I activity. FlagHis$_6$-AgrD-I was treated with AgrB-I proteoliposomes consisting of phospholipids at indicated compositions. Samples were withdrawn at indicated time points for SDS-PAGE analysis. **(a)** SDS-PAGE analysis of one representative time course and **(b)** kinetic plot of AgrD-I consumption from densitometric quantification of the gel are shown. Error bars = range (n = 2). **(c)** Proteolysis specificity between AgrB-AgrD pairs from *agr*-I and *agr*-II alleles. FlagHis$_6$-tagged AgrDs were treated with indicated liposomal reagents at 37^{o}C for 6hrs. The reactions were resolved on Tris-Tricine SDS-PAGE and proteins visualized with SYPRO-Ruby staining (Invitrogen). In panels **(a)** and **(c)**, the AgrD(1-32) fragments (linear and/or thiolactone) are collectively indicated as AgrD-I^{N} or AgrD-II^{N}.

5.4 AgrD processing by AgrB is a reversible process

In our initial reconstitution trials, complete consumption of AgrD was never observed, despite significant production of AgrD(1-32)-thiolactone. To account for this, we speculated that the proteolytic cyclization might have reached a balanced equilibrium under our reaction conditions. To test this idea, we treated AgrD-I (the FlagHis$_6$-tagged version, 20 μM) with proteoliposomes containing 1 μM AgrB-I and analyzed the reaction

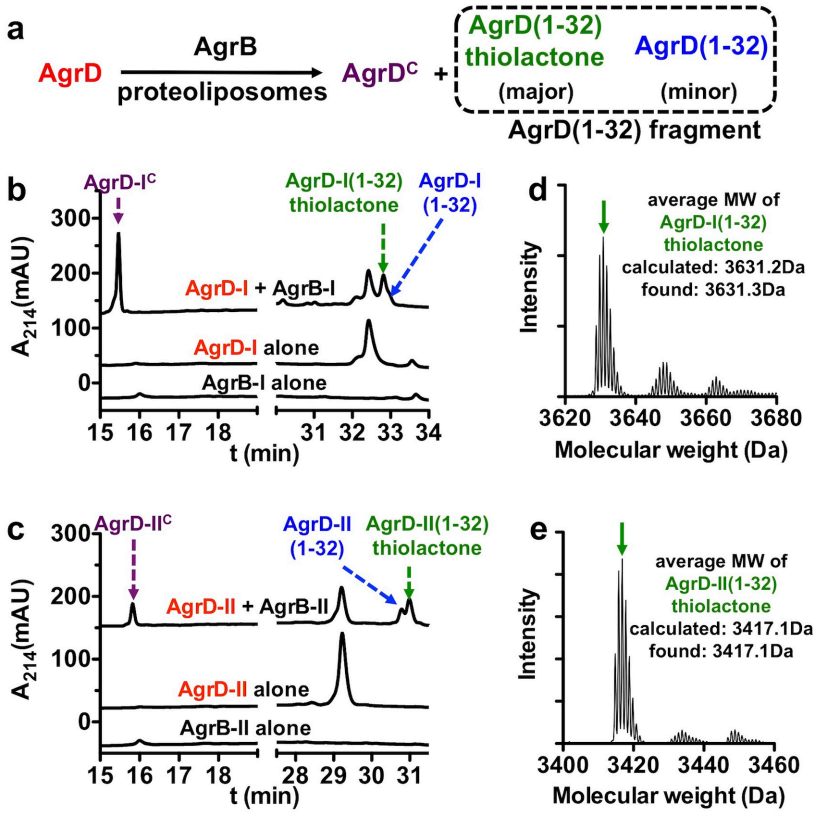


Figure 5.5 AgrD(1-32)-thiolactone is the major product of the AgrB-catalyzed proteolysis of AgrD.
(a) Schematic representation of the proteolysis reaction. **(b** and **c)** C4 RP-HPLC chromatograms of the proteolysis reactions of **(b)** AgrD-I and **(c)** AgrD-II. Product peaks are indicated with arrows. **(d** and **e)** Deconvoluted MS highlighting the isotopic profile of the AgrD(1-32)-thiolactone species (green arrows) generated from **(d)** AgrD-I and **(e)** AgrD-II.

mixture by RP-HPLC at a series of time points (Figure 5.6a). We observed time-dependent consumption of the starting material, as well as the production of three new species identified by MS as the $AgrD^{C}$ cleavage fragment, and the thiolactone- and linear AgrD(1-32) fragment. The amount of starting material and products were quantified based on integration of the RP-HPLC traces and the molarity of each species was normalized to the starting molarity of AgrD (Figure 5.6b). During the initial stage of the reaction ($t < 10$ minutes), the full-length precursor AgrD was efficiently cyclized to AgrD(1-32)-thiolactone (with the concomitant release of the $AgrD^{C}$ fragment), while only small amounts of linear AgrD(1-32) were generated. The cyclization then gradually slowed down, reaching a maximum by about 30 minutes, which is likely due to the dynamic equilibrium being reached (vide infra). In contrast to this equilibration behavior, the linear AgrD(1-32) was produced irreversibly at a slow but constant rate throughout the time course (Figure 5.6b). Intriguingly, its production after $t = 30$ minutes was coupled to a concerted decrease in the molar fraction of AgrD and AgrD(1-32)-thiolactone (Figure 5.6b). This kinetic observation can be understood in terms of a pre-equilibrium model, in which the proteolytic cyclization is a reversible process occurring much faster than the irreversible production of linear AgrD(1-32). In this model, the fast reaction reaches a quasi-equilibrium state that approximates the real thermodynamic equilibrium.

To further explore the apparent reversibility of the proteolytic cyclization, we synthesized an AgrD-I^{C} peptide with a C-terminal carboxyamide (AgrD-I^{C}-NH_2) and attempted to

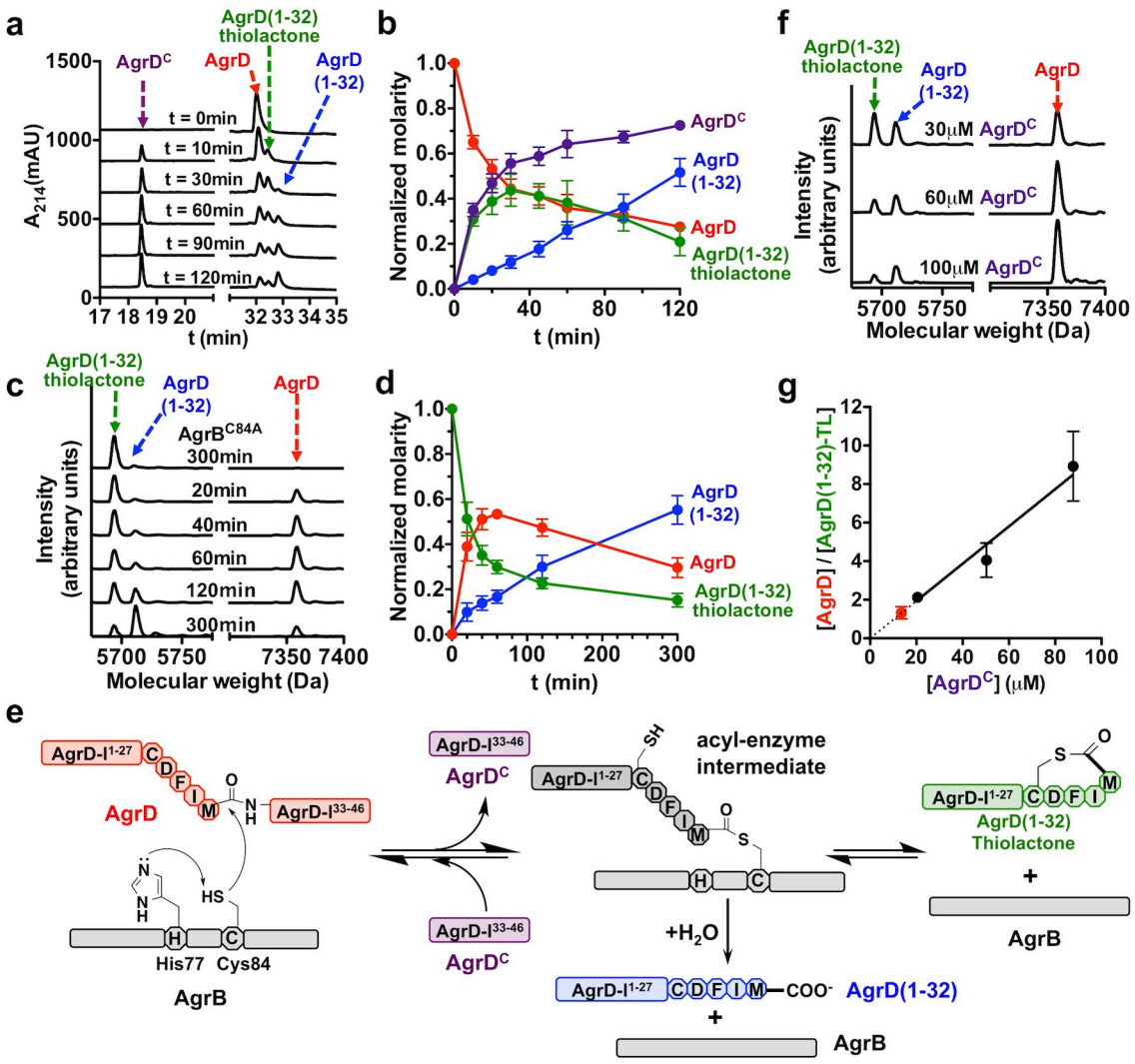


Figure 5.6 AgrD processing by AgrB is a reversible process.
(a and **b)** Time course of the proteolytic cyclization: a representative set of RP-HPLC curves **(a)** and a kinetic plot of indicated species generated from integrated peak areas thereof **(b)** are shown. In **(b)**, all molarity was normalized to the amount of full-length AgrD at t = 0. Error bars = SD (n = 3). **(c** and **d)** Time course of the ligation reaction: samples were resolved on RP-HPLC and amount of full-length AgrD and both AgrD(1-32) fragments was quantified using MS employing standards. A representative set of deconvoluted MS spectra **(c)** and a kinetic plot of indicated species **(d)** are shown. Error bars = SD (n = 3). **(e)** Scheme showing the mechanism of AgrB-catalyzed proteolysis of AgrD. **(f)** Ligation reactions starting at a series of $AgrD^C$ concentrations: reactions were analyzed at t = 120 min and deconvoluted mass spectra are shown as in panel **(c)**. **(g)** Plot showing the equilibrium molar ratio between AgrD and AgrD(1-32)-thiolactone as a function of the molarity of $AgrD^C$ in the ligation (black circles) and proteolysis (red circles) systems. Error bars = SD (n = 3).

ligate it to a recombinantly prepared, FlagHis$_6$-tagged AgrD-I(1-32)-thiolactone (Figure 7.1 and Appendix Figure 1) in the presence of AgrB-I proteoliposomes, thereby generating the corresponding full-length AgrD-I construct (Appendix Figure 2). Since this ligation product co-eluted with the AgrD(1-32)-thiolactone starting material on RP-HPLC, quantification of these two species relied on MS analysis employing ratiometric standards (Figures 5.6c and d). The ligation reaction occurred rapidly and a quasi-equilibrium state was achieved within 40 minutes. Importantly, the post-equilibration behavior of the ligation system (after t = 40 minutes) showed a marked similarity to that observed with the proteolysis reaction (compare Figures 5.6d to b): the molar fraction of both AgrD and AgrD(1-32)-thiolactone dropped in a concerted fashion, while the ratio between them stayed roughly constant. This result unambiguously demonstrated a balanced, dynamic equilibrium of AgrD proteolysis under our reaction conditions.

It has been proposed that the active-site cysteine within AgrB (Cys84) attacks the scissile peptide bond in AgrD to form an acyl-enzyme intermediate that subsequently undergoes an intramolecular trans-thioesterification to give the AgrD(1-32)-thiolactone (Figure 5.6e) (55). Hydrolytic resolution of this acyl-enzyme intermediate could potentially yield the linear AgrD(1-32) peptide. In accordance with this model, the inactive AgrB$^{\mathbf{C84A}}$ mutant is unable to ligate the AgrD$^{\mathbf{C}}$ peptide to AgrD(1-32)-thiolactone or to hydrolyze the thiolactone (Figure 5.6c, top spectrum and Figures 5.7a and b). Therefore, we conclude that the catalytic activity of AgrB supports the proteolysis/ligation equilibrium and the irreversible production of AgrD(1-32) (Figure 5.6e).

We next set out to determine the equilibrium constant of the proteolytic cyclization reaction. To this end, the pre-equilibrium positions in the presence of different concentrations of AgrD^{C} were monitored and plotted (Figures 5.6f and g, black data points). As expected, the pre-equilibrium position shifted towards the formation of full-length AgrD at higher AgrD^{C} concentrations. Linear regression of the plot in Figure 5.6g returned an equilibrium constant for the proteolytic cyclization, K_{eq} = 1.0x10^{-5}, or if one treats AgrD^{C} as a reversible ligand for the thiolactone, a dissociation constant (K_D) of 10 μM. This corresponds to a $\Delta G^{0'}$ of 7.1 kcal/mol, which is 2.6 kcal/mol more favorable than expected based on our initial estimate from model reactions. Gratifyingly, the pre-equilibrium state obtained using AgrD as the starting material (Figure 5.6b) conformed well to this equilibrium constant derived form the linear regression (Figure 5.6g, the red data point).

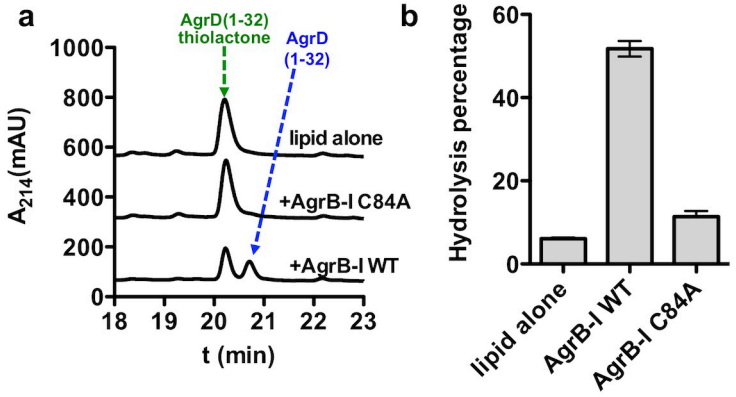


Figure 5.7. Generation of AgrD(1-32) is dependent upon AgrB catalysis.
The AgrD(1-32)-thiolactone was treated with indicated liposomal reagent at 37 °C for 1hr and then Sep-Pak processed and resolved on RP-HPLC. A representative set of chromatograms **(a)** and a bar graph showing the percentage of hydrolysis (corrected for SPE recovery) **(b)** are shown. Error bars = range (n = 2).

5.5 AgrD(1-32)-thiolactone is stabilized by phospholipids

Our experimentally measured equilibrium constant for proteolytic cyclization indicates the reaction is about 70-fold more favorable than might be expected based on model reactions (K_{eq} = 1.0 x 10^{-5} *vs.* 1.4 x 10^{-7}). In an effort to gain some insight into this discrepancy, we asked whether the macrocycle within the AgrD(1-32)-thiolactone was somehow stabilized under the reaction conditions. With this in mind, we compared the susceptibility of AIP-I and FlagHis$_6$-AgrD-I(1-32)-thiolactone to ring-opening thiolysis by NAC to give the respective linear thioesters (Figure 5.8a). Equilibration between thiolactone and linear thioester species was monitored at a series of NAC concentrations by RP-HPLC/MS. Treating NAC as a ligand that reversibly binds to the thiolactones, its dissociation constant, $K_{D, NAC}$, was calculated from the equilibrium positions. Importantly for the following discussion, we observed essentially unbiased partitioning of NAC between aqueous and 1-octanol phases at pH = 7.5 (logP = -0.03, Figure 5.8b). Thus, the effective concentration of NAC should stay constant across the bulk solution and the membrane environment.

To begin, we treated AIP-I with NAC ranging from 1.0 to 100 mM in aqueous solution buffered at pH 7.0 (Figure 5.8c, circles, dotted line and Table 5.1). The $K_{D, NAC}$ for AIP-I is 5.8 mM. Ring closure of AIP-I from its linear thioester is therefore less favorable than we estimated for the formation of an unstrained thiolactone, for which the equilibrium constant would be 24 mM (see Figure 5.1b). Since AgrD-I(1-32)-thiolactone is insoluble in the buffer system used to establish the thiolysis equilibrium for AIP-I, we dispersed this molecule in a suspension of lipid vesicles in the presence of different NAC

concentrations. With a calculated $K_{D,\,NAC}$ of 95 mM, this thiolactone appears to be much more stable than AIP-I (Figure 5.8c, squares, solid line and Table 5.1). The less favorable

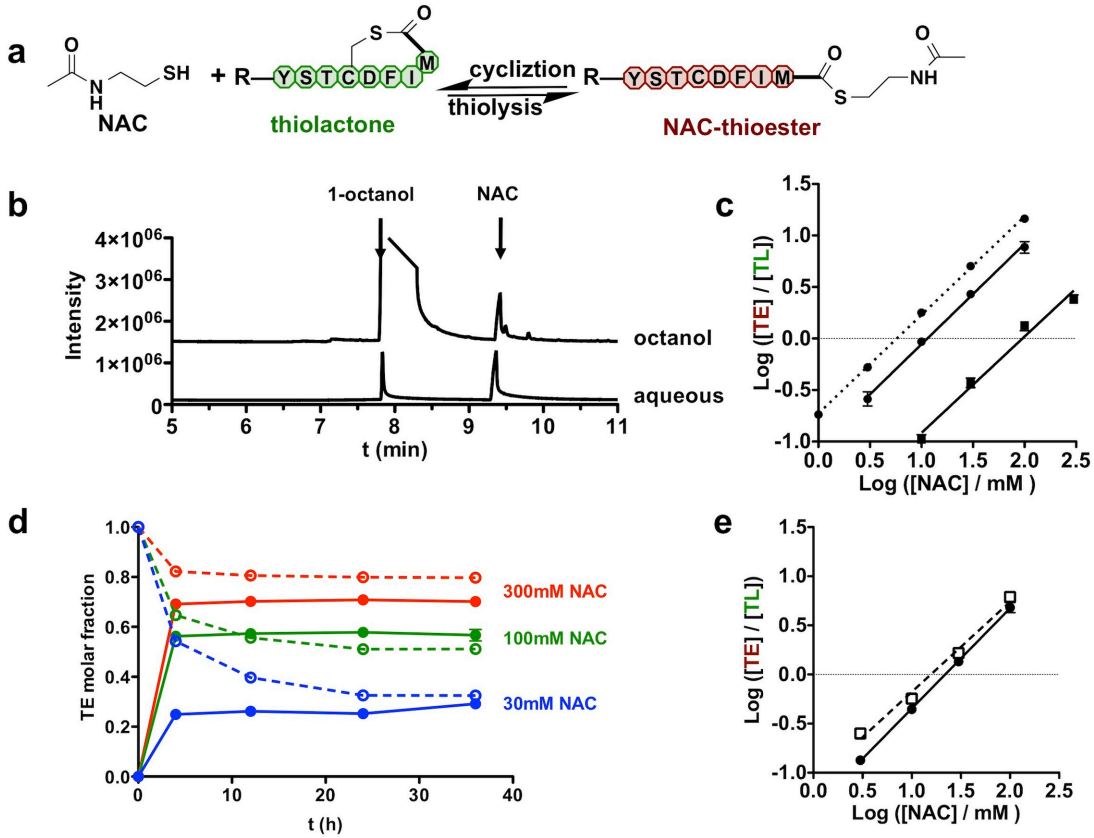


Figure 5.8 Ring-opening equilibrium of AIP-I and FlagHis$_6$-AgrD-I(1-32)-thiolactones.
(a) Schematic representation of the ring-opening equilibrium. **(b)** The octanol-water partitioning equilibrium of NAC: 100 μL NAC was added to 2 mL aqueous PBS buffer (pH = 7.5) and 2 mL 1-octanol and the system was allowed to equilibrate at RT. The panel shows GC-MS chromatograms of equivolume samples from the aqueous and 1-octanol phases. **(c)** Double-log plot showing equilibrium positions as a function of NAC concentration for AIP-I (circles) or AgrD-I(1-32)-thiolactone (squares) with the absence (dotted line) or presence (solid lines) of phospholipids. Error bars = SD (n = 4). TE, thioester; TL, thiolactone. **(d)** Time course of FlagHis$_6$-AgrD-I(1-32)-NAC ring closure in HBS buffer, pH = 7.0 containing 1.5 mM POPC and 0.5 mM POPG in the presence of indicated level of NAC. Plot shows the molar fraction of the thioester (TE) as a function of time (dashed lines). The ring-opening time courses (solid lines) under each NAC concentration are overlaid. **(e)** As in **(c)**, equilibrium positions were plotted for the AIP-I (closed circles, solid line) and FlagHis$_6$-AgrD-I(1-32)-thiolactone (open squares, dashed line) in 6M GuHCl buffered at pH = 7.0. For **(c)** and **(e)**, Error bars = range (n = 2) and are not visible due to the size of data symbols for most data points.

ring opening is not an effect of hysteresis, as reactions starting from the purified AgrD-I(1-32)-NAC thioester reached roughly the same equilibrium positions (Figure 5.8d). Intriguingly, the stability of AIP-I towards ring opening was only slightly enhanced (1-fold) in the presence of lipid vesicles (Figure 5.8c, circles, solid line and Table 5.1), despite the fact that it shares the identical macrocycle with AgrD-I(1-32)-thiolactone. It should also be noted that both thiolactones were equally susceptible to ring opening thiolysis in aqueous buffers containing chaotropic agents (Figure 5.8e and Table 5.1). Thus, the remarkable disparity in stability must originate from differential lipid-bilayer association of the macrocycle moiety. AgrD is thought to be targeted to the lipid bilayer by its N-terminal, amphipathic leader peptide (56). We confirmed this targeting effect using a co-migration assay on SEC, in which all AgrD constructs containing the N-terminal leader peptide, showed strong interaction to bilayer nanodiscs (Figures 5.9a-c). By contrast, AIP-I does not co-migrate with nanodiscs (Figures 5.9d and e). Based on this, we speculate that the membrane targeting property of the leader peptide places the macrocycle into proximity with the lipid bilayer and hence induces its membrane partitioning and thereby stabilizes it against ring opening thiolysis.

Table 5.1. Ring-closure equilibrium constants of thiolactones

Thiolactone	Condition*	$K_{D,\,NAC}$ (mM)
AIP-I	HBS	5.8 ± 0.2
	HBS, 2 mM POPC-POPG	11.4 ± 1.1
	6M GuHCl	21.9 ± 1.4
AgrD-I(1-32)	HBS, 2 mM POPC-POPG	95 ± 31
	HBS, 2 mM DGDG	56 ± 7
	6M GuHCl	15.6 ± 2.2
AIP-I tri-Ala**	HBS	8.0 ± 0.4
AgrD-I(1-32) tri-Ala**	HBS, 2 mM POPC-POPG	11.0 ± 0.4

* HBS, HEPES-buffer saline pH = 7.0; the POPC-POPG molar ratio is 3:1;

** Constructs bearing mutations F30A/I31A/M32A (in AgrD-I numbering).

To explore the molecular determinants of this putative partitioning interaction, we examined atomic-resolution structures available for *S. aureus* AIPs. We noticed that in a

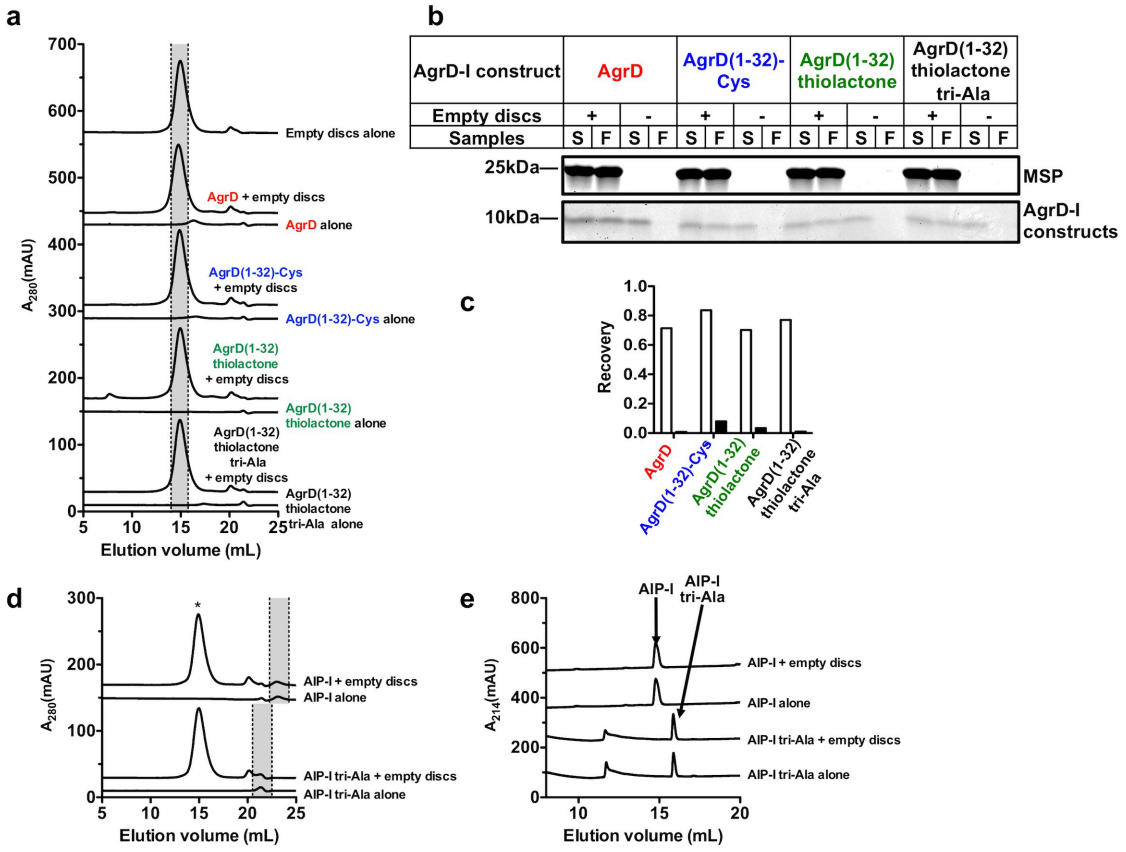


Figure 5.9. Nanodisc co-migration assays for AIP-I and AgrD-I constructs.
(a) SEC chromatograms: SEC sample (250 μL) contained 12 μM appropriate AgrD peptide and/or equimolar bilayer nanodiscs as indicated. The bilayer discs consisted of POPC-POPG in a molar ratio of 3:1. Shadow indicates the position at which nanodiscs (unbound or bound to AgrD) are eluted. FlagHis$_6$-AgrD(1-32)-Cys is a linear peptide prepared from treating the FlagHis$_6$-AgrD(1-32)-thiolactone with cysteine (see section 7.3.3), and was used as the surrogate of FlagHis$_6$-AgrD(1-32) in this experiment. **(b)** Recovery of AgrD peptides within eluate containing nanodiscs: for each sample containing AgrD peptide, the pre-SEC sample (denoted as **S**) and the combined fractions from the shadowed segment in **(a)** (denoted as **F**) were analyzed by SDS-PAGE. **(c)** Quantification of the gel in **(b)**. Bar graph depicts the recovery of indicated AgrD peptide from the shadowed segment of elution after resolved on SEC alone (close bars) or in the presence of bilayer nanodiscs (open bars). **(d)** SEC chromatograms of the co-migration assay of indicated AIP-I peptide. Positions where nanodiscs and AIP-I were eluted are indicated by asterisk and shadow, respectively. **(e)** RP-HPLC traces showing the recovery of either AIP-I peptide from the shadowed segment of elution in **(d)** to be unaffected by the presence of nanodiscs in the SEC samples.

crystal structure determined for a monoclonal antibody in complex to AIP-IV (174), three endocyclic AIP amides donate hydrogen bonds to a glutamate residue of the antibody, while side chains of residues Phe6, Ile7 and Met8 (all shared between AIP-IV and AIP-I) are buried in a large, contiguous hydrophobic cleft on the antibody surface (Figures 5.10a

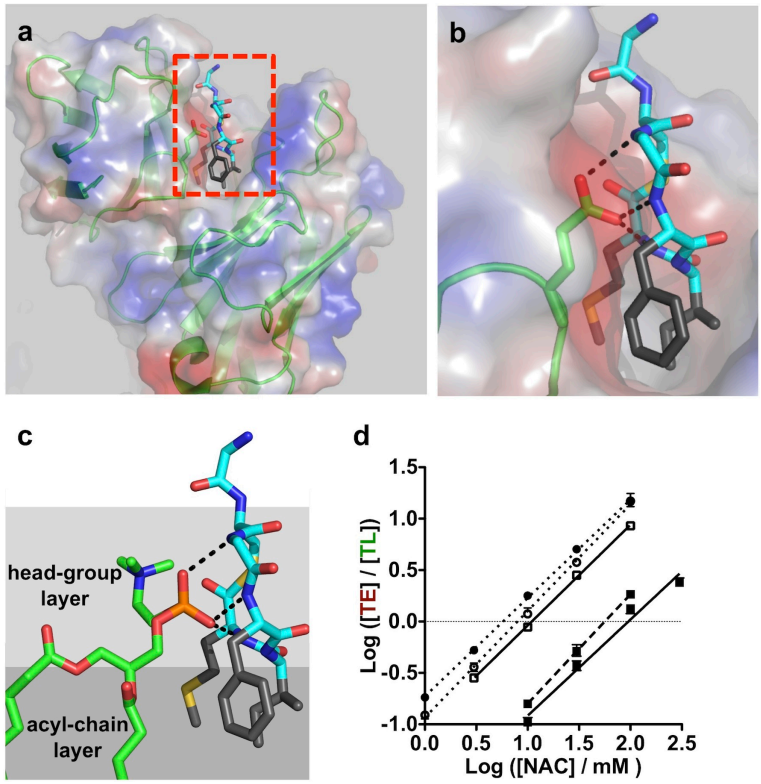


Figure 5.10 Hydrophobic side chains on the ring-forming residues are required for the liposome-dependent stabilization of the thiolactone.
(a and **b)** AIP-IV-antibody interface in a crystal structure (PDBID: 3QG6): An overview (panel **a**) and a zoom-in view at the interface (red rectangle in **a**, panel **b**) are shown. AIP-IV is presented as a stick model with the backbone carbons in cyan and side-chain carbons of Phe6, Ile7 and Met8 in gray. Side chains of Tyr1, Ser2, Thr3 and Asp4 are not shown for clarity. The ribbon diagram and electrostatic surface of the antibody are displayed. In **(b)**, Hydrogen bonds between a glutamate residue in the antibody (highlighted in green) and the AIP backbone are highlighted by black dashes. **(c)** Arbitrarily positioned stick models of AIP-I and POPC depicting the proposed interactions by which lipid vesicles stabilize the macrocycle: carbon atoms in POPC, are shown in green, and hydrogen bonds are highlighted by black dashes. Note that all displayed side chains in the AIP-I model are identical in AIP-IV. **(d)** As in Figure 5.8c, equilibrium positions were plotted for the mutant (open symbols) or wild type (closed symbols) AIP-I (circles) or AgrD-I(1-32)-thiolactone (squares) in aqueous buffer alone (dotted lines) or that containing phospholipids (solid lines) or glycolipids (dashed lines).

and b). Based on this crystal structure, we envisioned that the putative partitioning of the macrocycle in the membrane might be driven by analogous interactions, specifically: the anionic phosphodiester motif in lipid head groups might form hydrogen bonds to the AIP endocyclic amide groups and hydrophobic side chains could be embedded in the greasy interior of the bilayer (Figure 5.10c). In line with this model, mutating all three endocyclic hydrophobic residues to Ala in AgrD-I(1-32)-thiolactone abolished stabilization of the macrocycle in the presence of liposomes (Figure 5.10d and Table 5.1), even though this mutant still associated with membranes, presumably through its leader sequence that remains intact (Figures 5.9a-c). Indeed, this mutant exhibited thiolysis equilibrium positions similar to that of AIP-I, either the wild type peptide or that bearing the triple-Ala mutant macrocycle (Figure 5.10d and Table 5.1). In addition, liposomes primarily composed of a non-charged glycolipid, digalactosidyldiacylglycerol (DGDG) conferred slightly weaker stabilization to wild-type AgrD-I(1-32)-thiolactone than did POPC-POPG liposomes (Figure 5.10d and Table 5.1). Collectively, these observations are consistent with a role for both hydrophobic and polar interactions in the stabilization of the thiolactone macrocycle as a consequence of membrane targeting.

5.6 AgrD processing by AgrB is a reversible process *in vivo*

Finally, we asked whether the reversibility of proteolytic cyclization observed in our *in vitro* reconstituted system had relevance to AIP production in *S. aureus* cells. To this end, we constructed *S. aureus* plasmids for production of two AgrD-I^{C} peptides, one with wild-type sequence and the other a scrambled sequence. These AgrD-I^{C} peptides were fused to a Flag-tagged ubiquitin (Ub-$Flag_2$) for detection purposes (Figure 5.11a). Both

plasmids were introduced to a *S. aureus* strain carrying an *agr*-I locus, resulting in two transformants with similar growth rates (Figure 5.11b). Immuno-blotting revealed the presence of AgrD-I-Ub-Flag$_2$, the expected ligation product between the AgrD-I^C

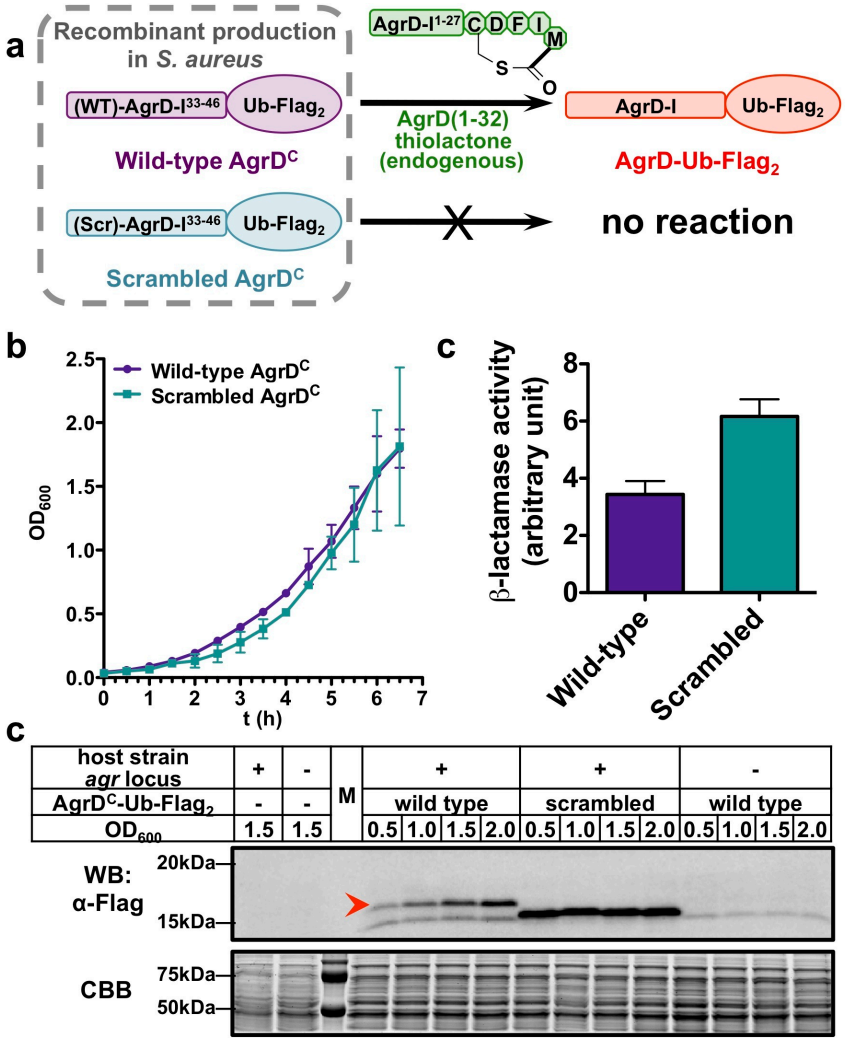


Figure 5.11. Reversibility of the proteolytic cyclization in vivo.
(a) Schematic showing the experimental design: production in *S. aureus* of wild-type AgrD-I^C peptide (purple), but not the scrambled control (teal), is expected to give rise to AgrD-I-Ub-Flag$_2$ (red) through the reverse activity of AgrB-I. **(b)** Growth curves of *agr*-I cells producing indicated AgrD-I^C peptides at 30^{o}C. Error bar = range (n = 2). **(c)** Western-blot analysis of cells producing AgrD-I^C peptides: cells with indicated genotypes were grown at 30 ^{o}C to indicated optical density (OD_{600}) prior to lysis. A high-molecular-weight slice of the gel was Coomassie-stained as a loading-control. **(d)** Bar graph showing AIP-I activity in the medium isolated from *agr*-I cells producing indicated AgrD-I^C peptides determined by an *agr*-I-specific reporter gene assay. Error bar = SD (n = 6).

construct and endogenous AgrD-I(1-32)-thiolactone, in cells expressing the wild-type construct (Figure 5.11c, lanes 4-7, red arrowhead) but, importantly, not the scrambled construct (Figure 5.11c, lanes 8-11). As expected, this ligation product was not present in transformed *agr*-knockout cells expressing the wild type AgrD-I^C construct (Figure 5.11c, lanes 13-15). Remarkably, AIP-I secretion was significantly lower in *agr*-I cells expressing the wild type AgrD-I^C construct as compared to those expressing the scrambled construct, as determined using an *agr*-I-specific reporter gene assay (Fig. 5C). We interpret this as resulting from sequestration of AgrD-I(1-32)-thiolactone as a consequence of AgrB-I-mediated ligation with wild-type AgrD-I^C-Ub-$Flag_2$, but not the scrambled control.

5.7 Summary and Discussion

5.7.1 Factors that facilitate AgrD proteolytic cyclization

When a polypeptide undergoes proteolysis, the peptidyl fragment N-terminal to the scissile bond is transferred to an acceptor nucleophile. While water is by far the most frequent acceptor, nature also efficiently harnesses a spectrum of *N*- or *O*-nucleophiles, leading to the formation of various branched, cyclic or C-terminally modified peptides or proteins (175). To the best of our knowledge, the first step in AIP biosynthesis is the only known proteolytic event where an *S*-nucleophile serves as the final acceptor of the peptidyl group – this definition excludes the transient, thioester-linked acyl-enzyme intermediates associated with the activity of cysteine proteases. It is noteworthy that nature typically couples ATP hydrolysis to the conversion of a carboxyl group (via an adenylation step) into a thioester, e.g. the activity of the ubiquitin activating enzyme, E1

(176). In this study, we broke down the thiolactone-forming proteolysis reaction into the following half-reactions; (1) thiolysis of AgrD at the scissile peptide bond, and (2) ring closure from the resulted linear thioester (Fig. 1B). Employing chemical equilibrium approaches, we experimentally determined the $\Delta G^{0'}$ for the overall proteolytic cyclization reaction (7.1 kcal/mol) and the ring closure from the linear AgrD(1-32)-NAC thioester (1.4 kcal/mol, calculated from $K_{\mathbf{D,\ NAC}}$ = 95 mM). The former reaction is 2.6 kcal/mol more favorable than our estimation based on model reactions, whereas the latter is more favorable by 0.9 kcal/mol. This comparison indicates that the initial thiolysis step is also more facile (by 1.7 kcal/mol) than might be expected. To account for this, we propose that the scissile bond in AgrD is more sensitive to proteolysis than a generic peptide bond. Indeed, it has been shown that the free energy of hydrolysis for an individual peptide bond may vary by a few kcal/mol as a function of sequence context in folded proteins (177). Conceivably, targeting of AgrD to the lipid bilayer by the N-terminal leader peptide could also help sensitize the scissile amide to proteolysis due to changes in the local chemical environment around this bond (e.g. changes in the dielectric of the medium, electrostatic effects and so on). Additional experiments will be required to test this idea.

5.7.2 Physiological implications of the $\Delta G^{0'}$ of thiolactone formation

Autoinduction of the *agr* locus in *S. aureus* features an abrupt up-regulation of the effector RNA levels during the exponential growth phase (29). A recent report showed that the pheromone concentration in a liquid-medium culture of a group-I strain reaches 5 μM in two hours (178). By definition, this generates the same number of $AgrD^{C}$

fragments in the cytoplasm, which leads to a considerably elevated intracellular concentration because of the much smaller volume of the cytoplasm compared to the medium. For example, AgrD^{C} production would reach an accumulative concentration of 5 mM in two hours, at a rate of 700 nM per second, assuming 1 mL cytoplasm is present in a one liter cell culture in late exponential phase (179). Constant and efficient AIP production therefore necessitates simultaneous removal of AgrD^{C} since, as we have shown, elevating the levels of this fragment can perturb the equilibrium of AgrB-mediated proteolytic cyclization (Fig. 5C) (Figure 5.12). Assuming that equimolar amounts of AgrD and AgrD(1-32)-thiolactone are present in a quasi-equilibrium with AgrD^{C} in post-induction *S. aureus* cells, the steady-state concentration of AgrD^{C} should be equal to the equilibrium constant for the proteolytic cyclization, K_{eq} = 10 μM. The

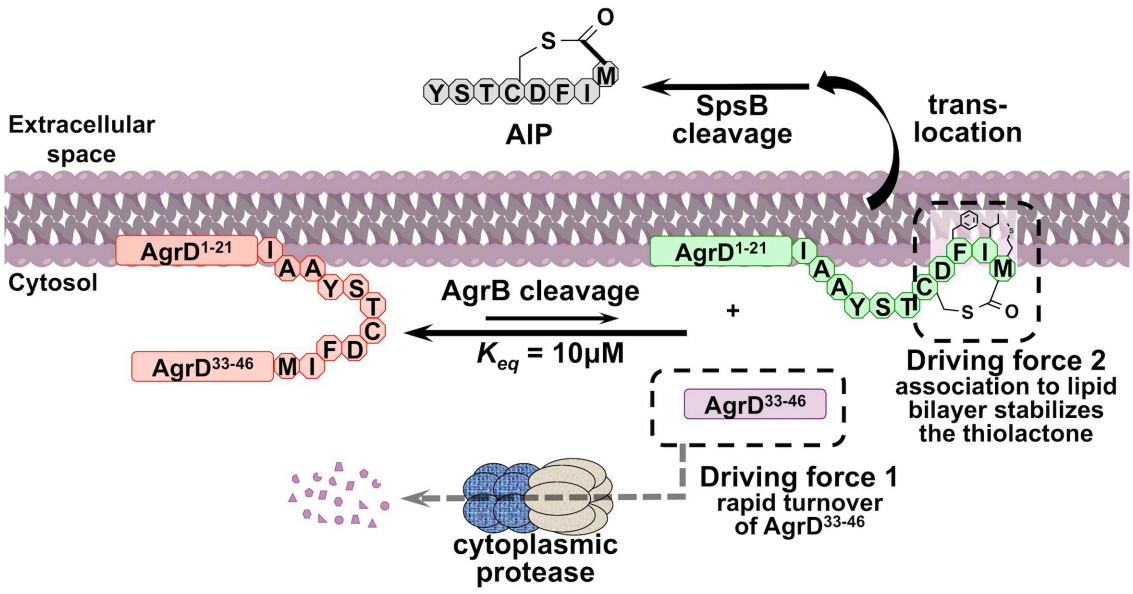


Figure 5.12 Summary of the driving forces that power the formation of the AgrD-(1-32)-thiolactone intermediate
The AAA+ protease proposed to degrade AgrD^{C}, ClpXP, is highlighted in a cartoon portraying its tertiary structure, with a hexameric ClpX ring stacking on two heptameric ClpP rings.

half-life of AgrD^{C} would therefore be on the order of 10 seconds given a production (and degradation) rate of 700 nM per second. This short lifetime of AgrD-I^{C} indicates a pathway in *S. aureus* that actively targets the peptide for degradation.

The above estimation of the *in vivo* lifetime of AgrD^{C} sheds light upon two long-standing puzzles with respect to the production of staphylococcal AIPs. The first of these involves the observation that AIP-production is greatly reduced as a consequence of genetic or chemical knockout of ClpP, the catalytic subunit of certain major AAA+ proteases in *S. aureus* (180, 181). While this phenotype has been known for more than a decade, the underlying mechanism has been elusive. We propose that ClpP constitutes the AAA+ protease that catalyzes the degradation of AgrD^{C} and thus fuels the proteolytic cyclization of AgrD (Figure 5.12). The second puzzle relates to the remarkable sequence conservation of AgrD^{C} (Figure 5.13). Surprisingly, mutations at most highly conserved positions on AgrD^{C} have no affect on the substrate recognition by AgrB (55). Given the necessity for rapid degradation of AgrD^{C} on *agr* induction, we hypothesize that a conserved motif within AgrD^{C} targets the peptide to the degradation machinery, either through direct binding or mediated by a dedicated adaptor protein. Experimental tests of this hypothesis are currently in progress.

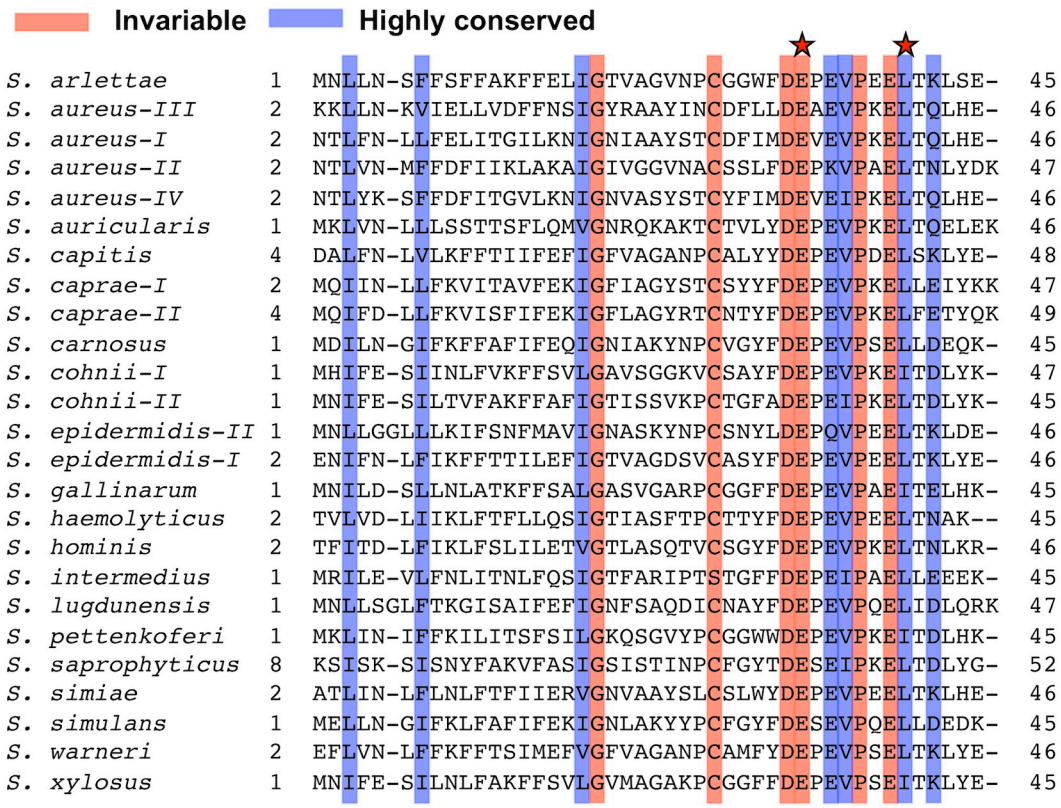


Figure 5.13 Multiple sequence alignment of Staphylococcal AgrDs.
All AgrD sequences from the genus *Staphylococcus* in the pfam database (05931) are subject to an identity cutoff at 80%. The resulting 25 non-redundant sequences were aligned to the hidden-Markov model of the AgrD family. Invariable and highly conserved positions are highlighted in red and blue, respectively. Two red stars indicate positions at which mutation to alanine abolishes AgrB-catalyzed processing *in vivo*.

Chapter Six: Outlook

The work presented in this dissertation established *in vitro* reconstitution systems for the quantitative biochemical and biophysical characterization of two pivotal membrane proteins, namely, AgrC and AgrB, involved in the QS in *S. aureus*. Prior to our study on AgrC, RHKs are in general believed to work as two-state switches that can be activated only with highly specific conformational input. The discovery of a gradual conformation-activity landscape for the AgrC-I HK domain employing the GCN4-AgrC-I chimera protein series raises an alternative model in which the signal-response properties of TCSs are more amenable to fine adjustments in evolutionary processes. Furthermore, guided by our thermodynamic analysis of the AgrB-mediated proteolytic cyclization of AgrD, we found that constant and efficient degradation of the C-terminal cleavage fragment, $AgrD^{C}$, is required to power the formation of the high-energy thiolactone N-fragment. In this chapter, opportunities to further explore the AgrC conformation-activity landscape are discussed and experiments are proposed for the identification of the protease(s) responsible for the turnover of $AgrD^{C}$. At the end of this chapter, a chemical tagging approach harnessing the reactivity of the AIP thiolactones to identify AIPs from other *Firmicutes* species is described.

6.1 The signal-response property of the AgrC-I HK domain

In Chapter 4, the conformation of the interdomain helical linkers of AgrC-I has been subject to symmetric rotational input imposed by a GCN4 coiled-coil fused to the N-terminus of the linkers. The auto-kinase activity changed gradually with the magnitude of the rotation, leading to the conclusion of a gradual conformation-activity landscape

(Figure 4.3). This understanding of the signal-response behavior of AgrC-I, however, is rather preliminary. Firstly, the auto-kinase activity was used as the sole readout in our analysis of the GCN4-AgrC-I chimera series, so there is no information concerning the response of, for instance, the trans-kinase (phospho-relay) activity to the rotation of interdomain linkers. Importantly, AIP-II, the inverse agonist of AgrC-I has been found to inhibit the trans-kinase activity while the underlying mechanism remains unknown. Secondly, given the vast diversity of sensor domains harnessed by RHKs in nature, it is very unlikely that they all impose symmetric, rotational input to the interdomain linkers that connect them to their respective HK domains. A fascinating project would therefore involve the systematic introduction of other types of conformational input to the AgrC-I linker pair followed by the analysis of the auto-kinase response. Thirdly, our inter-subunit disulfide crosslinking experiment on full-length AgrC-I favored the model in which the interdomain linker adopts a distinct, static conformation in each ligand state rather than distributes between two discrete conformations (Figures 4.4d and 4.9). This interpretation, however, needs to be restricted to the N-terminal half of the linker whereby Cys substitutions were performed (residues 205-209 out of 201-211). The HK domain, to which the C-terminus of the linker connects, on the other hand, is believed to distribute between as few as two conformations that corresponds to the kinase-off and kinase-on states captured in crystal structures (150). Thus, the interdomain linkers must somehow reconcile the discrepancy between the static conformation at its N-terminus and the conformation distribution at its C-terminus, by a yet unclear mechanism. Addressing the above three questions will significantly expand our knowledge on the signaling mechanism of TCSs.

6.1.1 The trans-kinase activity and the pHis stability in GCN4-AgrC-I chimeras

As was alluded to in section 4.5.2, the AgrC-AgrA TCS would be exceptionally amenable to the analysis of phospho-relay kinetics because AgrC lacks the phosphatase activity that may otherwise complicate such analysis (see Figure 3.5). In reality, however, one drawback of this system comes from the instability of AgrA: all recombinant AgrA proteins containing the phospho-acceptor REC domain are prone to precipitation even when presented at low-μM concentrations (data not shown). Moreover, while the trans-kinase kinetic data of all GCN4-AgrC-I chimeras are required to plot a conformation-activity landscape, direct reconstitution of the phospho-relay involving the auto-kinase-inactive chimeras is impractical because they cannot be pre-phosphorylated efficiently. To circumvent these issues, a three-component, phospho-equilibration system could be designed (Figure 6.1a), which harnesses the well-known thermodynamic reversibility of the His-to-Asp phospho-relay (84). This system contains equimolar quantities of a donor and an acceptor AgrC constructs, as well as a catalytic amount (less than 5% of the donor) of AgrA that shuttles the phosphoryl group between them (Figure 6.1a). Specifically, the phosphoryl donor should be a kinase-active AgrC-I construct that is phosphorylated to completion and purified from the auto-kinase reaction (so that it is nucleotide-free) and the acceptor is the subject GCN4-AgrC-I chimera. Assuming that the kinetics of the forward (AgrC to AgrA) and reverse (AgrA to AgrC) phospho-relays are much faster than the spontaneous dephosphorylation of AgrA (Figure 6.1a), this system is expected to reach a quasi-equilibrium state and the phosphorylation levels of the donor and the acceptor will approximate a thermodynamic equilibrium (Figure 6.1b, the equilibrium phase). Notably, the position of this equilibrium will reflect the

thermodynamic pHis stability on the subject chimera relative to the donor. These data, if successfully obtained, will reveal the dependence of the pHis stability on the conformation of the interdomain linkers. Meanwhile, phosphorylation of the catalyst, AgrA, is expected to reach a steady state shortly after the reaction begins (Figure 6.1b, the equilibration phase). The kinetic constant of the phospho-transfer from AgrA to the subject chimera will therefore be determined based on the reaction velocity and the steady-state phosphorylation level of AgrA. Kinetic parameters of the chimera-to-AgrA transfer will then be derived from the quasi-equilibrium position at the equilibrium phase.

Finally, it is important to point out that the faster kinetics of phospho-relays versus

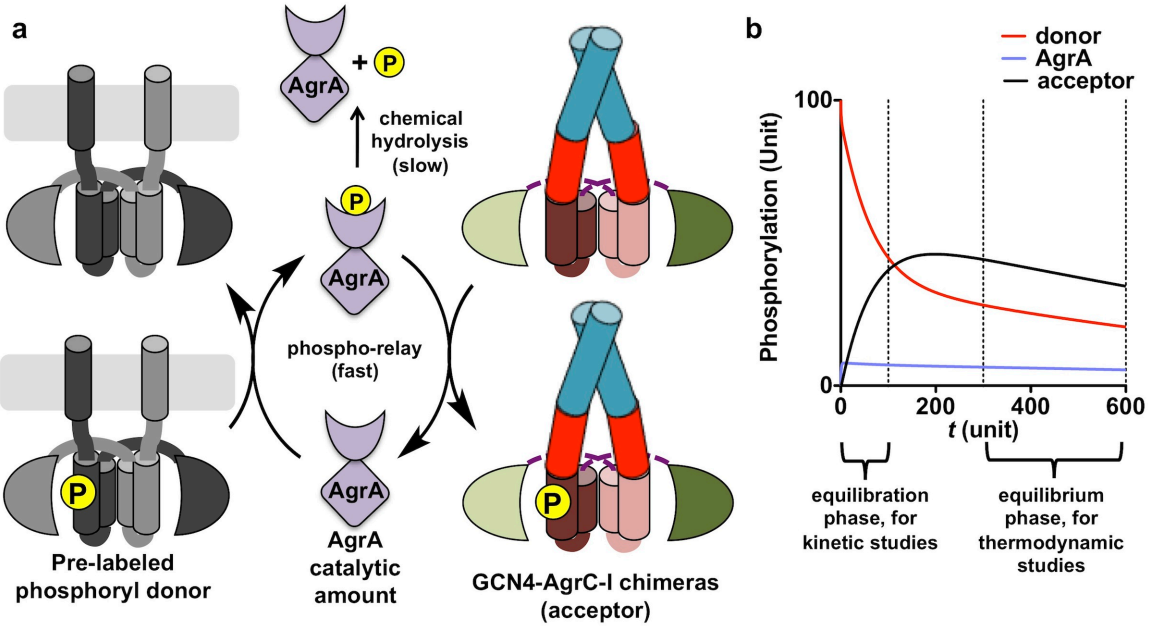


Figure 6.1 The three-component phospho-equilibration system
(a) The reaction scheme. See text for details. **(b)** A simulated kinetic plot depicting phosphorylation levels of all three components as a function of time. The phosphorylation of two His residues in each HK dimer are treated as independent events. In panel **(b)**, The simulation begins with an equimolar amount of the donor and the acceptor. The parameters are arbitrarily set such that the pHis in the acceptor is $0.3k_B$ more stable than that in the donor.

AgrA dephosphorylation and the AgrA stability in solution are both critical to the predicted behavior of this phospho-equilibration system. Preliminary experiments should be performed to confirm these prerequisites.

6.1.2 Introduction of scissoring motions to the AgrC-I interdomain linker helices

Geometrically, modifying the distance between the N-termini of the interdomain linkers will cause the scissoring motions (Figure 6.2a, top panel). Thus, a conformation-activity

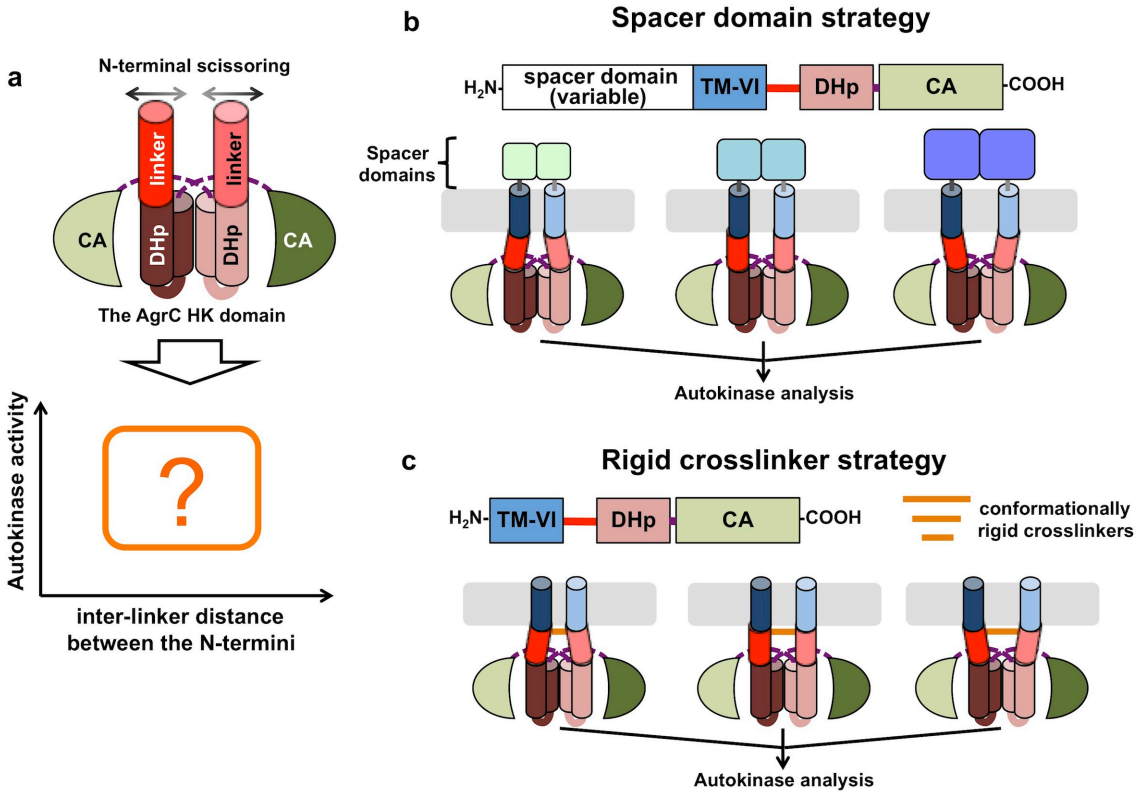


Figure 6.2 Strategies to introduce scissoring movements to the helical interdomain linkers.
(a) Color codes of AgrC-I intracellular structural elements (top panel) and the definition of the conformation-activity landscape about scissoring movements of the linkers (bottom panel). The conformation descriptor could be the inter-linker distance between any two positions close to the linker N-terminus. **(b)** The spacer-domain strategy: steric bulks of the spacer-domain dimers separate the N-termini of the TM domains by a series of distances, which is expected to translate into different scissoring conformations of the interdomain linkers. **(c)** The rigid-crosslinker strategy: the N-termini of the interdomain linkers are directly separated by a crosslinker acting as a molecular ruler.

landscape with such motions could be plotted in the form of auto-kinase activity versus a distance descriptor (Figure 6.2a, bottom panel). Modification of inter-linker distance can be achieved employing two strategies. The first strategy will rely on the use of homodimeric protein domains, herein referred to as spacer domains (Figure 6.2b). These domains are fused to the N-terminus of an AgrC-I construct to separate the interdomain linker pair by their steric bulk (Figure 6.2b). The fusion protein will also consist of the last TM domain (TM-VI) and the cytosolic portion of AgrC-I. TM-VI is included due to its critical role in the activation of the constitutive mutant, AgrC-I^{R238H}, which presumably occurs through promoting α-helicity in the interdomain linker (see Figures 4.1b and 4.7). Spacer domains are selected on two criteria: (i) that they homodimerize with mid-nM or higher affinity and (ii) that the C-termini in the homodimer structure are separated by a unique distance (Figure 6.2b). To chart the landscape, a series of spacer-AgrC-I chimeras will be prepared recombinantly, reconstituted to nanodiscs for auto-kinase analysis, and the distance between the interdomain linkers will be gauged through Förster resonance energy transfer (FRET)-based approaches. The second strategy will employ chemical crosslinkers, each containing two reactive groups separated by a conformationally rigid linker (Figure 6.2c, orange bars). These crosslinkers will be installed between the two subunits at an interdomain-linker position and are expected to separate the crosslinked positions by a defined distance (Figure 6.2c). For the same aforementioned reason, the AgrC-I construct used here will also include the TM-VI. Crosslinking may be achieved directly on the recombinantly prepared AgrC-I construct. Alternatively, the crosslinked protein could be prepared from the expressed chemical ligation (EPL) of two fragments: a synthetic N fragment containing the crosslinking site and a recombinant C-fragment. Crosslinkers will be installed to the N-fragment prior to the EPL reaction. Crosslinked AgrC-I dimers prepared in either way will be refolded and

incorporated to nanodiscs for autokinase analysis. The distance descriptor of scissoring will be calculated directly from the chemical structure of the crosslinkers.

6.1.3 Is the interdomain linker pair an analog-to-digital converter?

As was discussed in Section 4.5.1, the interdomain linker needs to transmit the conformational input imposed at its N-terminus to an output discernible for the HK domain. In the full-length AgrC-I bound to a certain ligand or a GCN4-AgrC-I chimera, the input conformation should be constant over time (Figure 6.3a). This input fits the definition of an analog signal in that it is a continuous function of time (see Figure 6.3c). By stark contrast, the HK-domain dimer to which the C-termini of the linker pair connect is believed to distribute between as few as two discrete conformational states (Figure 6.3b), each exhibiting a distinct level of auto-kinase activity (150). This notion is in part based on the packing principle of four-helix bundles, in which gradual rotation or tilting of the participating α-helices is not permitted (182). The auto-kinase regulation would therefore depend on the redistribution of the HK-dimer ensemble between/among its possible conformations (Figure 6.3b). At the single-molecule level, however, the HK domain conformation is expected to be a discrete function of time—a property reminiscent of a digital signal (see Figure 6.3c). In this regard, the interdomain linker pair acts as an analog-to-digital converter.

As a future endeavor of the Muir Lab, the above model will be further elucidated employing single-molecule FRET strategies (183), which have, to our best knowledge, not been applied to any RHK to study the conformational ensemble of its participant modules. The crystal structure of at least one GCN4-AgrC-I chimera is required to direct

the placement of fluorophores, and this work is currently in progress. To confirm our assumptions on the conformational states of the GCN4 coiled coil and the DHp four-helix bundle, the donor-acceptor pair will first be introduced to these modules and the FRET signal will be recorded as a function of time (Figure 6.3c). If the FRET data match our

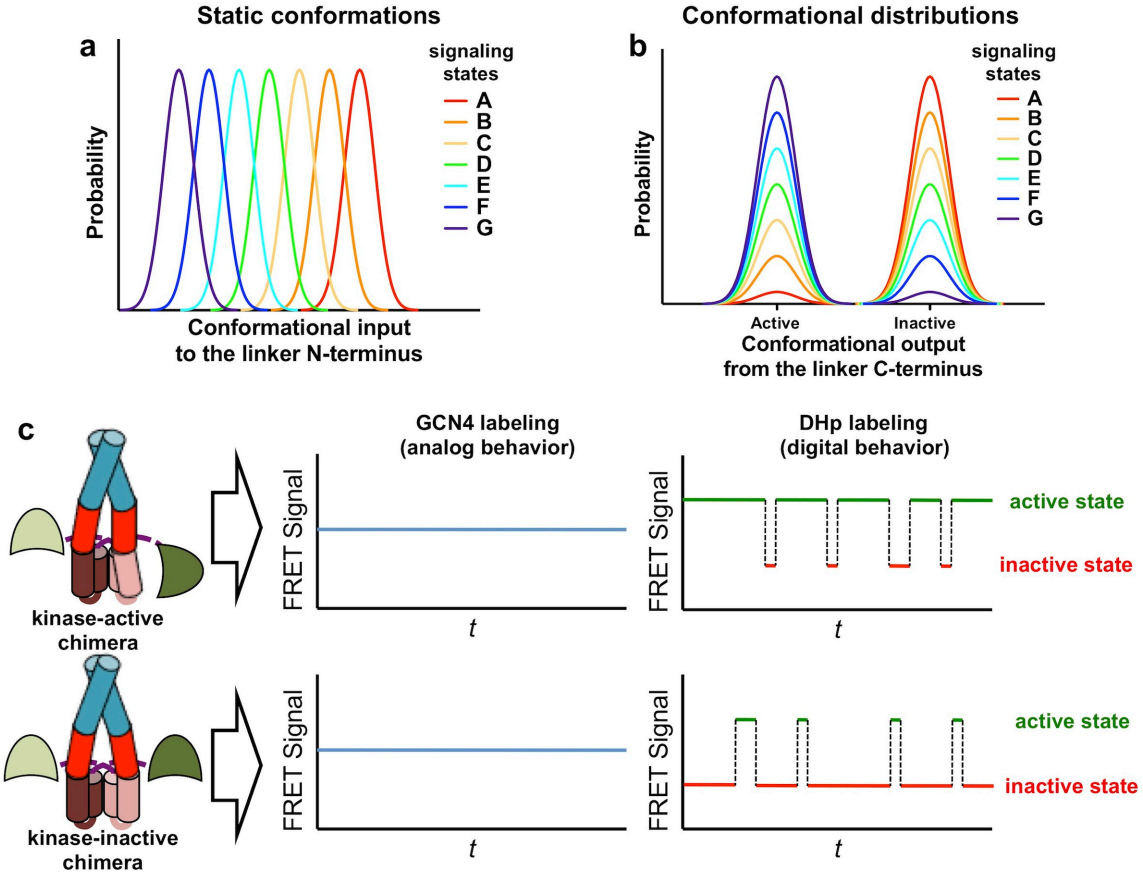


Figure 6.3 Putative conformational distributions of the two modules flanking the interdomain linker
(a) The static conformation introduced by the rigid-body GCN4 module. **(b)** The conformation distribution discernible by the four-helix DHp module. In panels **(a)** and **(b),** conformation distribution of seven hypothetical signaling states, A through G, are shown. In **(b)**, the output is arbitrarily distributed between two conformations corresponding to kinase-active and –inactive states, respectively. **(c)** Predicted FRET trace for two GCN4-AgrC-I chimeras, one kinase-active (top panels) and the other inactive (bottom panels), given the FRET pair is installed in the GCN4 (left plots) or DHp modules (right plots).

expectation on their respective analog and digital behaviors, a few more GCN4-AgrC-I chimeras exhibiting distinct auto-kinase activities will then be analyzed using the same sets of labeling sites (Figure 6.3c). Notably, the apparent kinase activity of a chimera should be equal to the average of all its major conformers weighed by their distribution probabilities. In this regard, the activity level of major DHp conformers could be solved employing simultaneous equations. If necessary, fluorophores may be moved to the N- or C- termini of the interdomain linkers to further elucidate the conformational states at these positions. Furthermore, once an input-output correlation of the wild-type linker is established, its native length or sequence could be altered through mutagenesis in order to investigate the influence of the perturbed factors on the linker's signal transmission property.

6.2 Degradation of the AgrD C-terminal cleavage fragment

As was discussed in section 5.7.2, the ClpXP AAA+ protease machinery is a favorable candidate for the degradation of $AgrD^{C}$. Several experiments are proposed to test this idea. Firstly, the $AgrD^{C}$-Ub-$Flag_2$ construct (section 5.6, Figure 5.11a) will be over-produced in the wild-type *S. aureus* or the $clpP^{S98A}$ mutant that expresses a catalytically inactive version of ClpP (Figures 6.4a and b). Assuming that $AgrD^{C}$-Ub-$Flag_2$ is also actively degraded by ClpXP, a stronger accumulation of this $AgrD^{C}$ construct should be seen in the mutant. Consequently, an increase of $AgrD^{FL}$-Ub-$Flag_2$ and a decrease of AIP production should also be expected (Figure 6.4b). In addition, the recombinant *S. aureus* ClpXP is expected to degrade the $AgrD^{C}$ peptide *in vitro* (Figure 6.4c). If both experiments support the role of ClpXP, efforts will be made to understand the recognition

of AgrD^{C} by ClpXP. Specifically, previous studies have shown that most conserved residues within AgrD^{C} are not required for substrate recognition by AgrB (55). Intuitively, these positions may conserve for the interaction that targets AgrD^{C} to ClpXP. In light of this idea, these positions will be Ala-scanned for mutations that confer resistance to ClpXP degradation (Figure 6.4d). Once such mutations are successfully identified, the role of the residue and its effect on the physical interaction with the AAA+ protease will be confirmed with immuno-precipitation (IP) studies (Figure 6.4e).

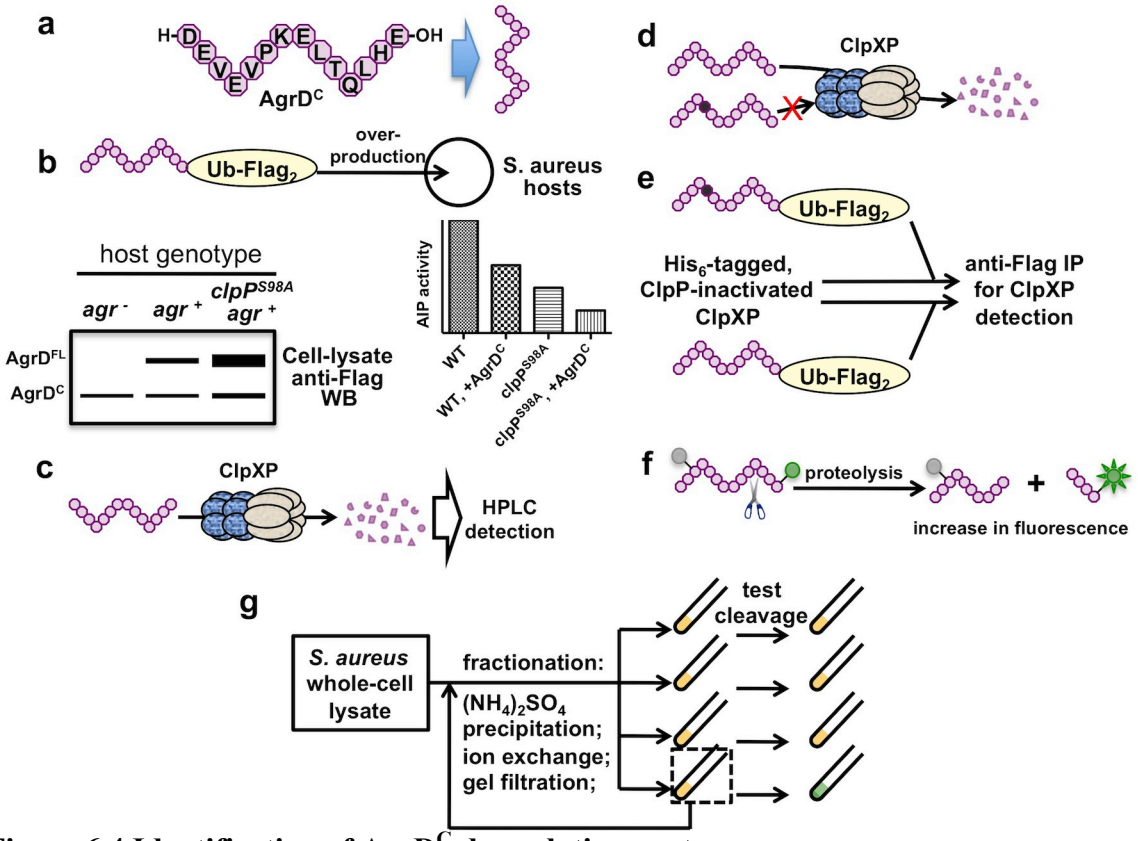


Figure 6.4 Identification of AgrD^{C} degradation proteases
(a) The sequence and cartoon presentation of AgrD-I^{C}. **(b** and **c)** Preliminary genetic **(b)** and *in vitro* biochemical **(c)** experiments to validate the AgrD^{C} degradation activity of ClpXP. See text for details. **(d** and **e)** The alanine-scan mutagenesis approach in search of AgrD^{C} residues essential for ClpXP targeting. The site of Ala-mutation is highlighted in black. **(f** and **g)** An unbiased strategy to enrich AgrD^{C}-degrading proteases through fractionation. In **(f)**, the fluorescein- and dabsyl- labels are indicated in green and gray, respectively. Any cleavage occurring between the two labeling sites is expected to generate green fluorescence.

While the ClpXP system is a strong candidate as an $AgrD^C$ protease, it is possible that other cellular proteases also contribute to the degradation of this peptide. Therefore, an unbiased, activity-guided fractionation strategy will be taken to identify them. Accordingly, we will use an $AgrD^C$ peptide modified with a fluorescein-dabsyl donor-quencher FRET pair as the substrate to follow the biochemical enrichment of protease activity in *S. aureus* cell-lysate fractions (Figures 6.4f and g). Once the protease activity has been suitably enriched, the responsible protease will be identified using mass spectrometry.

6.3 Detection of AIP thiolactones from other *Firmicutes* species

The phylum *Firmicutes* includes not only causative agents of various infective diseases but also a vast diversity of mutualistic symbionts of human beings. As a prime example, metagenomic analysis of human gut revealed that *Firmicutes* comprises nearly 40% of the bacterial population, and abnormalities in the abundance of key genus or species have been related to disease states (184). Intriguingly, homologs of the *agr* system can be identified in several prominent *Firmicutes* constituents in the gut microbiota (Figure 6.5a). Remarkably, a *Eubacterium rectale* strain possesses 7 AgrB homologs, each accompanied by a unique AgrD sequence sharing the prominent features required for AgrB recognition as well as the (thio)lactone formation (Figure 6.5a) (136). Inspired by the cross-group inhibitory effect of *S. aureus* AIPs, we hypothesize that the AIPs putatively produced by these AgrB-AgrD protease-substrate pairs are employed as weapons for the competition among *Firmicutes* species that shapes their ecological niche in the nutrient-limiting gut environment (185). Given the strong specificity of AgrC-AIP

recognition seen in *S. aureus*, the signaling of each putative AIP is expected to affect the fitness of a narrow spectrum of bacterial species. Thus, these AIPs may be harnessed as useful tools to probe the interaction among species or manipulate the composition of the microbiota in a phylotype-specific manner.

Because all AIP sequences are embedded in the middle of their AgrD precursors, it is necessary to determine the site of both proteolysis events required for their maturation. While the AgrB cleavage site can be identified with high confidence based on sequence

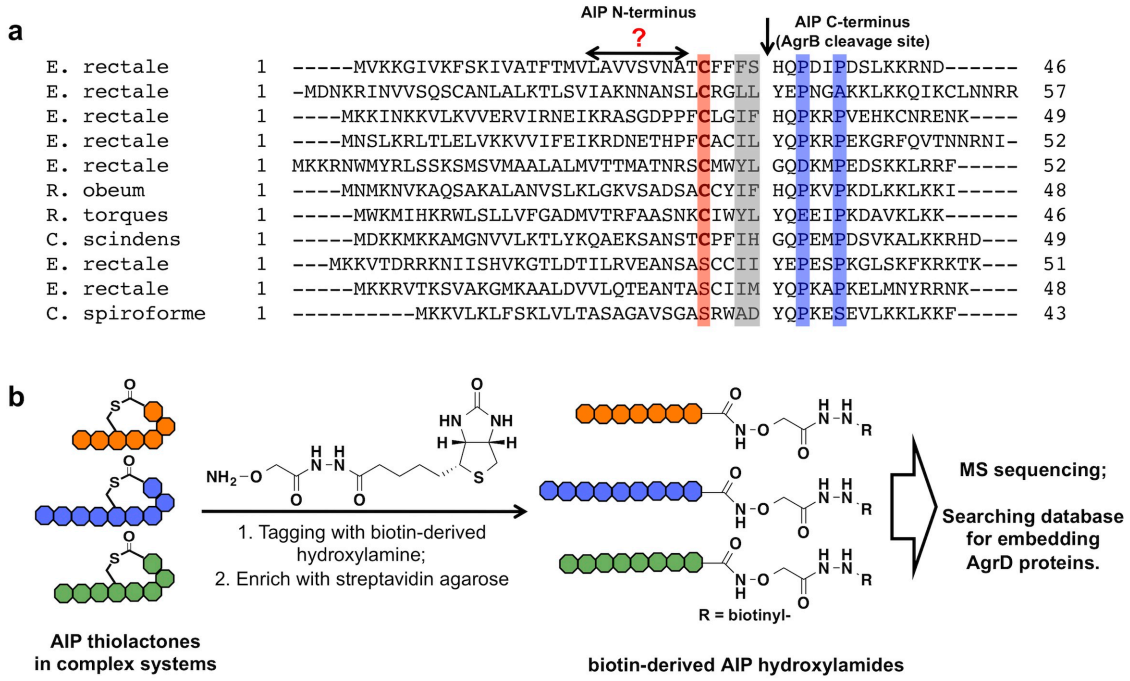


Figure 6.5 Identification of thiolactone AIPs from complex systems
(a) Gap-free sequence alignment of AgrD proteins from a few major *Firmicutes* species in human gut. The predicted AgrB cleavage site is shown with a blank in all sequences and indicated with an arrow. Three loci that defines the AgrB cleavage site, two enriched of Pro and one bearing the (thio)lactone-forming Cys or Ser are highlighted in blue and red, respectively. Two positions conserved for hydrophobic residues on the N-terminal side of the AgrB cleavage site are shadowed in grey. **(b)** A chemical tagging strategy that isolates AIP thiolactones from complex systems for MS sequencing. See text for details.

similarity (136), delineating the N-terminal cleavage junction is challenging due to its variability—this site is not conserved even among the closely-related *S. aureus* subgroups (refer to Figure 1.6a). To determine the exact sequence of AIPs produced by *Firmicutes* species in a complex system, a chemical-tagging approach harnessing the reactivity of the conserved thiolactone is proposed (Figure 6.5b). Thiolactone AIPs from human stool samples (or other complex systems) will be linearized by a biotin-tagged hydroxylamine and enriched with a streptavidin-agarose matrix. Peptides released from streptavidin will be sequenced based on mass spectrometry. A library of all AgrD genes encoded by the microbiome will be generated using the Human Microbiome Project database (hmpdacc.org/) and the sequences obtained for all tagged peptides will be mapped back to their host AgrD genes. It is noteworthy that, upon the ring-opening tagging reaction, AIPs will lose their strong proteolytic resistance conferred by the thiolactone macrocycle. In this regard, the samples will be pre-treated with protease-inhibitor cocktails, and casein will be included in the tagging and enrichment processes as a protective measure. Any AIP detected and sequenced in the above process will be synthesized and re-introduced to the complex system to dissect its impact on the fitness of its host species as well as other symbionts.

Chapter Seven: Methods

In this dissertation, we include the reconstitution of two disparate biochemical systems, namely the AgrC-AgrA TCS that mediates phospho-relay, and the AgrB peptidase that carries out proteolysis on AgrD. Nonetheless, protein production and reconstitution of these system shares considerable similarity, as the key component in both systems is integral membrane proteins. As a consequence, the molecular cloning and recombinant protein production methods for all four Agr proteins are described combined, while the biochemical assays on each system are documented separately.

7.1 Materials and Equipment

7.1.1 General Materials

All buffering salts, the LB broth, isopropyl-β-D-thiogalactopyranoside (IPTG), 1, 4-dithiothreitol (DTT), tris-(2-carboxyethyl) phosphine hydrochloride (TCEP), iodoacetamide, Coomassie brilliant blue, phenylmethylsulfonyl fluoride (PMSF) and nitrocellulous membranes were purchased from Fisher Scientific (Pittsburgh, PA). All medium ingredients for *S. aureus* cultures were from Difco Biosciences, Inc. (San Diego, CA). All antibiotics, non-radioactive nucleotides for AgrC-I-related assays, *N, N*-dimethylformamide (DMF), dimethylsulfoxide (DMSO), 2,2,2-trifluoroethanol (TFE), *N, N*-diisopropylethylamine (DIEA), piperidine, triisopropylsilane (TIS), β-mercaptoethanol (BME), N-Acetylcysteamine (NAC), lysostaphin, gluaraldehyde solution, guanidinium chloride (GuHCl) sodium 2-sulfanylethanesulfonate (MESNa), bovine serum albumin (BSA), and Eastman KODAK BioMax Film were purchased from Sigma-Aldrich (St. Louis, MO). All protected amino acids, coupling reagents and Rink-amide MBHA were

purchased from Novabiochem. [γ-^{32}P]-Adenosine triphosphate and Ultima Gold Cocktail was from PerkinElmer (Waltham, MA). All lipids were purchased from Avanti Polar lipids (Alabaster, AL) and all detergents were obtained from Anatrace (Maumee, OH). Trifluoroacetic acid (TFA) was from Halocarbon (North Augusta, SC). Anhydrous HF was from Matheson. Nickel-nitrilotriacetic acid (Ni-NTA) resin and KOD polymerase was from Novagen (Gibbstown, NJ). Talon cobalt resin was from Clontech (Mountain View, CA). Glutathione Sepharose 4 Fast Flow (GS4FF) resin was from GE Healthcare (Piscataway, NJ). The QuikChange Site-Directed Mutagenesis kit was from Agilent (La Jolla, CA). tC2 Sep-Pak cartridges were from Waters (Milford, MA). T4 DNA ligase, restriction enzymes and the Phusion High-Fidelity PCR kit were from New England Biolabs (Ipswich, MA). DNA purification kits (QIAprep spin minikit, QIAquick gel extraction kit, QIAquick PCR purification kit) were from Qiagen (Valencia, CA). Sub-cloning efficient DH5α competent cells and One Shot BL21(DE3) chemically competent *E. coli* were purchased from Invitrogen (Carlsbad, CA) and used to generate "in-house" high-competency cell lines. Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Codon-optimized cDNAs were generated synthetically and purchased from GENEWIZ (South Plainfield, NJ). Plasmid-sequencing services were ordered from GENEWIZ. DNA purification kits (QIAprep spin minikit, QIAquick gel extraction kit, QIAquick PCR purification kit) were from Qiagen (Valencia, CA). All denaturing gels, PVDF membrane (0.2 μm), Bio-Beads SM2 Adsorbent and goat anti-rabbit monoclonal antibody were from Bio-Rad (Hercules, CA). IRDye 800CW goat anti-rabbit IgG was from LI-COR Biotechnology (Lincoln, NE). The pan-specific antibody recognizing pHis (pHis) was purified from the anti-pHis polyclonal antiserum.

Some synthetic AIPs and their analogs were obtained from the Muir Lab storage freezers.

7.1.2 General Equipments

Solid-phase peptide synthesis (SPPS) was carried on a Liberty Synthesizer (CEM, Matthews, NC). Size exclusion chromatography was carried out on an ÄKTA FPLC system from GE Healthcare using a Superdex200 10/300 column or a Superose6 10/30 HR column. For all runs, proteins were eluted over one column volume of buffer (flow rate: 0.5 mL/min). For SEC-MALS experiments, the ÄKTA FPLC system was connected in tandem to a DAWN HELEOS-II multi-angle light scattering instrument and OptiLab TrEX differential refractometer. Ion-exchange chromatography was performed on the same FPLC instrument system using a Hiprep Q FF 16/10 column. RP-HPLC was performed on Hewlett-Packard 1100 and 1200 series instruments equipped with an analytical C_{18} or C_4 Vydac column (5μm, 4.6 x 150 mm) at a flow rate of 1 mL/min or a semiprep C_{18} or C_4 Vydac column (5μm, 10 x 250 mm) at a flow rate of 4 mL/min. All runs used 0.1 % TFA (trifluoroacetic acid) in water (solvent A) and 90 % acetonitrile in water with 0.1% TFA (solvent B). Electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) mass spectrometric analysis was performed on a Bruker Daltonics MicrOTOF-Q II and a PerSeptive Voyager DE STR MALDI-TOF spectrometer, respectively. Fluorescent anisotropy was measured using a fluorolog®-3 Model FL3-11 fluorimeter (Horiba Jobin Yvon). Coomassie stained gels and dot blots were imaged on a LI-COR Odyssey Infrared Imager. In-gel fluorescence, chemoluminescence and autoradiogram were detecteed using the GE ImageQuant LAS 4000. Scintillation counting was carried out on a Perkin Elmer 2450 microplate counter.

Concentration of DNA, peptide and protein samples was measured on a nanodrop2000 intrument from Thermo scientific or a UV-vis 8453 spectrophotometer from Agilent.

7.2 Molecular cloning

7.2.1 cDNA sequence of Agr proteins

Amino acid sequences of full-length AgrB-I (1-189), AgrD-I (1-46), AgrB-II (1-187) and AgrD-II (1-47), AgrC-I (1-430), AgrC-II (1-427), AgrC-III (1-430) and AgrA (1-238) were obtained from the NCBI protein database (Refseq/Genebank accession: YP_001332977.1, YP_001332978.1, NP_372560.1, NP_372561.1, YP_001332979.1, EFT84677.1, BAB95827.1 and YP_001332980.1). The cDNA sequences were back-translated from the amino acid sequences with optimized codons according to the default *E. coli* codon usage table (186). cDNA was purchased from Genewiz. The cDNA were received in pUC57 vectors.

7.2.2 Expression plasmids for AgrC and AgrA

Coding sequences corresponding to the desired construct of AgrC were amplified with KOD polymerase using primers harboring appropriate restriction sites and cloned between NdeI and XhoI sites of a pET24b or a pET15b vector, respectively. Similarly, the coding sequence of full-length AgrA was cloned into a pACYCDuet vector between BamHI and XhoI restriction sites. Site-directed mutagenesis of the AgrC-I construct was performed using a QuikChange Site-Directed Mutagenesis kit using standard protocols.

In all GCN4-AgrC-I chimeras, the N-terminal portion corresponds to residues 250-274 of the Gcn4p protein in *Saccharomyces cerevisiae* (Refseq accession: NP_010907), to which we refer as $GCN4^{1-25}$. The coding sequence of $GCN4^{1-25}$ was fused immediately upstream to the appropriate AgrC-I coding sequence employing the overlap-extension strategy and inserted between the NdeI and XhoI sites of the pET15b vector.

7.2.3 Expression plasmids for AgrB and AgrD constructs

AgrB-I-His_6 and AgrB-II-His_6

Coding sequences of full-length AgrB-I or AgrB-II were cloned between NdeI and XhoI sites of a pET24b vector (Novagen).

Non-tagged AgrD-I and AgrD-II

A 24-nucleotide (nt) sequence encoding seven histidines and a stop codon was inserted in-frame immediately downstream to the coding sequence of the *Mxe* GyrA intein in the pTXB1 vector (NEB) to generate pTXB1-His_7. Coding sequences of full-length AgrD-I or AgrD-II were cloned between NdeI and SapI sites in pTXB1-His_7.

FlagHis_6-AgrD-I and AgrD-II

FlagHis_6-AgrD-I and FlagHis_6-AgrD-II was fused to GST within an intervening sequence for thrombin (LDKLVPRGSS) or PreScission protease (LDLEVLFQGPGS) recognition, respectively. Coding sequences of these two constructs was inserted between NdeI and XhoI sites in pET24b, with a stop codon immediately downstream to the XhoI site. Coding sequence of GST was from pGEX4T1 (GE Healthcare).

FlagHis$_6$-AgrD-I(1-32)-thiolactone

The coding sequence of FlagHis$_6$-AgrD-I(1-32) was inserted into the pTXB1 vector. The amino acid sequence of the affinity tag is MDYKDDDDKHHHHHHGG.

Plasmids for wild-type or scrambled AgrD-I^C-Ub-Flag$_2$ production in *S. aureus*

Production of AgrD-I^C-Ub-Flag$_2$ was achieved by co-expression of two fusion proteins, Sumo-AgrD-I^C-Ub-Flag$_2$ and His$_6$-Ulp1 from the pCN51 shuttle plasmid (187). The insert sequence begins with the strong, constitutive *pblaZ^C* promoter (50), followed by coding sequences of AgrD-I^C-Ub-Flag$_2$ and His$_6$-Ulp1, both preceded by a consensus Shine-Dalgarno (SD) sequence and a short spacer. This insert sequence was synthesized by Genewiz and cloned between PstI and EcoRI sites of the vector. In a control plasmid, coding sequence of AgrD-I$^{33-46}$ (DEVEVPKELTQLHE) was substituted with that of a scrambled peptide, TLEVKEVQELPEHD. The above substitution and insertion were both achieved using an overlap extension PCR technique.

Site-directed mutagenesis

Expression plasmids bearing point mutations were generated from the parental, wild-type plasmids using a QuikChange Site-Directed Mutagenesis kit following the standard protocol.

7.3 Recombinant protein production and reconstitution

7.3.1 Soluble proteins

With an exception of AgrA, all soluble proteins in this dissertation were prepared as follows. *E. coli* BL21(DE3) cells transformed with the appropriate expression plasmid were grown at 37 °C in one liter of LB medium containing selection antibiotics at appropriate concentrations (100 μg/mL ampicillin, 50 μg/mL kanamycin or 35 μg/mL chloramphenicol). When the OD_{600} reached 0.6, the medium was cooled down to 22°C and overnight expression was induced by addition of 0.5 mM IPTG. Cells were harvested at 6000 G for 20 min and the cell pellets were resuspended in 20 mL of lysis buffer (20 mM Tris, 100 mM NaCl, 1mM TECP and 1mM PMSF, pH 8.0). Cells were lysed by four passages through a French-press homogenizer. After centrifugation at 30000 G for 30min, supernatant from the cell lysate was mixed with 4 mL of Ni-NTA resin. After incubation at 4 °C for 60 minutes, the resin was repacked in a 25-mL Bio-Rad disposable plastic column. The flow-through was discarded and the column was washed with 5 column volumes (CVs) of wash buffer 1 (20 mM Tris, 500 mM NaCl, 1mM TECP and 15 mM imidazole, pH 8.0) and 10 CVs of wash buffer 2 (20 mM Tris, 100 mM NaCl, 1mM TECP and 25 mM imidazole, pH 8.0). Bound protein was eluted with 1.5 CVs of elution buffer (wash buffer 2 with 300 mM imidazole). The elution containing up to 10 mg protein was concentrated to 1mL and further purified on Superdex200 size-exclusion chromatography, from which the desired peak fractions were collected. All GCN4-AgrC-I chimera proteins were treated with thrombin for His_6-tag removal prior to the size-exclusion chromatography step. For purification of AgrA, modification to the above

protocol involved a lowered expression temperature (at 16 °C) and inclusion of glycerol at 20% (v/v) in all buffers.

7.3.2 Integral membrane proteins

For membrane-bound AgrB and AgrC constructs, *E. coli* C43(DE3) cells transformed with the appropriate expression plasmids were grown in LB broth to an OD_{600} of 1.0 and then induced at 22°C with 0.5 mM IPTG overnight expression. Cells collected from one liter of culture were resuspended in 18 mL lysis buffer and disrupted in four passages through a French-press homogenizer. Cell-wall debris was spun down at 15000 G for 10 min and removed. Cell-membrane vesicles were then pelleted by ultracentrifugation at 200,000 G for 1 h. The membrane fraction was extracted using 5mL of buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM BME and the detergent of choice for 3 h at 4 °C (see below). Another ultracentrifugation step at 100,000 G for 20 min was undertaken to clear the membrane extract.

Membrane fractions containing recombinant AgrC proteins were extracted with a buffer containing 2% (w/v) Fos-choline-12 (FC-12). The cleared extract was subjected to Ni-NTA affinity chromatography as described above for soluble proteins, and the elution containing the target protein were further purified over a Superdex200 size-exclusion column. 0.05% (w/v) Fos-choline-14 (FC-14) was included in all buffers used in these chromatographic steps.

Extraction of AgrB out of *E. coli* membrane was performed with a buffer containing 2% (w/v) n-dodecyl-β-D-maltopyranoside (DDM). The cleared extract was subjected to Talon-cobalt affinity chromatography. After the loading and flow-through processes, the column was washed with 20 CVs of wash buffer (20 mM sodium phosphate, 100 mM NaCl, 5mM BME, 0.1% (w/v) FC-12 and 25 mM imidazole, pH 7.5). Bound protein was eluted with 3 CVs of elution buffer (wash buffer with 500 mM imidazole). The elution was concentrated to 1 mL and further purified on Superdex200 size-exclusion chromatography with running buffer (20 mM phosphate, 100 mM NaCl, 0.14% (w/v) FC-12, 1 mM TECP, pH 7.0), from which the desired peak fractions were collected.

7.3.3 AgrD constructs

Figure 7.1 shows a schematic presentation of the expression construct and the proteolytic processing site for all recombinant AgrD constructs. In general, these constructs are either soluble or directed to inclusion bodies when expressed in *E. coli*, and are affinity-purified under native or denatured conditions accordingly. After purification and refolding (if necessary), DDM is included in the buffers for stabilization purposes and these AgrD constructs, either before or after the autocatalytic or proteolytic processing, are handled similarly to integral membrane proteins.

Non-tagged AgrD-I and AgrD-II

Both constructs were expressed as C-terminal GyrA-His_7 fusion proteins in *E. coli* BL21(DE3) cells. Transformed cells harboring the appropriate expression vector were grown at 37^{o}C to an OD_{600} = 0.6 and then induced by addition of 1.0mM IPTG. The

culture was shaken at 37°C for 4 additional hours post-induction and cells were collected through centrifugation. Cells were resuspended in PBS (20 mM Phosphate pH=7.5, 100 mM NaCl). PMSF (100 mM stock in ethanol) was added to the cell suspension immediately prior to lysis at a final concentration of 1 mM. Cells were disrupted by four passages through a French-Press. The lysate was centrifuged at 30000 G for 1hr and the pellet resuspended in a buffer containing 7.5 M GuHCl, 50 mM phosphate buffer, pH = 7.5 and 2 mM TCEP. The suspension was homogenized, gently shaken at 4°C for 1hr and then spun at 30000 G for 1hr. The supernatant was loaded to a column packed with Ni-NTA resin (Qiagen), and the resin was washed with buffer containing 25 mM imidazole and 6 M GuHCl and eluted using a buffer containing 300 mM imidazole, 7.2 M urea and 0.05% (w/v) DDM. The eluted protein was immediately chilled in an ice-water bath and refolded through dialysis against a series of phosphate buffers with decreasing urea

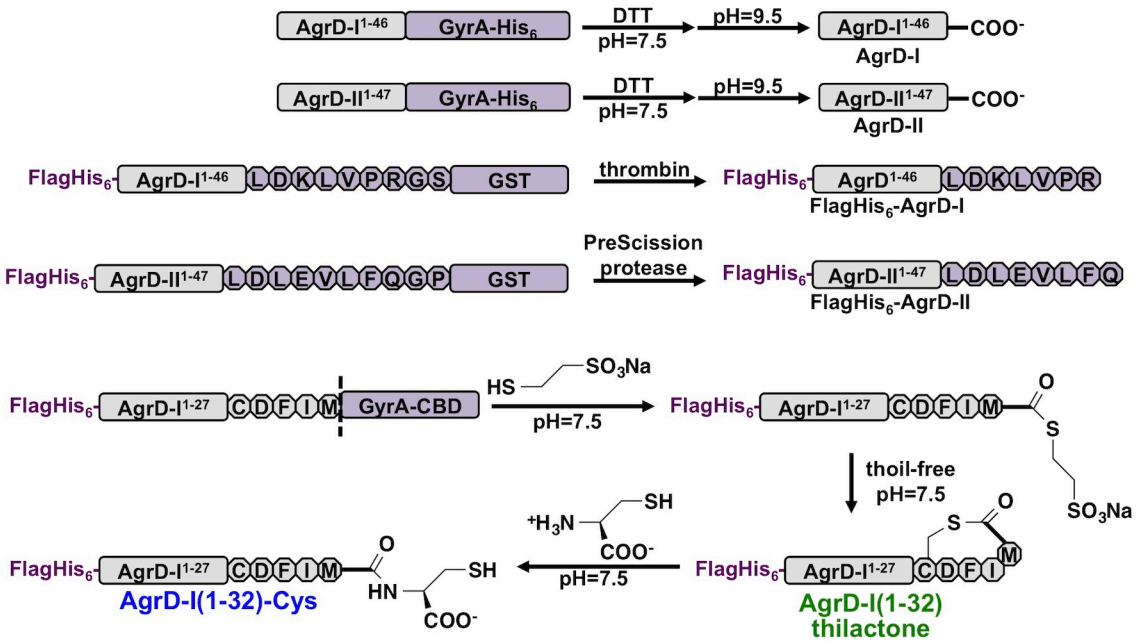


Figure 7.1 Preparation of recombinant AgrD constructs.
See text for detail protocols. GyrA, the *Mxe* GyrA intein; GST, glutathione S-transferase; DTT, dithiothreitol; MESNa, 2-mercaptoethanesulfonate (MESNa).

concentrations at 4°C. Upon dialysis, refolded AgrD-intein-His_7 proteins were cleaved by treatment with 50 mM DTT and 5 mM TCEP in the presence of 0.1% (w/v) DDM under an argon atmosphere at RT for 12 hrs. Free intein released from DTT treatment as well as uncleaved fusion protein was removed in a reverse-Ni-NTA affinity process. To the combined flow-through and wash fractions (containing predominantly AgrD-DTT esters) was then added 0.1% (v/v) FC-12. Residual urea in this solution was removed through two concentration-dilution cycles. In each cycle, the solution was concentrated by ten folds and and then diluted back to the original volume using a buffer containing 50 mM phosphate pH=7.5, 50 mM DTT and 0.07% (v/v) FC-12. The solution after the second dilution was concentrated to <5mL, to which GuHCl was added to a give a final concentration of 6 M. EDTA was added at 10 mM and NaOH was used to basify the solution to pH=9.5. The solution was incubated under argon atmosphere at 37°C for 4hrs for complete saponification of the DTT ester, acidified to pH < 2 using neat TFA, filtered, and purified over a semi-prep C4 RP-HPLC column.

FlagHis_6-AgrD-I and FlagHis_6-AgrD-II

Both constructs were expressed as C-terminal GST fusion proteins in *E. coli* BL21(DE3) host cells. The expression culture was grown at 37°C to an OD_{600} = 0.6 and then was cooled down to 16°C prior to induction by the addition of 0.4 mM IPTG. After overnight incubation, cells were isolated from the culture medium, resuspended in lysis buffer (20 mM sodium phosphate, 100 mM NaCl, pH 7.5), and lysed by four passages through a French-press homogenizer. After centrifugation at 30000 G for 30min, the cleared lysate was supplemented with 0.1% (v/v) DDM and then incubated with GS4FF resin in a

plastic column at 4 °C for 60 minutes. The lysate was then drained, and the resin was washed by 20 CVs of lysis buffer, resuspended in 2 CVs of lysis buffer, and treated with 10 units of thrombin for AgrD-I or 0.2 mg/mL GST-tagged HRV3C protease for AgrD-II at room temperature with gentle shaking for 3hrs(188). The released AgrD product was collected and allowed to flow through a 4-mL bed packed with Ni-NTA resin. The Ni-NTA bed was washed with 10 CVs of lysis buffer with 15 mM imidazole, 10 CVs of 6 M GuHCl buffered by Na-phosphate at pH = 7.5 and eluted with 3 CV of elution buffer (6 M GuHCl, 0.2 M acetic acid and 5 mM TCEP). Elution was purified over a semiprep-scale C18 RP-HPLC column.

FlagHis$_6$-AgrD-I(1-32)-thiolactone and its tri-Ala (F30A/I31A/M32A) mutant

Both constructs were expressed as C-terminal GyrA-CBD fusion proteins in *E. coli* BL21(DE3) host cells. Protein expression and cell lysis procedures were identical to that used for FlagHis$_6$-AgrD preparation. Cleared cell lysate was supplemented with 0.05% (w/v) DDM and subject to Ni-NTA purification. Upon flowing the lysate through the Ni-NTA column, the bed was washed with 5 CVs of wash buffer 1 (20 mM phosphate, 500 mM NaCl, 1 mM TECP 15 mM imidazole and 0.05% (w/v) DDM, pH 7.5) and 10 CVs of wash buffer 2 (20 mM phosphate, 100 mM NaCl, 1 mM TECP 25 mM imidazole and 0.05% (w/v) DDM, pH 7.5). Bound protein was eluted with 4 CVs of elution buffer (wash buffer 2 with 300 mM imidazole). The elution was dialyzed against a buffer containing (20 mM phosphate, 1 mM TECP and 0.01% (w/v) DDM, pH 7.5), and then cleaved after the addition of 100 mM MESNa and 5 mM TCEP. After an overnight incubation, a Ni-NTA affinity step was employed to remove the GyrA-CBD generated

during cleavage. Upon washing with MESNa-free buffers, the bound FlagHis$_6$-AgrD-I(1-32)-MESNa thioester spontaneously cyclized into the thiolactone. All His-tag-containing constructs, primarily the thiolactone and the intact fusion protein, were eluted in 100 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES)-Na, 50mM EDTA, 1mM TECP and 0.1% SDS, pH = 7.0. Concentrated elution was submitted, in 1-mL aliquots, to Superdex200 SEC purification with running buffer (20 mM 2-(N-morpholino)ethanesulfonic acid (MES)-Na, 100 mM NaCl, 0.1%SDS, 5mM L-methionine, pH = 6.1). Fractions free from intact fusion protein were pooled and dialyzed extensively against a buffer containing 0.1% (v/v) TFA and 5mM L-methionine until precipitate emerged. GuHCl was added to a final molarity of 6 M, and the solution was filtered and purified on a semiprep-scale C18 RP-HPLC column.

FlagHis$_6$-AgrD-I(1-32)-Cys and FlagHis$_6$-AgrD-I(1-32)-NAC

FlagHis$_6$-AgrD-I(1-32)-thioactone (1 mg) was dissolved in 1mL 7.5 M GuHCl buffered at pH = 7.5 with sodium phosphate. To the solution was added 50 μL 1 M TCEP together with 100 μL 1 M L-cysteine or N-acetylcysteamine (NAC). The reaction reached maximal conversion after 1-hr incubation at RT as analyzed by analytical RP-HPLC (not shown). The reaction was then purified using semi-prep C18 RP-HPLC.

7.3.4 Nanodisc reconstitution and SEC-MALS analysis

The membrane scaffold protein, MSP1E3D1, was prepared as reported (139). The His$_7$-tag on the recombinant protein was removed through TEV-protease digestion followed by anion exchange chromatography with a Hiprep Q FF 16/10 column (GE Healthcare).

DLPC (1,2-*di*l*l*auryl-sn-*g*lycero-3-*p*hosphocholine), DLPG (1,2-*di*l*l*auryl-sn-glycero-3-*p*hospho-(1'-rac-*g*lycerol)), DMPC, DMPG, and DMPS, POPC and POPG were dried from organic solvents as described on the Avanti Polar Lipids website and prepared in 50mM-stocks solubilized in a buffer containing 20 mM Tris pH = 8.0, 100 mM NaCl, 5 mM TCEP and 150 mM sodium cholate.

Nanodiscs were assembled according to the general considerations described previously (139). Lipid stocks were mixed at appropriate ratios to match the desired composition. For empty nanodisc reconstitution employing MSP1E3D1 as the scaffold, the MSP-to-lipid ratio is 1:200 for dilauryl (DL)-phospholipids, 1:150 for dimyristoyl (DM)-phospholipids and 1:120 for 1-palmitoyl-2-oleoyl (PO)-phospholipids. Each TM domain to be incorporated would displace six DL-phospholipids, five DM-phospholipids or four PO-phospholipids. Full-length AgrC and AgrB are considered to possess six and five TM domains, respectively. For instance, a typical reconstitution of full-length AgrC-I dimers entails two nanodiscs being formed for the incorporation of each copy of AgrC-I dimer. To calculate the scaffold-dimer-lipid molar ratio in the pre-assembly mixture, we first found that two empty nanodiscs would contain four copies of MSP1E3D1 and 600 DM-phospholipids. The incorporation of a 12-TM AgrC dimer is expected to display 60 lipid molecules. Therefore, the usage of MSPE3D1, AgrC-I dimer and DM-phospholipids should follow the molar ratio of 4:1:540.

Upon detergent removal using SM-2 Adsorbent Bio-Beads (Bio-Rad), the post-assembly mixture was subjected to Ni-NTA affinity purification. Elutions containing AgrC-I-

incorporated nanodiscs were resolved on a Superose6 column using the nanodisc buffer (20 mM HEPES, 100 mM NaCl, 1 mM TCEP, pH 7.0) from which peak fractions were pooled as product.

For SEC-MALS analysis, nanodisc samples were resolved on a Superose6 size-exclusion column equilibrated with the nanodisc buffer. The eluate was directed to a DAWN HELEOS-II multi-angle light scattering instrument in tandem with OptiLab TrEX differential refractometer for real-time analysis. Flow-cells of all instruments were set at 4 ^{o}C. Fitting was accomplished using the "protein conjugate" method in the ASTRA 6.0 software package, which calculates both the averaged particle size and the protein content utilizing UV, MALS and dRI signals simultaneously (159). Input for refractive index increment, dn / dc, was 0.185 mL / g for protein and 0.135 mL / g for the modifier (treating lipids as the modifier). Input for extinction coefficient of protein was 0.898 for empty discs (corresponding to MSP1E3D1), 0.806 in an equimolar mixture of MSP1E3D1 and AgrC-I for AgrC-I discs, and 0.840 as in an equimolar mixture of MSP1E3D1 and AgrB-II for AgrB-II discs, all in the unit of mL / (mg · cm).

7.3.5 AgrB proteoliposome assembly

A 3:1 mixture of POPC (1-*p*almitoyl-2-*o*leoyl-sn-*g*lycero-3-*p*hosphocholine) and POPG (1-*p*almitoyl-2-*o*leoyl-sn-glycero-3-*p*hospho-(1'-rac-*g*lycerol)) was dried from organic solvents and then hydrated with PBST buffer containing 20 mM sodium phosphate, 100mM NaCl and 2mM TCEP pH 7.5. The liposome suspension (1.0% w/v) was stored

at -20 ^{o}C, and thawed and ultrasound-treated in a sonicator water bath immediately prior to use.

For proteoliposome reconstitution, liposome suspension was mixed with an equal volume of 3.0% (w/v) 7-cyclohexylheptyl-β-D-maltopyranoside (Cymal-7) in PBST buffer and incubated at RT until the mixture turns clear. A typical reconstitution reaction was started by mixing 400 μL Cyaml-7-solubilized lipid with 200 μL of a 15 μM solution of SEC-purified AgrB-I-His_6. The mixture was incubated with 250 mg of Bio-Beads SM2 at RT with gentle shaking for 1hr. The cloudy suspension was withdrawn from Bio-Beads and the proteolipsomes were spun down at 17000 G at 4^{o}C for 30min. Upon removal of supernatant, the proteoliposomes were washed with 2 x 200 μL PBST buffer, re-suspended in 450μL PBST buffer and gently sonicated before being used for biochemical assays.

7.4 Characterization of the AgrC-AgrA TCS

7.4.1 Auto-kinase and trans-kinase assays

Detection of phosphorylated AgrC-I or AgrC-I variants: general considerations

In this work, we employed a few different readouts to detect and quantify the phosphorylation levels of AgrC-I or AgrC-I variants. Anti-pHis dot blot was the method of choice due to the high throughput of analysis (137). Therefore, most kinetic studies for the forward and reverse auto-phosphorylation of AgrC-I were performed based on dot-blot assays. We also employed anti-pHis western blotting in fixed time-point assays to identify the phosphorylated species based on band positions. For some experiments we

employed a more classic radiolabeling assay using [γ-^{32}P]-ATP as the phosphoryl donor and employing autoradiograghy as the readout. This was found to be especially useful when comparing the activity of different AgrC-I constructs and GCN4-AgrC-I chimeras due to concerns that immunodetection assays might be unduly influenced by subtle differences in the transfer or detection properties of each construct.

Auto-phosphorylation of AgrC-I

All auto-kinase reactions were performed in reaction buffer containing 50 mM Tris-HCl, 15 mM HEPES-Na pH = 7.8, 100 mM NaCl, 5 mM $MgCl_2$ and 1 mM TCEP. AIPs were included as indicated (from DMSO stocks), and DMSO was added to make the overall DMSO concentration up to 1% (v/v). Typically, an auto-kinase reaction contained 20 μM [γ-^{32}P]-ATP (1Ci/mmol) or 1 mM regular ATP or GTP and 1.4 μM dimeric receptor (full-length AgrC-I variants embedded in nanodiscs, or GCN4-AgrC-I chimeras) unless otherwise indicated. The reaction was incubated at 37 °C. In fixed time point assays, reactions were performed for 40min with [γ-^{32}P]-ATP or 3min with regular nucleotides. For time courses, samples were withdrawn from the reaction at indicated time points and quenched with gel loading buffer. All samples were processed, according to the specified method of detection, as described in the following sections.

Phospho-relay from AgrC-I to AgrA

AgrC-I discs were phosphorylated in the absence of AIP with 50 μM [γ-^{32}P]-ATP at 37^{o}C for 90 min, and subsequently exchanged into an ATP-free reaction buffer (50 mM Tris-HCl, 15 mM HEPES-Na pH = 7.8, 100 mM NaCl and 1 mM TCEP) using Bio-Rad

Micro Bio-Spin™ P-6 Gel Columns. This step gave rise to a stock containing 1.2 μM of AgrC-I disc with 0.25 μM of pHis as quantified through scintillation counting. The stock was equally divided into three portions, to which was added either AIP stocks or DMSO at equal volume. These pre-phosphorylated AgrC-I samples were mixed 1:1 with AgrA protein (3.6 μM stock in 50 mM Tris-HCl pH 8.0, 100 mM NaCl , 5 mM TCEP in 20% (v/v) glycerol) to initiate the phospho-relay. The AgrA buffer was used to mix with AgrC-I samples in mock reactions. All reactions were incubated at 37^{o}C from which samples were withdrawn at indicated time points and quantified through autoradiography.

Phospho-relay from GC214 to AgrA and AgrA dephosphorylation

GC214 (10 μM dimer) was treated with 100 μM [γ-^{32}P]-ATP in 20-μL volume at 37^{o}C for 90 min, and subsequently exchanged into an ATP-free reaction buffer. The product was diluted to 180 μL for a stock containing 0.40 μM GC214 dimer (RP-HPLC) with 0.73 μM of pHis (scintillation counting). To initiate the reaction, a mixture containing AgrA, AgrC-I discs and/or AIP peptide were pre-incubated at 37^{o}C for 2min, to which ^{32}P-labeled GC214 was then added. Final concentration of each component (if present) was 0.1 μM for GC214 dimer, 2.0 μM for AgrA, 2.0 μM for AgrC-I discs and 10 μM for AIP-I or AIP-II. All reactions were incubated at 37^{o}C from which samples were removed at indicated time points and quantified through autoradiography.

Reverse Auto-phosphorylation of AgrC-I

AgrC-I discs were phosphorylated in the absence of AIP with 2.0 mM of regular ATP at 37^{o}C for 60min and subsequently exchanged into an ATP-free reaction buffer (50 mM

Tris-HCl, 15 mM HEPES-Na pH = 7.8, 100 mM NaCl and 1 mM TCEP) using a Superose6 size-exclusion column (Figure 3.10c). This step gave rise to a stock containing 4.0 μM of AgrC-I disc and 4.2 μM of pHis as quantified relative to an intermediate standard standardized with scintillation counting. The pHis level of this stock was stable at 4°C over 2 months (data not shown). The stock was diluted 4 fold in reaction buffer in the presence of varying levels of ADP (0-2 mM), $MgCl_2$ (5mM) and either AIP-I (10 μM from a DMSO stock) or DMSO (1% v/v). The reaction was incubated at 37°C from which samples were withdrawn at indicated time points and analyzed using anti-pHis dot blot.

Scintillation counting

Samples from each reaction were spotted on nitrocellulose membranes in triplicate. Samples from a mock reaction (in the absence of AgrC-I) under the same conditions were also spotted and used as the background. Each spot was 5 μL in volume. The membranes were air-dried and then washed 3x5min with TBST buffer (50mM Tris pH 8.5, 150 mM NaCl and 0.1% (v/v) Tween-20) and air-dried again. Each piece of membrane was subsequently transferred to a 4mL counting vial containing 3.5mL of Ultima Gold™ Cocktails (Perkin Elmer) and luminescence from the vial was quantified in a scintillation counter. For quantification purposes, an aliquot of the [γ-^{32}P]-ATP stock used in the reactions was diluted into a series of samples of known concentration. These samples were spotted (5 μL) in duplicate on nitrocellulose membranes and counted after drying without TBST-wash. Correlations between count numbers (in *c*ount *p*er *m*inute, CPM) and concentrations of [γ-^{32}P]-ATP dilutions were plotted to give a working curve which

was used to calculate the absolute pHis concentration in the background corrected experimental samples.

Autoradiography

For detection using autoradiography, samples from auto-kinase or phospho-relay reactions were mixed with 4x SDS sample buffer (1x concentration: 50 mM Tris-HCl pH 8.0, 2 % (w/v) SDS, 10 (v/v) % glycerol, 10 mM dithiothreitol and 0.01 (w/v) % bromophenol blue) and immediately resolved on a 15% Tris-HCl SDS-polyacrylamide gel. The gel was then dried, and incubated with Eastman Kodak BIOMAX MR film at RT for an appropriate amount of time.

For quantification purposes, a highly phosphorylated AgrC-I stock was diluted into a series of internal standards. pHis concentrations of these standards were assigned in arbitrary units based on the dilution process. These standards were resolved on a denaturing gel and exposed, along with gels bearing all experimental samples, to the same piece of film. Importantly, loading volumes of all experimental and standard samples was identical. The normalized, background-subtracted intensity of all bands on the autoradiogram was calculated using the ImageQuant TL program, and used to generate a working curve from which pHis levels (arbitrary units) of experimental samples were determined.

Anti-pHis western blot

As in the autoradiography protocol, samples from auto-kinase reactions were mixed with 4x SDS sample buffer and resolved on a 15% Tris-HCl SDS-polyacrylamide gel. Then, anti-pHis western blot was performed as reported (137). Briefly, the resolved proteins were electroblotted onto a PVDF membrane at 100 V for 1hr. The membrane was blocked with 3% BSA in TBST for 30min and incubated with anti-pHis antibody diluted 1:500 in TBST with 3% BSA for 1 hr at RT. After washing with TBST (3 x 5 min), the membrane was then incubated with goat anti-rabbit IgG-HRP conjugate (diluted 1:5000 in wash buffer with 3% BSA) for 1 hr at RT, followed by washing with TBST (3 x 5 min). The membrane was incubated with ECL chemiluminescence solution for 1 min at RT and chemiluminescence from the membrane was imaged using ImageQuant LAS 4000. The membrane was washed with water and stained with Coomassie blue for the visualization of loading control.

Anti-pHis dot blot

Nitrocellulose membrane (8cm x 8cm) was gridded 10 x 10 and heated on a 50^{o}C heat block prior to use. 2-μL samples from forward or reverse auto-kinase reactions were directly spotted at the center of each small square. Internal standards of known absolute or normalized pHis levels were also spotted (2 μL) in duplicate on the same membrane. The membrane was allowed to dry, cooled to RT and blocked with 3% BSA in TBST for 30min. Anti-pHis antibody diluted 1:500 in TBST with 3% BSA was added for 1 hr at RT. After washing with TBST (3 x 5 min) the membrane was incubated with IRDye 800CW goat anti-rabbit IgG (diluted 1:15000 in wash buffer with 3% BSA) for 1 hr at

RT, followed by washing with TBST (3 x 5 min). The dot blot was imaged on a LI-COR Odyssey Infrared Imager and the intensity of each dot was calculated using the built-in densitometry software. The dot blot was washed with water and stained with Coomassie blue for the visualization and quantification of loading. The pHis concentration of each sample was determined using a working curve generated from the intensity-to-concentration relationship of standard samples.

7.4.2 Kinetic data processing

For kinetic studies, time course experiment of a reaction was performed in multiplicate (n = 3 or 4). The time courses were averaged and fit to certain kinetic models, returning kinetic parameters of the subject reaction. Curve fitting were performed using Graphpad Prism Software.

Phospho-relay from AgrC-I to AgrA (quantified using autoradiography)

$$\text{AgrC-I}\sim\text{P} + \text{AgrA} \rightarrow \text{AgrA}\sim\text{P} + \text{AgrC-I}$$

$$\text{AgrA}\sim\text{P} + \text{H}_2\text{O} \rightarrow \text{AgrA} + \text{Pi} \qquad \text{(Scheme 7.1)}$$

where, AgrC-I~P and AgrA~P represent phosphorylated AgrC-I and AgrA, respectively. To study the decay of AgrC-I~P, only the kinetic equation of the first reaction is considered:

$$\frac{\mathrm{d}[\text{AgrC}\sim\text{I}\sim\text{P}]}{\mathrm{d}t} = -k[\text{AgrC}\sim\text{I}\sim\text{P}][\text{AgrA}] \qquad \text{(Equation 7.1)}$$

where [AgrC-I~P] is the concentration of AgrC-I~P monomers, [AgrA] is the concentration of AgrA, k is the second-order kinetic constant and t is time. In the experimental setup, [AgrC-I~P] << [AgrA] at t = 0. Therefore, progress of the reaction has negligible affect on [AgrA] such that the phospho-relay follows near-first-order

kinetics with respect to AgrC-I~P. Hence, [AgrC-I~P] (in arbitrary unit) was normalized to the [AgrC-I~P] at t = 0, and the time course was fit to the Equation 7.2:

$$\frac{[\text{AgrC–I}\sim\text{P}]}{[\text{AgrC–I}\sim\text{P}]\big|^{t=0}} = (1 - plateau) \bullet \exp\left(-k_{app} \bullet t\right) + plateau \qquad \text{(Equation 7.2)}$$

where k_{app} is the apparent first-order kinetic constant and *plateau* represents the un-reactable fraction of AgrC-I~P at t = 0. The apparent half-life of AgrC-I~P, $t_{1/2}$, was obtained from Equation 7.3:

$$t_{1/2} = \frac{\ln 2}{k_{app}} \qquad \text{(Equation 7.3)}$$

Phospho-relay from GC214 to AgrA and self-catalyzed dephosphorylation of AgrA (quantified using autoradiography)

GC214~P + AgrA → GC214 + AgrA~P

AgrA~P + H_2O → AgrA + Pi (Scheme 7.2)

where, GC214~P and Pi represents phosphorylated GC214 and inorganic phosphate, respectively. This reaction scheme is similar to Scheme 7.1 but quantitatively different because the unmodified GC214 is present in much smaller amount. This drives the phospho-relay reaction to near completion in approximately one minute (Figure 3.6c). Thereafter, no change in [GC214~P] is observed and therefore the change in [AgrA~P] accounts only to the self-catalyzed dephosphorylation (the second reaction). Therefore, [AgrA~P] values (in arbitrary unit) after t = 1min were fit to the first-order decay model for the kinetic constant. The apparent half-life of AgrA-I~P was calculated based on Equation 7.3.

Auto-kinase reaction of AgrC-I (quantified using autoradiography)

AgrC-I + ATP → AgrC-I~P + ADP (Scheme 7.3)

The initial velocity of the reaction follows Equation 7.4:

$$v_i = \left.\frac{\mathrm{d[AgrC{-}I{\sim}P]}}{\mathrm{d}t}\right|^{t=0} = k_1[\mathrm{AgrC{-}I}] \qquad \text{(Equation 7.4)}$$

where [AgrC-I~P] is the concentration of AgrC-I~P monomers, [AgrC-I] is the concentration of free AgrC-I monomers, v_i is the initial velocity and k_1 is the apparent rate constant. The concentration of AgrC-I discs was 1.4 μM (2.8 μM of each monomer) and [γ-^{32}P]-ATP was used at 20 μM. Thus, v_i should be proportional to k_1. In all reactions, [AgrC-I~P] (in arbitrary unit) increased linearly over the first 40 minutes; $R^2 > 0.90$. Therefore, v_i was calculated as the slope of the [AgrC-I~P] time course. For ease of comparision, v_i found under different conditions was normalized to the v_i in the absence of AIP and plotted versus the equivalence of AIP-I (Figure 3.8a).

Auto-kinase reaction of AgrC-I (quantified using anti-pHis dot blot)

Taking the influence of substrate concentration, [ATP] into account, the initial velocity of the reaction is given in Equation 7.5:

$$v_i = \left.\frac{\mathrm{d[AgrC{-}I{\sim}P]}}{\mathrm{d}t}\right|^{t=0} = k_1[\mathrm{AgrC{-}I}] = k_{\max}\frac{[\mathrm{ATP}]}{K_m + [\mathrm{ATP}]}[\mathrm{AgrC{-}I}] \qquad \text{(Equation 7.5)}$$

where k_1 is the first-order rate constant at a given ATP concentration, k_{max} is the first-order rate constant assuming that ATP saturates the binding sites on AgrC-I and K_m is the apparent Michaelis constant with respect to ATP. Under the conditions employed, consumption of ATP was negligible. Therefore, the reaction was first-order with respect

to AgrC-I. Determination of K_m and k_{max} entails k_1 values at a series of ATP concentrations to be calculated from v_i and [AgrC-I]. v_i values were either obtained from linear regression as described in section B or, in the case of saturable time courses, using Equation 7.6:

$$[\mathrm{AgrC{-}I{\sim}P}] = plateau \bullet \left(1 - \exp\left(-k \bullet (t - t_0)\right)\right) \qquad \text{(Equation 7.6)}$$

where *plateau* is the saturation level of pHis, and t_0 is a lag time used to offset the systematic error on the reaction time elapsed before sampling the first time point.

The apparent rate constant k_1 was calculated using Equation 7.7:

$$k_1 = \frac{plateau \bullet k}{[\mathrm{AgrC{-}I}]\Big|^{t=0}} \qquad \text{(Equation 7.7)}$$

The k_1-[ATP] relationship of AgrC-I discs in different ligand states was fit to the Michaelis-Menten equation (Equation 7.8), returning k_{max} and K_m values as fitting parameters:

$$k_1 = k_{\max} \frac{[\mathrm{ATP}]}{K_m + [\mathrm{ATP}]} \qquad \text{(Equation 7.8)}$$

D. Reverse auto-kinase reaction (quantified using anti-pHis dot blot)

AgrC-I~P + ADP → AgrC-I + ATP (Scheme 7.4)

Taking the influence of substrate concentration, [ADP], into account, the initial velocity of the reaction is given in Equation 7.9:

$$v_i = -\left(\frac{\mathrm{d}[\mathrm{AgrC{-}I{\sim}P}]}{\mathrm{d}t}\Big|^{t=0}\right) = k_{-1}[\mathrm{AgrC{-}I{\sim}P}] = k_{\max}\frac{[\mathrm{ADP}]}{K_m + [\mathrm{ADP}]}[\mathrm{AgrC{-}I{\sim}P}] \qquad \text{(Equation 7.9)}$$

where k_{-1} is the first-order rate constant at a given ADP concentration, k_{max} is the first-order rate constant assuming that ADP saturates the binding sites on AgrC-I and K_m is the apparent Michaelis constant with respect to ATP. In contrast to the forward reaction, the reverse reaction is first-order with respect to [AgrC-I~P] at any practical ADP concentration. Absolute quantification of [AgrC-I~P] is hence not required. Therefore, the starting material, pre-phosphorylated AgrC-I discs, was used for internal standardization, and [AgrC-I~P] at all time points was quantified relative to the starting point. As all time courses saturated at the end of the assays, they were fit to Equation 7.10:

$$[\text{AgrC–I}\sim\text{P}]=(Y0-plateau)\bullet\exp(k_{-1}\bullet t)+plateau \qquad \text{(Equation 7.10)}$$

The k_{-1}-[ADP] relationship of AgrC-I discs in different ligand states was fit to the Michaelis-Menten equation (Equation 7.11), returning k_{max} and K_m values as fitting parameters:

$$k_{-1}=k_{\max}\frac{[\text{ADP}]}{K_m+[\text{ADP}]} \qquad \text{(Equation 7.11)}$$

7.4.3 Determination of the dissociation constants between AgrC-I and AIP

Measurement of fluorescent anisotropy

Fluorescence measurements were performed in measurement buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 1mM TCEP, 2 μM BSA and 100 nM of empty nanodiscs). All buffers were freshly degassed before use. Measurement of fluorescent anisotropy was performed on a Fluorolog-3 instrument (HORIBA Jobin Yvon) equipped with automated dual polarizers and using a Semi-Micro Fluorometer Cell (Starna Cells) with a 10-mm

path length. The measurement chamber was held at a constant temperature of 30.0°C by an Advanced Series AC200 thermostat (Thermo Scientific) connected to an Arctic series refrigerated circulating water bath (Thermo Scientific). The excitation wavelength was 490 nm and emission wavelength was recorded at 520 nM, both with a bandwidth of 5 nm. Six to eight measurements were taken per titration point with an integration time of 2 sec for V/V, V/H, H/H and H/V polarizer settings (excitation/emission, V: vertical polarization, H: horizontal polarization). The final anisotropy, *r,* was calculated from $r = (I_{VV} - G \times I_{VH})/(I_{VV} + 2 \times G \times I_{VH})$, with $G = I_{HV}/I_{HH}$.

For titrations of FAM-AIP-I with AgrC-I discs, 750 μL of measurement buffer containing 2.0 nM FAM-AIP-I was transferred to the cuvette and the fluorescent anisotropy was measured as a reference. AgrC-I discs were then added to the required concentration and the mixture allowed to equilibrate for 5 minutes before taking the anisotropy measurement. This process was repeated until measurement of the last titration point (at 1.5 μM of AgrC-I discs). Each titration of FAM-AIP-I was performed simultaneously with a control titration, in which 2.0 nM of FAM-(lin)-AIP-I was included at the starting point.

For competitive titrations with AIP-I or AIP-II, the fluorescent anisotropy of 750 μL of measurement buffer containing 20 nM FAM-AIP-I was first measured as a reference. The starting-point mixture was then prepared by addition of AgrC-I discs to a final concentration of 150 nM. The desired AIP was titrated into the cuvette and the anisotropy measured after 5 minutes equilibration.

Data processing of fluorescent-anisotropy-based binding assays

Curve fitting were performed using Graphpad Prism Software.

A. Titrations of FAM-AIP-I with AgrC-I discs

The fluorescent anisotropy change, Δr, was calculated for each titration point as:

$$\Delta r = r - r_r \qquad \text{(Equation 7.12)}$$

where r_r is the reference anisotropy measured in the absence of AgrC-I discs. To correct the anisotropy increase due to the light-scattering effect of AgrC-I discs, observable only at high-nanomolar and above, the anisotropy change in the control titration (titration of FAM-(lin)-AIP-I), Δr_c, was subtracted from the anisotropy change in the experimental titration, Δr_o at each AgrC-I concentration. This resulting correction, Δr_o - Δr_c, reflects the anisotropy change caused by binding of FAM-AIP-I to the AgrC-I discs. Therefore, Δr_o - Δr_c (referred to as ΔSSA^c in Figures 3.8b and 4.5c) was plotted against [AgrC-I disc]. In titrations with AgrC-I^{S109VS116I} mutant discs, no significant increase of Δr_o - Δr_c was observed so that data points were connected with a smoothed line. In titrations with AgrC-I WT discs, the dataset was fit to the Hill equation treating AgrC-I as the ligand (Equation 7.13):

$$\frac{\Delta r_0 - \Delta r_C}{\Delta r_{\max}} = \frac{[\mathrm{AIPI}_{\mathrm{bound}}]}{[\mathrm{AIPI}_{\mathrm{total}}]} = \frac{[\mathrm{AgrC\text{–}I\ disc}]^h}{[\mathrm{AgrC\text{–}I\ disc}]^h + K_d} \qquad \text{(Equation 7.13)}$$

where [AIPI$_{bound}$] is the concentration of FAM-AIP-I bound to AgrC-I, [AIPI$_{total}$] is the total concentration of FAM-AIP-I, K_d is the dissociation constant, h is the Hill coefficient with respect to AgrC-I, and Δr_{max} is the Δr_o - Δr_c value assuming 100% binding of FAM-AIP-I to AgrC-I discs. Δr_{max}, K_d, and h were obtained as fitting parameters.

To obtain the Hill coefficient with respect to FAM-AIP-I, we considered the Hill equation treating FAM-AIP-I as ligand:

$$\frac{[\text{AgrC}-\text{I}_{\text{bound}}]}{[\text{AgrC}-\text{I}_{\text{free}}]}=\frac{[\text{AIPI}_{\text{free}}]^{h}}{K_d} \qquad \text{(Equation 7.14)}$$

where $[\text{AgrCI}_{\text{free}}]$ and $[\text{AgrC-I}_{\text{bound}}]$ are the concentration of AgrC-I sites unoccupied and bound to FAM-AIP-I, respectively, and $[\text{AIPI}_{\text{free}}]$ is the unbound FAM-AIP-I concentration. Importantly, AIP-I was found to bind AgrC-I in 2:2 stoichiometry. For each titration point, $[\text{AgrC-I}_{\text{free}}]$, $[\text{AgrC-I}_{\text{bound}}]$ and $[\text{AIPI}_{\text{free}}]$ could be calculated as following:

$$[\text{AIPI}_{\text{free}}]=2.0\text{nM}\times\frac{750\mu\text{L}}{V_{total}}\times\left(1-\frac{\Delta r_0-\Delta r_C}{\Delta r_{\max}}\right) \qquad \text{(Equation 7.15)}$$

$$[\text{AgrC}-\text{I}_{\text{bound}}]=[\text{AIPI}_{\text{bound}}]=2.0\text{nM}\times\frac{750\mu\text{L}}{V_{total}}\times\frac{\Delta r_0-\Delta r_C}{\Delta r_{\max}} \qquad \text{(Equation 7.16)}$$

$$[\text{AgrC}-\text{I}_{\text{free}}]=2\times[\text{AgrC}-\text{I disc}]-[\text{AgrC}-\text{I}_{\text{bound}}] \qquad \text{(Equation 7.17)}$$

Where V_{total} is the total volume of the titration system including the volume of titrant added.

$\frac{[\text{AgrC-I}_{\text{bound}}]}{[\text{AgrC-I}_{\text{free}}]}$ was then plotted with $[\text{AIPI}_{\text{free}}]$. To avoid large relative errors in $[\text{AIPI}_{\text{free}}]$ and $[\text{AgrC-I}_{\text{bound}}]$, only data points meeting the criterion of $0.15<\frac{\Delta r_0-\Delta r_C}{\Delta r_{\max}}<0.85$ were included in the dataset. Fitting the dataset to the eq.14 returned values of K_d and h. K_d and h from four individual titrations were pooled and shown as mean ± SD in the main text.

B. Competitive titrations with native AIPs

The fluorescent anisotropy change, Δr (referred to as ΔSSA in Figure 3.8d), was calculated for each titration point and plotted with the concentration of [AIP] in a semi-log plot. Since FAM-AIP-I was found to bind AgrC-I non-cooperatively, we treated AgrC-I as independent monomers. The dataset was fit to Equation 7.18, which was simplified from the literature (189):

$$\Delta r = \frac{(S+B_{\max}) - \sqrt{(S+B_{\max})^2 + 4\times\Delta r_{\max}\times B_{\max}}}{2},$$

where $$S = \left(\frac{K_{d,\mathrm{AIP}} + [\mathrm{AIP}]}{K_{d,\mathrm{AIP}}} \bullet \frac{K_{d,\mathrm{FAM-AIPI}}}{20\mathrm{nM}} + 1\right) \bullet \Delta r_{\max} \qquad \text{(Equation 7.18)}$$

In the above, $K_{\mathbf{d,\ FAM\text{-}AIP}}$, the dissociation constant of FAM-AIP-I, was 122nM. $\Delta r_{\mathbf{max}}$, the anisotropy change assuming 100% binding of FAM-AIP-I to AgrC-I disc, was calculated from the Δr and the concentration of FAM-AIP-I and AgrC-I disc at the starting point. $B_{\mathbf{max}}$ was calculated as:

$$B_{\max} = \Delta r_{\max} \bullet \frac{[\mathrm{AgrC-I_{total}}]}{[\mathrm{FAM-AIPI_{total}}]} \qquad \text{(Equation 7.19)}$$

$K_{d,\ \mathbf{AIP}}$, the dissociation constant of native AIP, was the only variable parameter in the fitting process. For each AIP, K_d values obtained from three individual titrations were pooled and shown as mean ± SD in the main text.

7.4.4 Other methods related to the AgrC-AgrA TCS:

Co-migration assays employing FAM-AIP-I

AgrC-I discs or empty discs (1.5 nmol) were incubated with 10 nmol FAM-AIP-I in 500 μL of nanodisc buffer at 30°C for 30min. The mixture was then resolved on Superose6

size-exclusion chromatography running with nanodisc buffer. The eluate was collected in 0.25-mL size fractions, and fluorescence intensity (λ_{ex} = 490 nm, λ_{em} = 520nm) of the fractions was recorded on a Molecular Devices Spectramax M3 micro plate reader. The fluorescence intensity was plotted with the elution volume of the fraction and connected with a smoothed line. This fluorescence chromatogram overlaid with the UV chromatogram recorded by the UV detector (λ= 280 nm) is shown in Figure 3.2b.

AIP Stability Assays

AIP-I or AIP-II (2 nmol) was incubated in 100 μL solution buffered at pH = 7.0 for 3hrs (AIP-I) or at pH = 8.0 for 1hr (AIP-II) at 37^{o}C in the presence or absence of 0.2 nmol of AgrC-I discs. Control samples without incubation were also prepared. To each sample was added 50 μL of acetonitrile, and the mixture was centrifuged at 17000 G for 5 minutes. Supernatant was analyzed by RP-HPLC on a C18 column using a 0-70B% gradient. The percentage of intact AIPs remaining after incubation with buffer or AgrC-I discs was determined from the peak area and plotted as mean ± SD (n=3) in Figure 3.7c.

Iodoacetamide capture of the hypothetic AgrC-AIP-I covalent complex

FAM-AIP-I (100 pmol) was incubated with AgrC-I discs (40 pmol) in 20 μL buffer (20 mM HEPES pH 7.0, 100 mM NaCl and 1 mM TCEP) at 37^{o}C for 30 minutes. Iodoacetamide was then added to 50 mM (final concentration) and the mixture was incubated at 37°C for additional 30 minutes. The mixture was then denatured with 4 x (v/v) thio-free loading buffer (50 mM Tris-HCl pH 7.5, 2 % (w/v) SDS, 10 % (v/v) glycerol and 0.01 % (w/v) bromophenol blue). Two control samples were prepared under

identical procedures in the absence of either FAM-AIP-I or iodoacetamide. All samples were resolved on Criterion 15% Tris-HCl SDS-polyacrylamide gel, which was first scanned for in-gel fluorescence and then stained with Coomassie blue to confirm the equal loading of AgrC-I protein.

Circular-dichroism (CD) analysis of the TMH-DHp linker peptides

CD spectra were taken on a Chirascan^{TM} CD spectrometer equipped with a Quantum TC125 temperature control unit at 20^{o}C. Path length of the cuvette was 1mm. The HPLC-purified peptide was dissolved in a 100-μM stock solution. This stock was diluted into 15-μM samples containing 10 mM sodium phosphate pH = 7.0, 50 mM NaF, and 0%, 10%, 20%, 30%, 40% or 50% TFE (v/v). For each peptide, CD spectra of six samples were corrected for buffer background and Helical content (*H*) was calculated as:

$$H = \frac{\theta_{222}}{\theta_{\max}} = \frac{\theta_{222}}{-40000 \bullet (1 - 2.5/n) + 100 \bullet T} \qquad \text{(Equation 7.20)}$$

where θ_{222} is the mean residue ellipticity at 222 nm in (deg·cm^2·dmol^{-1}), *n* is the length of the peptide (in residue) and *T* is temperature in ^{o}C (190).

Inter-subunit disulfide formation within AgrC-I cysteine point mutant dimers

Oxidized glutathione (GSSG) was dissolved at 60 mM in a buffer containing 20 mM HEPES and 100 mM NaCl, and pH of the solution was readjusted to 7.0. Nanodiscs incorporated with WT or cysteine-mutant AgrC-I proteins were prepared at 1.0 μM in a buffer containing 20 mM HEPES pH 7.0, 100 mM NaCl and 0.2 mM TCEP. These nanodisc stocks were incubated with either AIP-I, AIP-II or vehicle (DMSO). The final

concentration of each AIP was 10 μM. To initiate oxidation, 10 μL from each aliquot was mixed with 2μL GSSG solution. These mixtures were incubated at 37^{o}C for 10min before mixing with 4 x thiol-free SDS sample buffer. A 'non-oxidized' control reaction was prepared by mixing 10 μL from the AIP-free aliquot with SDS sample buffer and 2 μL GSSG solution simultaneously; this procedure controlled for spurious oxidation post-denaturation. Oxidation reactions were resolved on Criterion 15% Tris-HCl SDS-polyacrylamide gel. The gel was stained with Coomassie blue, scanned with an LI-COR Odyssey Infrared Imager and quantified with the build-in densitometry software. Intensity of each AgrC dimer band was normalized to the MSP band in the same lane. Level of disulfide formation was calculated by subtracting normalized AgrC-I dimer band intensity of the mock-treated discs from the oxidized discs for each AgrC-I mutant.

Lipidation of the AgrC-I$^{208-430}$ construct

To introduce lipid anchors using maleimide chemistry, an AgrC-I intracellular construct bearing an N-terminal cysteine need to be prepared. To ensure specificity, the native cysteine at position 371 was mutated to serine. The AgrC-I$^{208-430}$ (M208C/C371S) with either a wild-type or an R238H mutant HK domain was expressed as an N-terminal His$_6$-Sumo fusion and purified using Ni-NTA affinity chromatography following standard methods (section 7.3.1). The elutate was first dialyzed to imidazole-free and then treated with the Sumo protease, Ulp1, at 16oC overnight. The His$_6$-Sumo released upon proteolysis was removed through a reverse Ni-NTA step, and the AgrC-I$^{208-430}$ was concentrated and further purified on Superdex200 size-exclusion chromatography with running buffer (20 mM HEPES, 100 mM NaCl, and 1mM TECP, pH 7.0)

For lipidation, the purified AgrC-I$^{208-430}$ (M208C/C371S) proteins were adjusted to a concentration between 70-80uM (in dimer). 200uL of the substrate was added to 200uL of lipidation reagents containing 50mM of HEPES pH = 7.6, 2.5 mM of TCEP (pre-neutralized), 10 mM DPPE:MCC and 5% (w/v) FC-12. The reaction was incubated at RT for 40min and quenched with 20 mM of BME (final concentration). Analytical HPLC-MS revealed excellent conversion and specificity. The reaction was purified on Superdex200 size-exclusion chromatography with running buffer containing 0.05% FC-14, and fractions containing the lipidated AgrC-I construct was used for nanodisc reconstitution following standard methods (section 7.3.4).

Limited proteolysis of AgrC-I$^{205-430}$ and GCN4-AgrC-I chimeras

All substrate proteins were diluted to 100 μM (in dimer, 5-6 mg/mL) in cleavage buffer (20 mM Phosphate pH=7.5, 100 mM NaCl). Trypsin (5 mg/mL prepared in 1mM HCl) was serial-diluted in cleavage buffer to five concentrations, i.e. 1.0, 0.5, 0.25, 0.125 and 0.062 in the unit of mg / mL. Each trypsin dilute (5 μL) was mixed with a 15-μL aliquot from each substrate protein, resulting in altogether 20 proteolysis reactions. Each reaction (20uL system) was incubated on ice for 20min and then stopped by adding 5 μL of 1M TFA. 5 μL of the acid-treated reaction was immediately transferred into 20 μL of 100 mM phosphpate pH 7.5 containing cOmplete protease inhibitor cocktail (Roche) at 10X working concentration. To the rest 20 μL of each reaction was added 180 μL 7.5M guanidinium chloride to dissolve the emerging precipitate.All inhibitor-treated samples were denatured by the addition of 15 μL of 4X SDS-PAGE sample buffer. These samples

were subsequently analyzed on Criterion 15% Tris-HCl SDS-polyacrylamide gel with Coomassie staining.

The guanidinium-chloride-solubilzed samples were used for HPLC analysis. RP-HPLC was performed with analytical C18 columns using a 0–70% B gradient over 20 minutes. For two reactions in which AgrC-I$^{205-430}$ and GC206 was digested with the 0.5 mg/mL trypsin dilute, primary peaks eluted between 10 and 20 minute from the HPLC were analyzed with mass spectrometry. Trptic fragments were identified from the primary sequences of AgrC-I$^{205-430}$ and GC206 using the FindPept tool.

7.5 Characterization of the AgrB-AgrD peptidase-substrate pair

7.5.1: Biochemical assays based on AgrB-I prteoliposomes

Proteolysis/hydrolysis assays

For a typical 100 μL reaction, to 45 μL PBST buffer was added 5μL of a 400 μM solution of AgrD-I or FlagHis$_6$-AgrD(1-32)-thiolactone substrate in DMSO. At t = 0, 50 μL AgrB-I proteoliposomes (containing 2 μM AgrB-I monomer as quantified by SDS-PAGE) were added with rigorous mixing and the reaction tube was immediately transferred into a 37^{o}C water bath. 20μL aliquots were withdrawn at indicated time points and mixed with 4 X SDS sample buffer before resolved on 16.5% Tris-Tricine SDS-PAGE. For HPLC-MS analysis, the reaction could be linearly scaled up. At the indicated time points, 100 μL of the reconstitution system was mixed with 400 μL of a solution containing 7.5 M GuHCl, 0.1 M TFA and then subject to Sep-Pak solid phase extraction (SPE, see below). The Sep-Pak eluate was lyophilized, dissolved in the GuHCl-TFA

solution and resolved on a C_4 (for non-tagged constructs) or C_{18} (for $FlagHis_6$-tagged constructs) analytical RP-HPLC column. For proteolysis reactions, a linear gradient with increasing buffer B percentage from 0 to 90% between t = 5 min and 35 min was used. For hydrolysis reactions, a two-segment gradient was employed with buffer B percentage increasing from 0 to 50% in the first 5 min and from 50 to 90% between t = 5 min and 25 min. Peaks with significant absorbance at 214 nm were collected for ESI-MS analysis.

Ligation assays

Immediately prior to ligation assays, AgrB-I proteolipsomes were spun down and re-suspended in PBST buffer containing AgrD-I^{C}-NH_2 peptide at 2X final concentration. This suspension was mixed with PBST buffer and 400 μM DMSO stock of $FlagHis_6$-AgrD-I(1-32)-thiolactone at a volume ratio of 50:45:5 to initiate the reaction. At indicated time points, 100 μL aliquots from the reaction were acidified, Sep-Pak-processed and analyzed using RP-HPLC-MS as described in the previous section.

Sep-Pak SPE

The tC2 Sep-Pak cartridge (100 mg adsorbant, Waters) was activated with 3 x 1 mL of HPLC solvent B and then equilibrated with 2 x 1 mL of HPLC solvent A. Samples acidified with the GuHCl-TFA solution (see section **7-5**) were then load on to the cartridge. The cartridge was subsequently washed with 2 x 1 mL of HPLC solvent A and bound protein was eluted consecutively with 0.5 mL 50% (v/v) acetonitrile in water, 0.5 mL 67% (v/v) acetonitrile in water and 1 mL of 75% (v/v) acetonitrile in water. All

solvents used for elution were buffered by 10 mM triethylammonium-acetate at pH = 8.0. The combined eluate was lyophilized.

7.5.2 Ring-opening of FlagHis$_6$-AgrD-I(1-32)-thiolactone and AIP-I

A 3 M aqueous solution of N-acetylcysteamine (NAC) was prepared and buffered at pH = 7.0 using 100mM HEPES-Na. For each reaction, a 10X NAC solution was prepared from this concentrated stock through serial dilution using 100 mM HEPES-Na pH = 7.0. The following three buffers were used in the reaction:

HBST buffer: 100 mM HEPES-Na, 100 mM NaCl, 2 mM TCEP, pH=7.0;

Liposome buffer: 1 volume of the 1% (w/v) liposomal suspension diluted in 4 volumes of HBST buffer;

GuHCl buffer: 100 mM HEPES-Na, 7.5 M GuHCl, 2 mM TCEP, pH=7.0;

Each reaction was initiated by combining 50 μL 1M TCEP-Na (buffered at pH = 7.0), 100μL 10X NAC solution and 50 μL of a 400 μM DMSO stock of FlagHis$_6$-AgrD-I(1-32)-thiolactone or AIP-I with 800 μL of one of the above three buffers. For reactions performed in HBST or GuHCl buffers, aliquots (100 μL for AIP-I or 20 μL for FlagHis$_6$-AgrD-I(1-32)-thiolactone) were removed and analyzed by C_{18} analytical RP-HPLC at indicated time points. For reactions containing liposomes, aliquots were added to 0.5 mL 0.1 M aqueous TFA and freeze-dried. The lyophilized sample was triturated with 150 μL GuHCl-TFA solution. The extract was then cleared through centrifugation and 100 μL of the supernatant was analyzed on RP-HPLC. Ring-opening reactions of the tri-Ala mutants of FlagHis$_6$-AgrD-I(1-32)-thiolactone and AIP-I were initiated, processed and analyzed using the same procedures.

Analytical HPLC columns and gradients:

Samples from reactions involving AIP-I: solvent B percentage 20% to 50% over 15min;

Samples from reactions involving tri-Ala mutant of AIP-I: solvent B percentage 0% to 30% with 15min;

Samples from reactions involving wild-type or tri-Ala mutant of FlagHis$_6$-AgrD-I(1-32)-thiolactone: two-segment gradient such that solvent B percentage increases from 0% to 50% in 5 min and then to 80% in 15min.

7.5.3 Quantification methods

Quantification of all species in the proteolysis reaction from HPLC analysis

The proteolysis reaction begins with 20 μM pure FlagHis$_6$-AgrD-I. This AgrD-I construct has two tyrosines, both located in the N-fragment that results from AgrB-I cleavage. Therefore, the full-length AgrD-I, the linear and the thiolactone N-fragment should have identical extinction coefficients at 280 nm. Peak areas of these species in the 280-nm HPLC chromatogram are hence proportional to their respective molarities. The molar fraction of each species (out of the sum of three) was calculated accordingly after factoring in their respective SPE recovery (see below). The molarity of AgrD-I^C is given by the sum of the molarity of both linear and thiolactone N-fragments as a consequence of mass balance.

To determine the relative SPE recovery, FlagHis$_6$-AgrD-I, FlagHis$_6$-AgrD-I(1-32)-Cys [as a surrogate of FlagHis$_6$-AgrD-I(1-32)] and the FlagHis$_6$-AgrD-I(1-32)-thiolactone N-fragment were mixed at a moderate ratio, and part of the mixture was diluted to an

overall molarity of 20μM in a buffer containing phospholipids. This sample was incubated at 37^{o}C for 30min and then subject to the Sep-Pak SPE following the standard method provided in section 7.5.2. The SPE eluate and the mixture prior to dilution were resolved on HPLC and the molar fractions of each species were plotted (Figure 7.2a). Relative Sep-Pak recovery of each species is calculated as the ratio between its post- and pre-SPE molar fraction relative to the sum of three.

Quantification of the ligation reaction based on HPLC-MS analysis

The ligation reaction starts with 20 μM FlagHis$_6$-AgrD-I(1-32)-thiolactone and a variable concentration of AgrD-I^{C}-NH_2. The starting thiolactone overlaps, on the RP-HPLC chromatogram, with the ligation product, but is separable from the linear FlagHis$_6$-AgrD-I(1-32). Again, due to the lack of a chromogenic residue in the AgrD-I^{C} segment, the extinction coefficient at 280 nm is expected to be the same for the starting thiolactone, the linear AgrD(1-32) and the ligation product AgrD. Accordingly, the linear AgrD(1-32) was quantified based on RP-HPLC peak areas as described in the previous section. The molar ratio between the starting thiolactone and the ligation product was determined using MS. Briefly, MS spectra were recorded for the eluate containing AgrD and both AgrD(1-32) fragments and the profile between m/z = 600 and 2000 was de-convoluted using the "maximum entropy" method (191). Target MS range and instrument resolution were set at 3000-8000 Da and 10000, respectively. The peaks corresponding to the starting thiolactone (5694.2Da) and ligation product (7359.2Da) in the de-convoluted profile were integrated. The molar ratio between the two species was calculated from

their MS intensity after factoring in a coeffecient that accounts for their difference in ionization efficiency (see below).

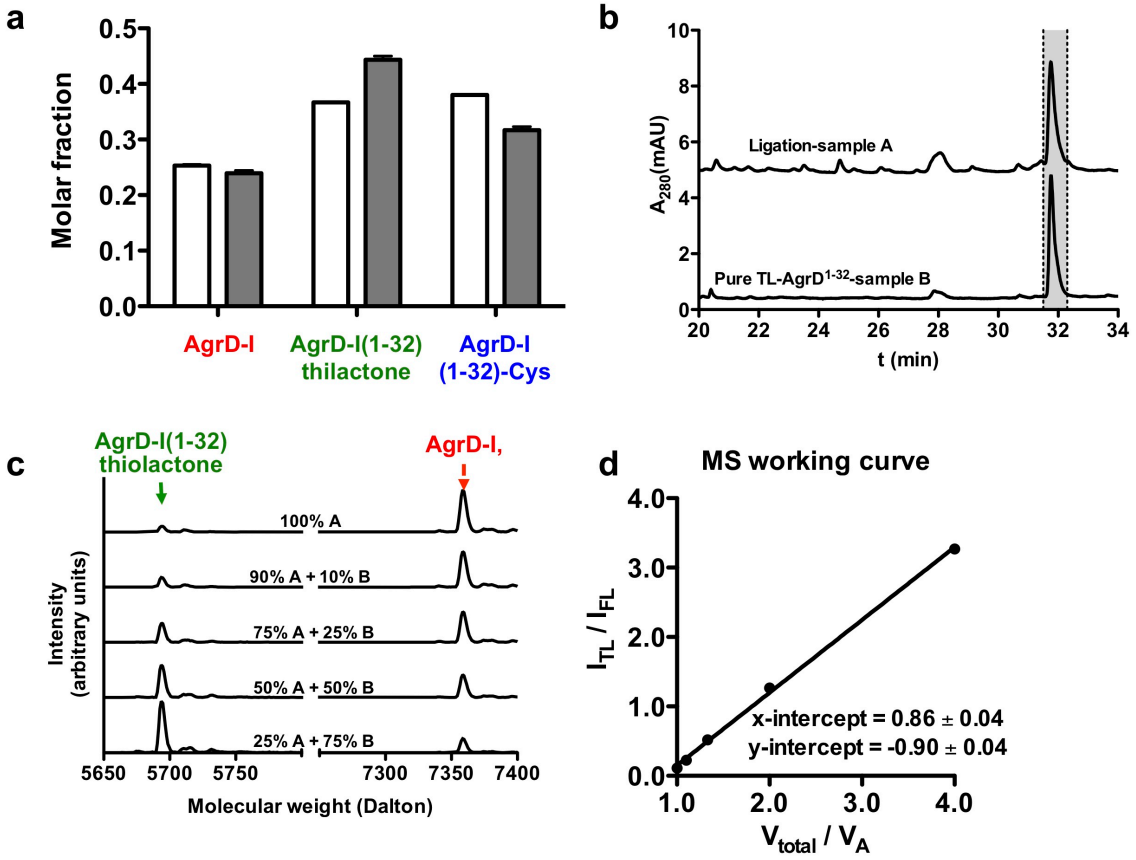


Figure 7.2 Correction of the systematic errors introduced by the Sep-Pak procedure or the ionization bias of the MS

(a) The relative Sep-Pak recovery among FlagHis$_6$-AgrD-I, FlagHis$_6$-AgrD-I(1-32)-thiolactone, and FlagHis$_6$-AgrD-I$^{1-32}$-Cys: bar graph shows the molar fraction of each species in a mixture of three before (open bars) or after (closed bars) solid-phase extraction (SPE) using Sep-Pak. Pre- and post-SPE samples were quantified using HPLC. Error bars = range (n = 2). **(b** to **d)** The relative ionization efficiency of FlagHis$_6$-AgrD-I with respect to FlagHis$_6$-AgrD-I(1-32)-thiolactone. **(b)** RP-HPLC preparation of standard samples A and B: 20μM of FlagHis$_6$-AgrD-I(1-32)-thiolactone was treated with 300μM AgrD^C in the presence of AgrB-I proteoliposomes or mock-treated and then subject to RP-HPLC. Shaded segments of elution were collected from the ligation (as sample A) and the mock treatment (as sample B). Both samples were diluted to the same molarity; **(c)** Deconvoluted mass spectra for mixtures of samples A and B at indicated ratios. **(d)** The working curve: plot shows the MS peak intensity ratio between FlagHis$_6$-AgrD-I(1-32) and FlagHis$_6$-AgrD-I as a function of the reciprocal of the weight of sample A in the mixture. The inverse reciprocal of the y-intercept, *k*, converts MS peak intensity ratio to the molar ratio by simple multiplication. Error bars (range between two parallels) are not visible due to the size of data symbols.

Relative MS ionization efficiency of FlagHis$_6$-AgrD-I with respect to FlagHis$_6$-AgrD-I(1-32)-thiolactone

This procedure was used to empirically determine a coeffecient, *k*, that converts the MS peak intensity ratio between the starting thiolactone (**TL**) and the full-length ligation product (**FL**) ($I_{\mathbf{TL}}/I_{\mathbf{FL}}$) into the molarity ratio ($c_{\mathbf{TL}}/c_{\mathbf{FL}}$). Note that these two species cannot be resolved on RP-HPLC.

$$k\frac{I_{\mathrm{TL}}}{I_{\mathrm{FL}}}=\frac{c_{\mathrm{TL}}}{c_{\mathrm{FL}}} \qquad \text{(Equation 7.21)}$$

For simplicity, we prepared sample B containing pure **TL** and sample A as a mixture containing predominantly **FL**. To this end, 20μM of FlagHis$_6$-AgrD-I(1-32)-thiolactone was treated with 300μM AgrD$^{\mathrm{C}}$ in the presence of AgrB-I proteoliposomes or mock-treated and then subject to RP-HPLC. The whole HPLC peak was collected from the ligation (as sample A) and the mock treatment (as sample B) (Figure 7.2b). Based on their peak areas on the 280-nm RP-HPLC chromatogram, both samples were diluted to the same molarity. Assuming the molar fraction of **FL** in sample A to be x (x<1), the **TL**-to-**FL** molarity ratio in a A-B mixture is given by equation 22:

$$\frac{c_{\mathrm{TL}}}{c_{\mathrm{FL}}}=\frac{V_{\mathrm{B}}+(1-x)V_{\mathrm{A}}}{V_{\mathrm{A}}x} \qquad \text{(Equation 7.22)}$$

where $V_{\mathbf{A}}$ and $V_{\mathbf{B}}$ are the volume of A and B in the mixture. Equation 7.21 and Equation 7.22 could be transformed into:

$$\frac{I_{\mathrm{TL}}}{I_{\mathrm{FL}}}=\frac{1}{kx}\frac{V_{\mathrm{B}}+V_{\mathrm{A}}}{V_{\mathrm{A}}}-\frac{1}{k} \qquad \text{(Equation 7.23)}$$

Therefore, after plotting the MS intensity ratio between **TL** and **FL** versus the overall volume of the A-B mixture divided by the participating sample A volume, linear

regression of the plot returns *k* as the inverse reciprocal of the y-intercept (Figure 7.2c and d).

Quantification of the thioester/thiolactone in the ring-opening assays

Molar ratios of AIP-I-NAC thioester versus AIP-I or $FlagHis_6$-AgrD(1-32)-NAC thioester versus $FlagHis_6$-AgrD(1-32)-thiolactone were quantified based on integration of 280-nm RP-HPLC chromatograms.

7.5.4 Construction of *S. aureus* strains and western blotting analysis thereof

Shuttle plasmids for the *pblaZ*^{C}-driven co-expression of Sumo-AgrD^C-Ub-$Flag_2$ and His_6-Ulp1 were first electroporated into strain RN4220 (a restriction-deficient host) and then transduced to strain RN6390 (agr-I wild type) or RN7206 (RN6390 carrying an *agr* locus deletion) *via* phage 80 using standard protocols (192). Transductants were grown in CYGP medium containing 10 μg/mL erythromycin at 37^{o}C. Upon reaching the stationary phase, cells from a 1 mL culture were collected, washed twice with ice-cold water and finely resuspended in fresh CYGP medium on ice. 20 mL CYGP medium containing 10 ug/mL erythromycin was inoculated with 40 μL cell suspension (c.a. $2x10^8$ cfu) and then shaken at 250 rpm at 30^{o}C. Cell density was measured every 30 min. When OD_{600} exceeded 0.60, cell culture was diluted using fresh CYGP medium to give an OD_{600} of between 0.20 and 0.60 prior to measurement, and the readout was then corrected for the dilution.

For the detection of $AgrD^{C}$-Ub-$Flag_2$ and AgrD-Ub-$Flag_2$ using anti-Flag western blot, 1 mL cell culture was sampled when the culture OD_{600} reached 0.50, 1.00, 1.50 and 2.00. Cells were pelleted and resuspended in lysis buffer (10mM Tris-HCl 8.0, 1mM TCEP, 100μM $ZnCl_2$, supplemented with 50 μg/mL lysostaphin and 1X protease inhibitor cocktail). The volume of suspension was proportional to the cell density of the original sample (50 μL per 1.0 OD_{600}). The suspension was incubated at 37°C for 30 min, mixed with appropriate amount of 3X SDS-PAGE loading dye and boiled at 95°C for 10min. Samples were then resolved on 15% Tris-HCl SDS-PAGE. To avoid interference of western blotting by endogenous protein A, the PAGE gel was cut at the level of the 37.5kDa standard band and the upper half was stained with Coomassie blue for loading control. Proteins in the lower half of the gel were transferred to a PVDF membrane and blotted with Flag-specific mouse monoclonal antibody (Sigma-Aldrich) using standard protocols. After blotting the membrane with goat anti-mouse IgG HRP conjugate (Bio-Rad), bands were visualized through chemiluminescence.

Extracellular AIP-I activity in plasmid-free or transformed RN6390 culture was determined using a reporter gene assay (39). Briefly, RN9222 (an *agr*-I reporter strain harboring a β-lactamase gene driven by a P3 promoter) was grown in CYGP medium at 37°C to OD_{600} = 0.60. A 50-μL aliquot of the culture was mixed with 50μL fresh CYGP (for background) or cell-free medium isolated from RN6390 samples. The mixture was cultured at 37°C for 1hr and cell density (OD_{600}) was measured. Then, 50 μL of the mixture was transferred to a 96-well plate and mixed with 50 μL 200 μM nitrocephin in assay buffer (100 mM phosphate-Na pH = 5.9, 27% (v/v) propylene glycol, 5 mM

sodium azide). Progress of the substrate hydrolysis was monitored at 37°C in a plate reader. The slope of the linear increase of absorbance at 490 nm during the initial period of the reaction was designated as the β-lactamase activity in (0.001 absorbance unit/min). The final result was background-subtracted and normalized by the pre-reaction cell density of RN9222.

7.5.5 Other experiments related to AgrB and AgrD

Glutaraldehyde crosslinking of AgrB-II

Crosslinking was performed on SEC-purified AgrB-II protein solubilized by FC-12. The protein was diluted in PBST buffer (20 mM sodium phosphate, 100 mM NaCl and 2 mM TCEP pH 7.5) supplemented with 0.14% FC-12 to 2.5 μM. 20 mM glutaraldehyde was prepared in 100 mM phosphate buffer (pH = 7.5) immediately prior to use. The crosslinking was performed at room temperature. 90μL of protein and 10μL crosslinker were mixed at t = 0, and 20μL aliquots were withdrawn at indicated time points and quenched by mixing with 2μL 1M Tris buffered at pH=8.7. The t = 0 control was prepared by adding first 2μL 1M Tris pH=8.7 and then 2μL crosslinker to 18μL protein stock. All quenched reactions were mixed with 4X SDS sample buffer and resolved on 4-20% TGX SDS-PAGE.

Proteolysis assays using AgrB-II nanodiscs

SEC-purified AgrB-II nanodiscs were diluted in PBST such that each nanodisc sample contains 2 μM AgrB-II (counted as monomers). FlagHis$_6$-AgrD-II was dissolved in DMSO at 0.5 mM, from which 1 μL was mixed with 25μL AgrB-II nanodisc sample.

Reactions were incubated at 37^{o}C for 2hr before mixing with 4 X SDS sample buffer and resolved on Criterion 16.5% Tris-Tricine SDS-polyacrylamide gel. The gel was stained with Coomassie blue.

Appendix

Appendix Table 1. All RHK hits from the database search for HPK10 homologs in Chapter 2

Entry	Accession Number	Length	TMH	"G1-box" Residue	Phylum	Species
1	WP_006147676.1	203	0	Asn	Firmicutes	Streptococcus infantis
2	WP_010734516.1	224	0	Asn	Firmicutes	Enterococcus mundtii
3	YP_004561998.1	233	0	Asn	Firmicutes	Lactobacillus kefiranofaciens ZW3
4	YP_003602088.1	230	0	Asn	Firmicutes	Lactobacillus crispatus ST1
5	WP_016623240.1	228	0	Asn	Firmicutes	Enterococcus faecalis
6	YP_006621934.1	387	0	Asn	Firmicutes	Desulfosporosinus meridiei DSM 13257
7	WP_002471760.1	220	0	Asn	Firmicutes	Staphylococcus pettenkoferi
8	WP_009384255.1	240	0	Asn	Firmicutes	Staphylococcus massiliensis
9	WP_010241734.1	256	1	Asn	Firmicutes	Clostridium arbusti
10	WP_003037768.1	274	1	Asn	Firmicutes	Streptococcus anginosus
11	WP_006701462.1	232	1	Asn	Firmicutes	Facklamia ignava
12	WP_010170763.1	404	1	Asn	Firmicutes	Bacillus coahuilensis
13	WP_010241389.1	291	2	Asn	Firmicutes	Clostridium arbusti
14	WP_016230456.1	303	2	Asn	Firmicutes	Lachnospiraceae bacterium 10-1
15	WP_010275906.1	283	2	Asn	Firmicutes	Leuconostoc pseudomesenteroides
16	WP_005723496.1	343	2	Asn	Firmicutes	Lactobacillus crispatus
17	YP_006920935.1	282	2	Asn	Firmicutes	Thermacetogenium phaeum DSM 12270
18	YP_006462487.1	435	2	Asn	Firmicutes	Solibacillus silvestris StLB046
19	WP_006837370.1	437	2	Asn	Firmicutes	Bacillus sp. SG-1
20	WP_010861419.1	435	2	Asn	Firmicutes	Lysinibacillus sphaericus
21	YP_001697754.1	435	2	Asn	Firmicutes	Lysinibacillus sphaericus C3-41
22	YP_002771205.1	450	2	Asn	Firmicutes	Brevibacillus brevis NBRC 100599
23	WP_005833804.1	441	2	Asn	Firmicutes	Brevibacillus
24	WP_003094496.1	317	3	Asn	Firmicutes	Streptococcus vestibularis
25	WP_004271011.1	316	3	Asn	Firmicutes	Lactobacillus curvatus
26	WP_003069924.1	308	3	Asn	Firmicutes	Streptococcus intermedius
27	WP_001872392.1	334	3	Asn	Firmicutes	Streptococcus agalactiae
28	WP_002300816.1	352	3	Asn	Firmicutes	Enterococcus faecium
29	WP_007126042.1	334	3	Asn	Firmicutes	Lactobacillus ultunensis
30	WP_009999978.1	355	3	Asn	Firmicutes	Leuconostoc lactis
31	YP_001699604.1	353	3	Asn	Firmicutes	Lysinibacillus sphaericus C3-41
32	YP_003842650.1	379	4	Asn	Firmicutes	Clostridium cellulovorans 743B
33	WP_009854992.1	454	4	Asn	Firmicutes	Streptococcus gallolyticus
34	WP_004271266.1[c]	385	4	Asn	Firmicutes	Lactobacillus curvatus
35	WP_004634648.1	430	4	Asn	Firmicutes	Dolosigranulum pigrum
36	WP_016264937.1	398	4	Asn	Firmicutes	Lactobacillus sakei
37	WP_010752094.1	365	4	Asn	Firmicutes	Enterococcus villorum
38	WP_009557583.1	377	4	Asn	Firmicutes	Lactobacillus sp. 66c
39	WP_005725782.1	395	4	Asn	Firmicutes	Lactobacillus crispatus
40	WP_002482782.1	365	4	Asn	Firmicutes	Staphylococcus saprophyticus
41	WP_009443765.1	404	5	Asn	Firmicutes	Streptococcus sp. oral taxon 056
42	WP_010757654.1	397	5	Asn	Firmicutes	Enterococcus pallens
43	YP_006991522.2	430	5	Asn	Firmicutes	Carnobacterium maltaromaticum LMA28
44	WP_009535076.1	435	5	Asn	Firmicutes	Oribacterium
45	WP_010781286.1	433	5	Asn	Firmicutes	Enterococcus gilvus
46	WP_010742519.1	441	5	Asn	Firmicutes	Enterococcus malodoratus
47	WP_016179641.1	441	5	Asn	Firmicutes	Enterococcus avium
48	WP_016176172.1	436	5	Asn	Firmicutes	Enterococcus saccharolyticus
49	YP_003926231.1	420	5	Asn	Firmicutes	Lactobacillus plantarum subsp. plantarum ST-III
50	WP_018373329.1	443	5	Asn	Firmicutes	Streptococcus merionis
51	WP_016624566.1	426	5	Asn	Firmicutes	Enterococcus faecalis
52	WP_018380719.1	447	5	Asn	Firmicutes	Streptococcus thoraltensis
53	WP_010736347.1	438	5	Asn	Firmicutes	Enterococcus mundtii
54	WP_019782329.1	447	5	Asn	Firmicutes	Streptococcus sobrinus
55	WP_004225929.1	442	5	Asn	Firmicutes	Streptococcus criceti
56	YP_005004303.1	423	5	Asn	Firmicutes	Pediococcus claussenii ATCC BAA-344
57	YP_003431422.1	435	5	Asn	Firmicutes	Streptococcus gallolyticus UCN34
58	WP_003638473.1	422	5	Asn	Firmicutes	Lactobacillus pentosus

59	WP_004242076.1	448	5	Asn	Firmicutes	Streptococcus mitis
60	WP_006917176.1	427	5	Asn	Firmicutes	Lactobacillus coleohominis
61	WP_004049770.1	430	5	Asn	Firmicutes	Lactobacillus murinus
62	WP_016623999.1	381	5	Asn	Firmicutes	Enterococcus faecalis
63	WP_003665154.1	425	5	Asn	Firmicutes	Lactobacillus reuteri
64	WP_007126043.1	432	5	Asn	Firmicutes	Lactobacillus ultunensis
65	WP_009431292.1	366	5	Asn	Firmicutes	Peptoniphilus sp. oral taxon 375
66	WP_016307854.1	439	5	Asn	Firmicutes	Lachnospiraceae bacterium A2
67	YP_001307802.1	420	5	Asn	Firmicutes	Clostridium beijerinckii NCIMB 8052
68	YP_004090305.1	433	5	Asn	Firmicutes	Ruminococcus albus 7
69	WP_002847577.1	432	5	Asn	Firmicutes	Ruminococcus albus
70	WP_005945720.1	424	5	Asn	Firmicutes	Faecalibacterium prausnitzii
71	WP_016220195.1	440	5	Asn	Firmicutes	Dorea sp. 5-2
72	WP_001639479.1	362	5	Asn	Firmicutes	Staphylococcus aureus
73	YP_005119571.1	406	5	Asn	Firmicutes	Bacillus cereus F837/76
74	WP_002579318.1	439	6	Asn	Firmicutes	Clostridium butyricum
75	WP_003410367.1	435	6	Asn	Firmicutes	Clostridium butyricum
76	YP_001311644.1	439	6	Asn	Firmicutes	Clostridium beijerinckii NCIMB 8052
77	WP_003446201.1	445	6	Asn	Firmicutes	Clostridium pasteurianum
78	WP_010766320.1	430	6	Asn	Firmicutes	Enterococcus moraviensis
79	WP_004071405.1	433	6	Asn	Firmicutes	Clostridium sp. ASF502
80	WP_010234544.1	451	6	Asn	Firmicutes	Clostridium arbusti
81	WP_010772171.1	431	6	Asn	Firmicutes	Enterococcus caccae
82	WP_016622707.1	434	6	Asn	Firmicutes	Enterococcus faecalis
83	WP_016216882.1	440	6	Asn	Firmicutes	Dorea sp. 5-2
84	WP_016226170.1	431	6	Asn	Firmicutes	Lachnospiraceae bacterium 3-2
85	WP_016226738.1	428	6	Asn	Firmicutes	Lachnospiraceae bacterium 10-1
86	WP_016292271.1	436	6	Asn	Firmicutes	Lachnospiraceae bacterium 28-4
87	YP_003781661.1[b]	449	6	Asn	Firmicutes	Clostridium ljungdahlii DSM 13528
88	YP_001786573.1	450	6	Asn	Firmicutes	Clostridium botulinum A3 str. Loch Maree
89	WP_010750902.1	424	6	Asn	Firmicutes	Enterococcus villorum
90	WP_010768675.1	430	6	Asn	Firmicutes	Enterococcus phoeniculicola
91	WP_010719091.1	428	6	Asn	Firmicutes	Enterococcus faecium
92	WP_016215903.1	434	6	Asn	Firmicutes	Eubacterium sp. 14-2
93	WP_010779009.1	443	6	Asn	Firmicutes	Enterococcus gilvus
94	WP_009732097.1	434	6	Asn	Firmicutes	Streptococcus sp. F0442
95	WP_006596699.1	445	6	Asn	Firmicutes	Streptococcus australis
96	WP_002885486.1	439	6	Asn	Firmicutes	Streptococcus salivarius
97	WP_002889753.1	445	6	Asn	Firmicutes	Streptococcus salivarius
98	YP_007988158.1	420	6	Asn	Firmicutes	Lactobacillus plantarum subsp. plantarum P-8
99	WP_016287036.1	434	6	Asn	Firmicutes	Lachnospiraceae bacterium 3-1
100	YP_004287481.1	435	6	Asn	Firmicutes	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069
101	WP_016304832.1	434	6	Asn	Firmicutes	Lachnospiraceae bacterium A2
102	WP_002604771.1	437	6	Asn	Firmicutes	Clostridium hathewayi
103	WP_002280560.1	441	6	Asn	Firmicutes	Streptococcus mutans
104	WP_010739752.1	433	6	Asn	Firmicutes	Enterococcus malodoratus
105	YP_003064445.1	419	6	Asn	Firmicutes	Lactobacillus plantarum JDM1
106	WP_018164077.1	450	6	Asn	Firmicutes	Streptococcus henryi
107	WP_010744242.1	440	6	Asn	Firmicutes	Enterococcus raffinosus
108	WP_002943511.1	445	6	Asn	Firmicutes	Streptococcus suis
109	WP_018378535.1	444	6	Asn	Firmicutes	Streptococcus ovis
110	WP_004058876.1	441	6	Asn	Firmicutes	Eubacterium plexicaudatum
111	WP_003702872.1	426	6	Asn	Firmicutes	Lactobacillus salivarius
112	WP_002313965.1	421	6	Asn	Firmicutes	Enterococcus faecium
113	YP_006069688.1	438	6	Asn	Firmicutes	Streptococcus salivarius JIM8777
114	WP_004255283.1[c]	441	6	Asn	Firmicutes	Streptococcus mitis
115	WP_019787615.1	448	6	Asn	Firmicutes	Streptococcus sobrinus
116	YP_878135.1	437	6	Asn	Firmicutes	Clostridium novyi NT
117	WP_005880245.1	451	6	Asn	Firmicutes	Enterococcus durans
118	WP_009553709.1[b]	422	6	Asn	Firmicutes	Lactobacillus saerimneri
119	WP_003376164.1	439	6	Asn	Firmicutes	Clostridium botulinum
120	WP_018375371.1	436	6	Asn	Firmicutes	Streptococcus orisratti
121	WP_018376391.1	435	6	Asn	Firmicutes	Streptococcus orisratti
122	WP_020992797.1	441	6	Asn	Firmicutes	Streptococcus sp. 2_1_36FAA
123	WP_018373287.1	436	6	Asn	Firmicutes	Streptococcus merionis

124	WP_018164876.1	440	6	Asn	Firmicutes	Streptococcus henryi
125	WP_010817845.1	437	6	Asn	Firmicutes	Enterococcus faecalis
126	YP_002561848.1	439	6	Asn	Firmicutes	Streptococcus uberis 0140J
127	WP_008535279.1	446	6	Asn	Firmicutes	Streptococcus sp. C150
128	WP_009754525.1	447	6	Asn	Firmicutes	Streptococcus sp. oral taxon 056
129	WP_004240956.1	444	6	Asn	Firmicutes	Streptococcus mitis
130	WP_004195387.1	454	6	Asn	Firmicutes	Streptococcus suis
131	YP_006067616.1	444	6	Asn	Firmicutes	Streptococcus salivarius 57.I
132	WP_003139433.1	458	6	Asn	Firmicutes	Lactococcus raffinolactis
133	YP_005173501.1	436	6	Asn	Firmicutes	Leuconostoc mesenteroides subsp. mesenteroides J18
134	WP_016224073.1	435	6	Asn	Firmicutes	Lachnospiraceae bacterium 3-2
135	WP_004049925.1	441	6	Asn	Firmicutes	Lactobacillus murinus
136	WP_016184462.1	457	6	Asn	Firmicutes	Enterococcus columbae
137	WP_016226189.1	438	6	Asn	Firmicutes	Lachnospiraceae bacterium 3-2
138	WP_010815375.1	428	6	Asn	Firmicutes	Enterococcus faecalis
139	WP_003646469.1[b]	446	6	Asn	Firmicutes	Lactobacillus plantarum
140	YP_005004345.1	432	6	Asn	Firmicutes	Pediococcus claussenii ATCC BAA-344
141	WP_007124555.1	453	6	Asn	Firmicutes	Lactobacillus antri
142	WP_003641980.1[b,c]	442	6	Asn	Firmicutes	Lactobacillus plantarum
143	YP_007986684.1	420	6	Asn	Firmicutes	Lactobacillus plantarum subsp. plantarum P-8
144	YP_003621756.1	435	6	Asn	Firmicutes	Leuconostoc kimchii IMSNU 11154
145	WP_004070366.1	454	6	Asn	Firmicutes	Clostridium sp. ASF502
146	WP_003637608.1	440	6	Asn	Firmicutes	Lactobacillus pentosus
147	WP_003637610.1	442	6	Asn	Firmicutes	Lactobacillus pentosus
148	WP_007060938.1	432	6	Asn	Firmicutes	Clostridium carboxidivorans
149	YP_003842389.1	434	6	Asn	Firmicutes	Clostridium cellulovorans 743B
150	WP_016293822.1	457	6	Asn	Firmicutes	Lachnospiraceae bacterium M18-1
151	WP_002834212.1	434	6	Asn	Firmicutes	Pediococcus pentosaceus
152	WP_010751948.1	433	6	Asn	Firmicutes	Enterococcus villorum
153	WP_003695050.1	439	6	Asn	Firmicutes	Lactobacillus ruminis
154	WP_016184472.1	427	6	Asn	Firmicutes	Enterococcus columbae
155	YP_005047496.1	441	6	Asn	Firmicutes	Clostridium clariflavum DSM 19732
156	WP_004908337.1	434	6	Asn	Firmicutes	Leuconostoc citreum
157	WP_002328140.1	427	6	Asn	Firmicutes	Enterococcus faecium
158	WP_008469780.1	444	6	Asn	Firmicutes	Lactobacillus hominis
159	YP_006796412.1	432	6	Asn	Firmicutes	Leuconostoc carnosum JB16
160	WP_003708561.1	430	6	Asn	Firmicutes	Lactobacillus salivarius
161	WP_007126030.1	445	6	Asn	Firmicutes	Lactobacillus ultunensis
162	WP_010690030.1	420	6	Asn	Firmicutes	Lactobacillus animalis
163	WP_003665180.1	420	6	Asn	Firmicutes	Lactobacillus reuteri
164	NP_964473.1	419	6	Asn	Firmicutes	Lactobacillus johnsonii NCC 533
165	WP_003692144.1	416	6	Asn	Firmicutes	Lactobacillus ruminis
166	WP_003699415.1	430	6	Asn	Firmicutes	Lactobacillus salivarius
167	YP_193512.1	426	6	Asn	Firmicutes	Lactobacillus acidophilus NCFM
168	WP_010018224.1	435	6	Asn	Firmicutes	Lactobacillus farciminis
169	WP_005720992.1	442	6	Asn	Firmicutes	Lactobacillus crispatus
170	WP_009558353.1	440	6	Asn	Firmicutes	Lactobacillus sp. 66c
171	WP_020806829.1	439	6	Asn	Firmicutes	Lactobacillus gasseri
172	WP_002878748.1	443	6	Asn	Firmicutes	Lactobacillus delbrueckii
173	WP_003616411.1	451	6	Asn	Firmicutes	Lactobacillus delbrueckii
174	YP_004293001.1	442	6	Asn	Firmicutes	Lactobacillus acidophilus 30SC
175	WP_010727102.1	419	6	Asn	Firmicutes	Enterococcus faecium
176	WP_016624980.1	432	6	Asn	Firmicutes	Enterococcus faecalis
177	WP_010816586.1	414	6	Asn	Firmicutes	Enterococcus faecalis
178	WP_003616444.1[b]	417	6	Asn	Firmicutes	Lactobacillus delbrueckii
179	WP_002373740.1	498	6	Asn	Firmicutes	Enterococcus faecium
180	WP_010751677.1	412	6	Asn	Firmicutes	Enterococcus villorum
181	WP_009461668.1	423	6	Asn	Firmicutes	Lachnospiraceae bacterium 2_1_46FAA
182	WP_007714334.1	438	6	Asn	Firmicutes	Clostridium asparagiforme
183	WP_009982749.1	402	6	Asn	Firmicutes	Ruminococcus flavefaciens
184	WP_010761975.1	430	6	Asn	Firmicutes	Enterococcus haemoperoxidus
185	WP_009172438.1	418	6	Asn	Firmicutes	Clostridium sp. DL-VIII
186	WP_010718901.1	420	6	Asn	Firmicutes	Enterococcus faecium
187	WP_010752108.1	420	6	Asn	Firmicutes	Enterococcus villorum

188	WP_016250704.1	435	6	Asn	Firmicutes	Enterococcus cecorum
189	WP_005338206.1	435	6	Asn	Firmicutes	Dorea formicigenerans
190	WP_009982643.1	427	6	Asn	Firmicutes	Ruminococcus flavefaciens
191	WP_005215399.1	427	6	Asn	Firmicutes	Clostridium celatum
192	WP_005923174.1	438	6	Asn	Firmicutes	Faecalibacterium prausnitzii
193	YP_001558714.1	446	6	Asn	Firmicutes	Clostridium phytofermentans ISDg
194	WP_004610893.1	408	6	Asn	Firmicutes	Clostridium spiroforme
195	WP_007867475.1	435	6	Asn	Firmicutes	Clostridium citroniae
196	YP_001332979.1[a, c]	430	6	Asn	Firmicutes	Staphylococcus aureus subsp. aureus str. Newman
197	WP_000387802.1	430	6	Asn	Firmicutes	Staphylococcus aureus
198	WP_002464788.1	429	6	Asn	Firmicutes	Staphylococcus simiae
199	WP_002488053.1	429	6	Asn	Firmicutes	Staphylococcus hominis
200	WP_002469368.1	432	6	Asn	Firmicutes	Staphylococcus epidermidis
201	YP_007384612.1	429	6	Asn	Firmicutes	Staphylococcus warneri SG1
202	WP_002481722.1	433	6	Asn	Firmicutes	Staphylococcus simulans
203	WP_002509116.1	432	6	Asn	Firmicutes	Staphylococcus arlettae
204	WP_002511407.1	431	6	Asn	Firmicutes	Staphylococcus equorum
205	YP_004149866.1	430	6	Asn	Firmicutes	Staphylococcus pseudintermedius HKU10-03
206	WP_007474155.1	435	6	Asn	Firmicutes	Listeria fleischmannii
207	WP_000263507.1	431	6	Asn	Firmicutes	Bacillus cereus
208	YP_006464114.1	433	6	Asn	Firmicutes	Solibacillus silvestris StLB046
209	WP_002598585.1[b]	428	6	Asn	Firmicutes	Clostridium colicanis
210	WP_010631760.1	437	6	Asn	Firmicutes	Sporolactobacillus vineae
211	WP_002200091.1	447	6	Asn	Firmicutes	Bacillus cereus
212	NP_346725.1	447	7	Asn	Firmicutes	Clostridium acetobutylicum ATCC 824
213	YP_003845678.1	434	7	Asn	Firmicutes	Clostridium cellulovorans 743B
214	WP_010294543.1	445	7	Asn	Firmicutes	Clostridium sp. JC122
215	WP_007062817.1	465	7	Asn	Firmicutes	Clostridium carboxidivorans
216	YP_007454888.1	454	7	Asn	Firmicutes	Clostridium saccharoperbutylacetonicum N1-4(HMT)
217	WP_018371070.1	455	7	Asn	Firmicutes	Streptococcus massiliensis
218	WP_009170117.1	441	7	Asn	Firmicutes	Clostridium sp. DL-VIII
219	YP_006787706.1	416	7	Asn	Firmicutes	Clostridium acidurici 9a
220	YP_007939695.1	444	7	Asn	Firmicutes	Clostridium pasteurianum BC1
221	WP_010826789.1	442	7	Asn	Firmicutes	Enterococcus faecalis
222	WP_002314968.1	425	7	Asn	Firmicutes	Enterococcus faecium
223	YP_001451369.1[c]	453	7	Asn	Firmicutes	Streptococcus gordonii str. Challis substr. CH1
224	YP_003822135.1	437	7	Asn	Firmicutes	Clostridium saccharolyticum WM1
225	WP_009253846.1	441	7	Asn	Firmicutes	Lachnospiraceae bacterium 3_1_57FAA_CT1
226	YP_007939348.1	444	7	Asn	Firmicutes	Clostridium pasteurianum BC1
227	WP_010720219.1	425	7	Asn	Firmicutes	Enterococcus faecium
228	WP_018380441.1	444	7	Asn	Firmicutes	Streptococcus thoraltensis
229	WP_010723115.1	426	7	Asn	Firmicutes	Enterococcus faecium
230	WP_005590517.1	444	7	Asn	Firmicutes	Streptococcus cristatus
231	WP_003075490.1[c]	441	7	Asn	Firmicutes	Streptococcus intermedius
232	WP_018380443.1	452	7	Asn	Firmicutes	Streptococcus thoraltensis
233	WP_002830271.1	432	7	Asn	Firmicutes	Pediococcus acidilactici
234	WP_003071652.1	440	7	Asn	Firmicutes	Streptococcus intermedius
235	WP_000345272.1	439	7	Asn	Firmicutes	Streptococcus sp. oral taxon 058
236	WP_002907623.1	444	7	Asn	Firmicutes	Streptococcus sanguinis
237	WP_005591383.1	447	7	Asn	Firmicutes	Streptococcus cristatus
238	WP_001054986.1	439	7	Asn	Firmicutes	Streptococcus sp. SK140
239	WP_001048121.1[b]	439	7	Asn	Firmicutes	Streptococcus mitis
240	WP_004455149.1	446	7	Asn	Firmicutes	Clostridium pasteurianum
241	WP_006149098.1	440	7	Asn	Firmicutes	Streptococcus infantis
242	WP_018367818.1	447	7	Asn	Firmicutes	Streptococcus entericus
243	WP_003365114.1	440	7	Asn	Firmicutes	Clostridium botulinum
244	WP_000362886.1	439	7	Asn	Firmicutes	Streptococcus sp. C300
245	WP_006531892.1	439	7	Asn	Firmicutes	Streptococcus infantarius
246	WP_004261048.1	449	7	Asn	Firmicutes	Streptococcus mitis
247	WP_006595333.1[b]	441	7	Asn	Firmicutes	Streptococcus australis
248	WP_000402607.1	445	7	Asn	Firmicutes	Streptococcus sp. SK643
249	YP_006904098.1	444	7	Asn	Firmicutes	Streptococcus dysgalactiae subsp. equisimilis AC-2713
250	WP_010758329.1	433	7	Asn	Firmicutes	Enterococcus pallens

251	WP_019278667.1	438	7	Asn	Firmicutes	Clostridium botulinum
252	WP_018369941.1	444	7	Asn	Firmicutes	Streptococcus marimammalium
253	WP_009004071.1	466	7	Asn	Firmicutes	Clostridium sp. D5
254	YP_006003161.1	447	7	Asn	Firmicutes	Streptococcus thermophilus ND03
255	YP_002743992.1	444	7	Asn	Firmicutes	Streptococcus equi subsp. zooepidemicus
256	YP_006082817.1	447	7	Asn	Firmicutes	Streptococcus suis D12
257	WP_010756667.1	423	7	Asn	Firmicutes	Enterococcus pallens
258	WP_003643808.1	446	7	Asn	Firmicutes	Lactobacillus plantarum
259	WP_003025428.1	435	7	Asn	Firmicutes	Streptococcus anginosus
260	WP_016220081.1	432	7	Asn	Firmicutes	Dorea sp. 5-2
261	WP_000231528.1	441	7	Asn	Firmicutes	Streptococcus mitis
262	WP_003073361.1	438	7	Asn	Firmicutes	Streptococcus intermedius
263	WP_016287680.1	458	7	Asn	Firmicutes	Lachnospiraceae bacterium 3-1
264	YP_003621251.1[b]	437	7	Asn	Firmicutes	Leuconostoc kimchii IMSNU 11154
265	WP_017285146.1	444	7	Asn	Firmicutes	Streptococcus agalactiae
266	WP_016307158.1	454	7	Asn	Firmicutes	Lachnospiraceae bacterium A2
267	WP_016282345.1	450	7	Asn	Firmicutes	Lachnospiraceae bacterium A4
268	WP_016220284.1	454	7	Asn	Firmicutes	Dorea sp. 5-2
269	WP_010770640.1	422	7	Asn	Firmicutes	Enterococcus caccae
270	WP_016219150.1	513	7	Asn	Firmicutes	Dorea sp. 5-2
271	YP_006744460.1	435	7	Asn	Firmicutes	Leuconostoc gelidum JB7
272	WP_016295538.1	448	7	Asn	Firmicutes	Lachnospiraceae bacterium M18-1
273	WP_010707474.1	475	7	Asn	Firmicutes	Enterococcus faecalis
274	WP_003639720.1	442	7	Asn	Firmicutes	Lactobacillus pentosus
275	WP_004079905.1	446	7	Asn	Firmicutes	Clostridium sp. ASF502
276	WP_016623052.1	430	7	Asn	Firmicutes	Enterococcus faecalis
277	WP_016625521.1	424	7	Asn	Firmicutes	Enterococcus faecalis
278	WP_004635172.1	439	7	Asn	Firmicutes	Dolosigranulum pigrum
279	WP_004635179.1	442	7	Asn	Firmicutes	Dolosigranulum pigrum
280	WP_010770253.1	409	7	Asn	Firmicutes	Enterococcus caccae
281	YP_536800.1	429	7	Asn	Firmicutes	Lactobacillus salivarius UCC118
282	WP_004049677.1	442	7	Asn	Firmicutes	Lactobacillus murinus
283	YP_007988137.1	445	7	Asn	Firmicutes	Lactobacillus plantarum subsp. plantarum P-8
284	YP_005861339.1[b]	443	7	Asn	Firmicutes	Lactobacillus johnsonii DPC 6026
285	WP_010771509.1	451	7	Asn	Firmicutes	Enterococcus caccae
286	YP_003602296.1	447	7	Asn	Firmicutes	Lactobacillus crispatus ST1
287	YP_194634.1	440	7	Asn	Firmicutes	Lactobacillus acidophilus NCFM
288	WP_009896878.1	452	7	Asn	Firmicutes	Clostridium difficile
289	WP_005718161.1	468	7	Asn	Firmicutes	Lactobacillus crispatus
290	WP_010762052.1	436	7	Asn	Firmicutes	Enterococcus haemoperoxidus
291	WP_008680295.1	416	7	Asn	Firmicutes	Clostridium sp. 7_2_43FAA
292	WP_016208990.1	419	7	Asn	Firmicutes	Clostridium sartagoforme
293	YP_007828746.1	442	7	Asn	Firmicutes	Ruminococcus champanellensis 18P13
294	WP_009985441.1[b]	440	7	Asn	Firmicutes	Ruminococcus flavefaciens
295	WP_008423253.1	420	7	Asn	Firmicutes	Clostridium sp. Maddingley MBC34-26
296	YP_004104667.1	425	7	Asn	Firmicutes	Ruminococcus albus 7
297	WP_006445121.1	445	7	Asn	Firmicutes	Clostridium hylemonae
298	WP_005361465.1	425	7	Asn	Firmicutes	Eubacterium ventriosum
299	WP_002586250.1	434	7	Asn	Firmicutes	Clostridium clostridioforme
300	YP_007458077.1	416	7	Asn	Firmicutes	Clostridium saccharoperbutylacetonicum N1-4(HMT)
301	WP_007863964.1	445	7	Asn	Firmicutes	Clostridium citroniae
302	WP_006782132.1	432	7	Asn	Firmicutes	Clostridium hathewayi
303	WP_006863055.1	460	7	Asn	Firmicutes	Marvinbryantia formatexigens
304	WP_005331796.1	435	7	Asn	Firmicutes	Dorea formicigenerans
305	WP_019679521.1	434	7	Asn	Firmicutes	Ruminococcus flavefaciens
306	WP_003480616.1	420	7	Asn	Firmicutes	Clostridium perfringens
307	WP_002847701.1	439	7	Asn	Firmicutes	Ruminococcus albus
308	YP_004103515.1	448	7	Asn	Firmicutes	Ruminococcus albus 7
309	WP_005353871.1	444	7	Asn	Firmicutes	Eubacterium siraeum
310	YP_004545147.1[b]	439	7	Asn	Firmicutes	Desulfotomaculum ruminis DSM 2154
311	YP_007790596.1	440	7	Asn	Firmicutes	butyrate-producing bacterium SSC/2
312	WP_018702136.1	445	7	Asn	Firmicutes	Anaeromusa acidaminophila
313	WP_016227154.1	438	7	Asn	Firmicutes	unclassified Lachnospiraceae
314	WP_004801384.1	421	7	Asn	Firmicutes	Eggerthia catenaformis
315	WP_016184116.1	434	7	Asn	Firmicutes	Enterococcus columbae

316	WP_016316494.1	445	7	Asn	Firmicutes	Anaerotruncus sp. G3(2012)
317	WP_018579518.1	453	7	Asn	Firmicutes	Erysipelothrix tonsillarum
318	YP_003820510.1	434	7	Asn	Firmicutes	Clostridium saccharolyticum WM1
319	WP_000447891.1	427	7	Asn	Firmicutes	Staphylococcus aureus
320	WP_002477479.1[c]	429	7	Asn	Firmicutes	Staphylococcus epidermidis
321	YP_003471243.1	429	7	Asn	Firmicutes	Staphylococcus lugdunensis HKU09-01
322	YP_006677599.1	436	7	Asn	Firmicutes	Listeria monocytogenes SLCC2540
323	WP_002151242.1	439	7	Asn	Firmicutes	Bacillus cereus
324	WP_007124101.1	450	7	Asn	Firmicutes	Lactobacillus antri
325	YP_001621389.1	444	7	Asn	Firmicutes[d]	Acholeplasma laidlawii PG-8A
326	YP_003821953.1	650	8	Asn	Firmicutes	Clostridium saccharolyticum WM1
non-HPK10 hits						
327	WP_017024785.1	555	2	Asp	Proteobacteria	Vibrio rumoiensis
328	WP_004396401.1	551	2	Asp	Proteobacteria	Vibrio metschnikovii
329	WP_016545474.1	802	0	Asp	Spirochaetes	Leptospira wolffii
330	YP_004340943.1	433	0	Asp	Euryarchaeota	Archaeoglobus veneficus SNP6

a: The query sequence for the first round of database search;

b: Query sequences for the second round of database search;

c: Founding members of the HPK10 subfamily (125). 7 sequences in the table covers 12 out of 15 founding with at >75% sequence identity. The rest 3 are not deposited in the Refseq database.

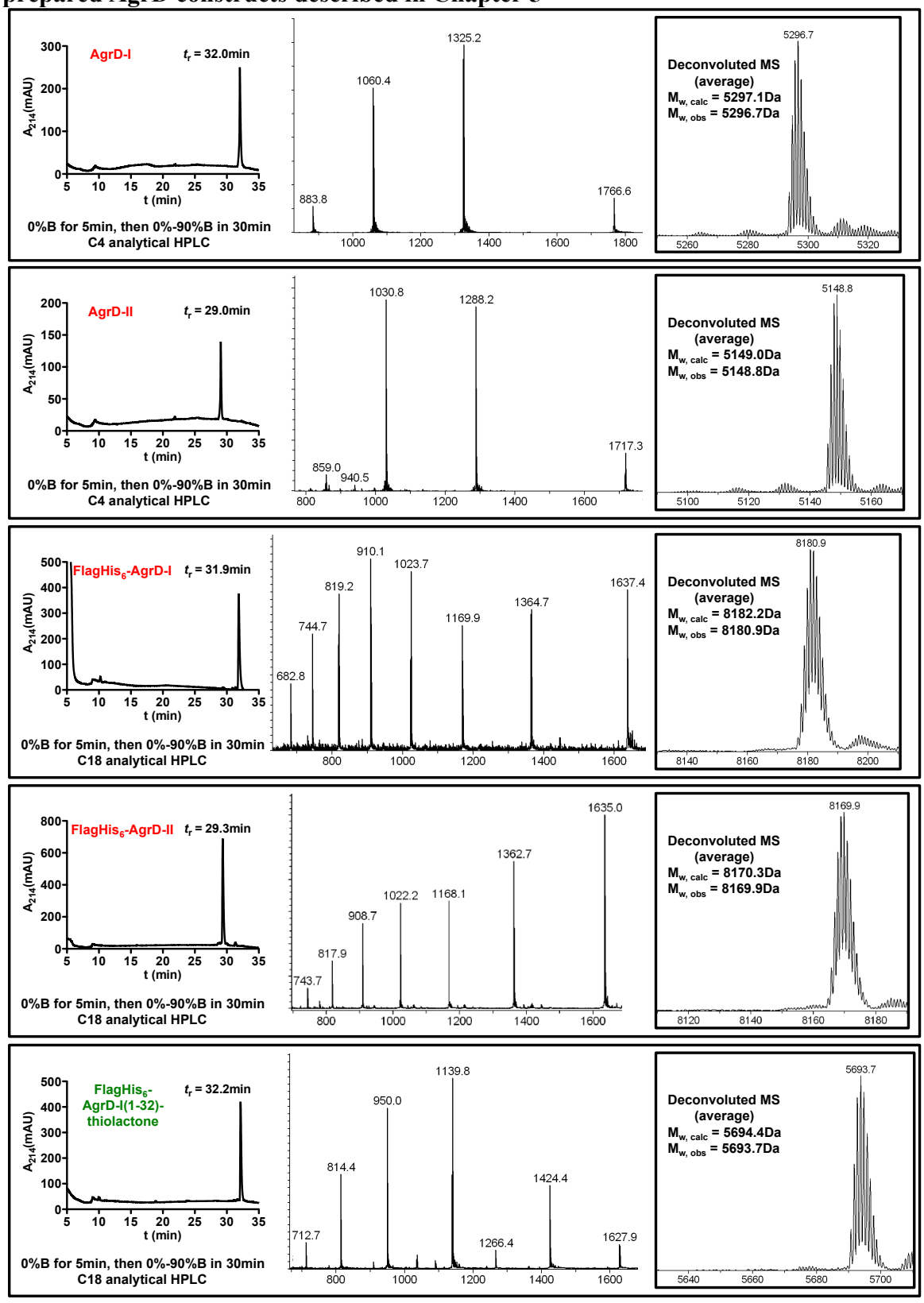
d: There is controversy on whether the class *Mollicutes* should be placed in the phylum *Firmicutes* or *Tenericutes* (193).

Appendix Table 2. AgrD constructs prepared as starting materials or detected as products in Chapter 5.

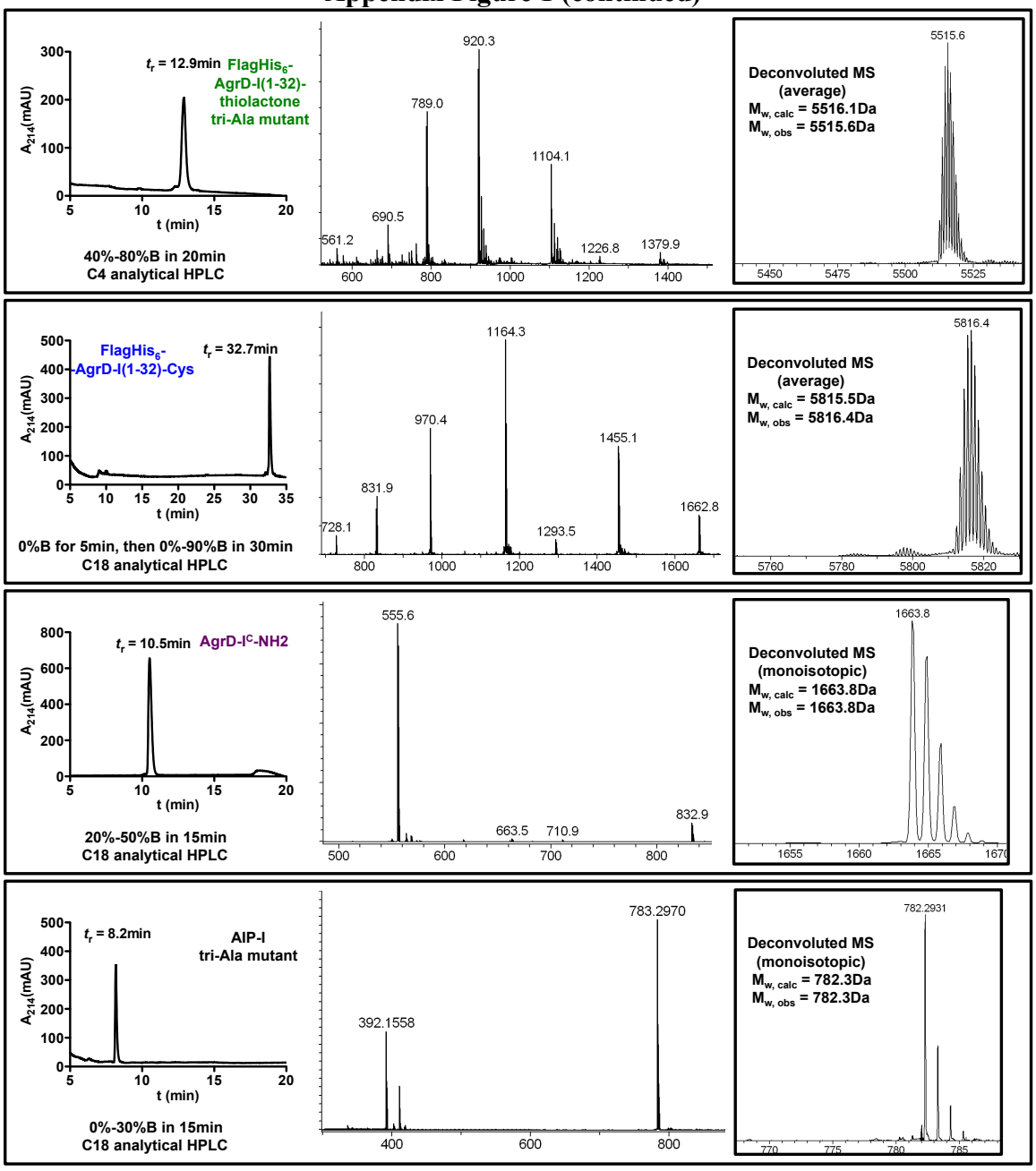
Entry	Construct Name	Tags/Terminal Groups
	R_1-AgrD-**R_2**	
1*	AgrD-I	**R_1**= H-, **R_2** = -OH
2*	AgrD-II	**R_1**= H-, **R_2** = -OH
3*	FlagHis$_6$-AgrD-I	**R_1**= H-MDYKDDDDKHHHHHHGG-; **R_2** = -LDKLVPR-OH
4*	FlagHis$_6$-AgrD-II	**R_1**= H-MDYKDDDDKHHHHHHGG-; **R_2** = -LDLELVFQ-OH
5	FlagHis$_6$-AgrD-I-NH_2	**R_1**= H-MDYKDDDDKHHHHHHGG-; **R_2** = -NH_2
	AgrD^C-**R_2**	
6	AgrD-I^C	**R_2** = -OH
7	AgrD-II^C	**R_2** = -OH
8	AgrD-I^C-LDKLVPR	**R_2** = -LDKLVPR-OH
9**	AgrD-I^C-NH_2	**R_2** = -NH_2
	R_1-AgrD(1-32)-thiolactone	
10	AgrD-I(1-32)-thiolactone	**R_1**= H-,
11	AgrD-I(1-32)-thiolactone	**R_1**= H-,
12*	FlagHis$_6$-AgrD-I-(1-32)-thiolactone	**R_1**= H-MDYKDDDDKHHHHHHGG-;
	R_1-AgrD(1-32)	
13	AgrD-I(1-32)	**R_1**= H-,
14	AgrD-II(1-32)	**R_1**= H-,
15	FlagHis$_6$-AgrD-I(1-32)	**R_1**= H-MDYKDDDDKHHHHHHGG-;
	Others	
16**	AIP-I	H-YST(CDFIM)***
17**	AIP-I tri-Ala mutant	H-YST(CDAAA)***
18*	FlagHis$_6$-AgrD-I-(1-32)-thiolactone tri-Ala mutant	H-MDYKDDDDKHHHHHHGGMNTLFNL LFELITGILKNIGNIAAYST(CDAAA)***

The sequence of full-length AgrD-I (residues 1-46) is
MNTLFNLFFDFITGILKNIGNIAAYSTCDFIMDEVEVPKELTQLHE
The sequence of full-length AgrD-II (residues 1-47) is
MNTLVNMFFDFIIKLAKAIGIVGGVNACSSLFDEPKVPAELTNLYDK
*: constructs prepared from recombinant proteins;
**: constructs synthesized chemically;
***: parenthesis indicates residues that form the thiolactone macrocycle.

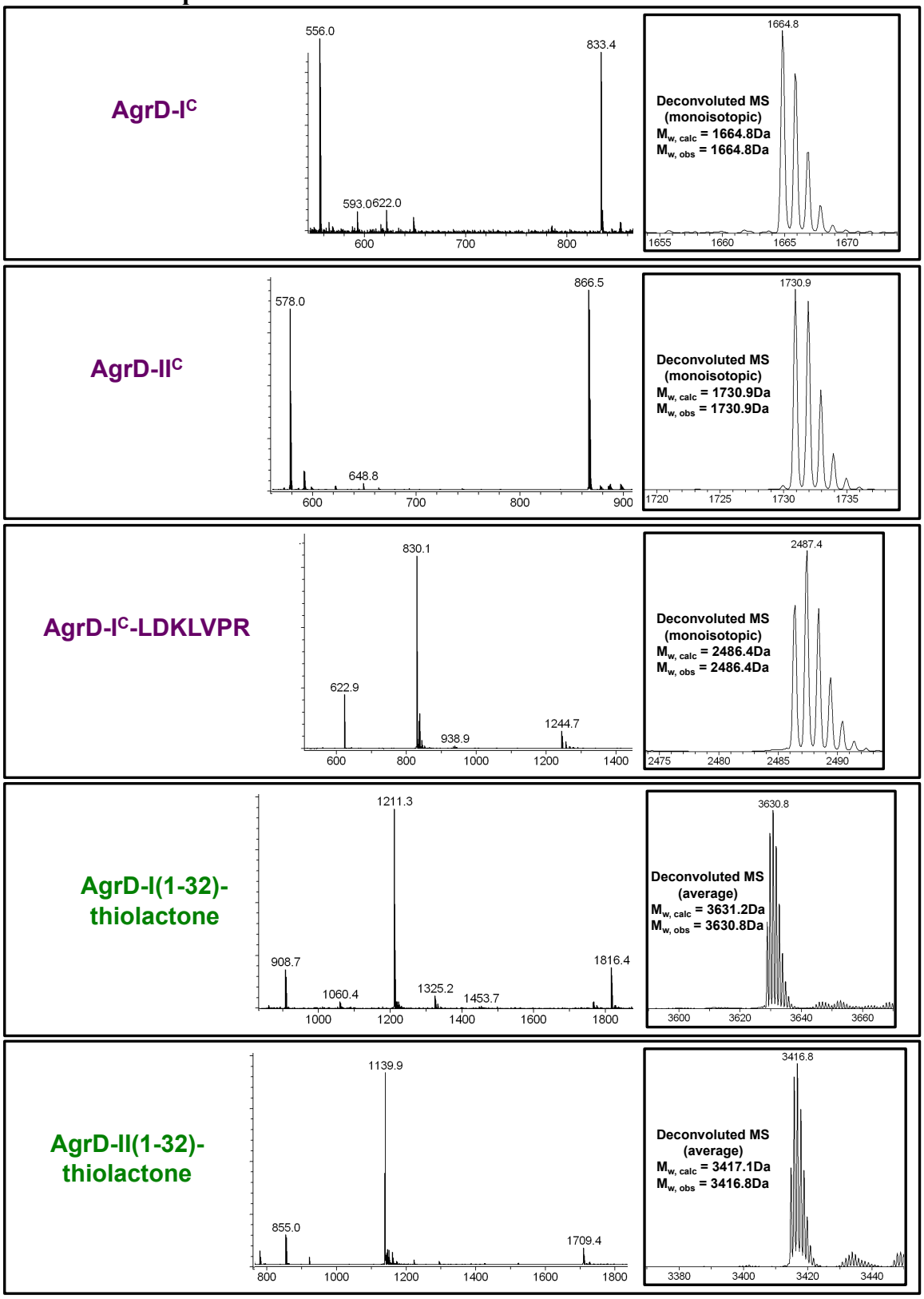
Appendix Figure 1. RP-HPLC-MS results of all synthetic or recombinantly prepared AgrD constructs described in Chapter 5



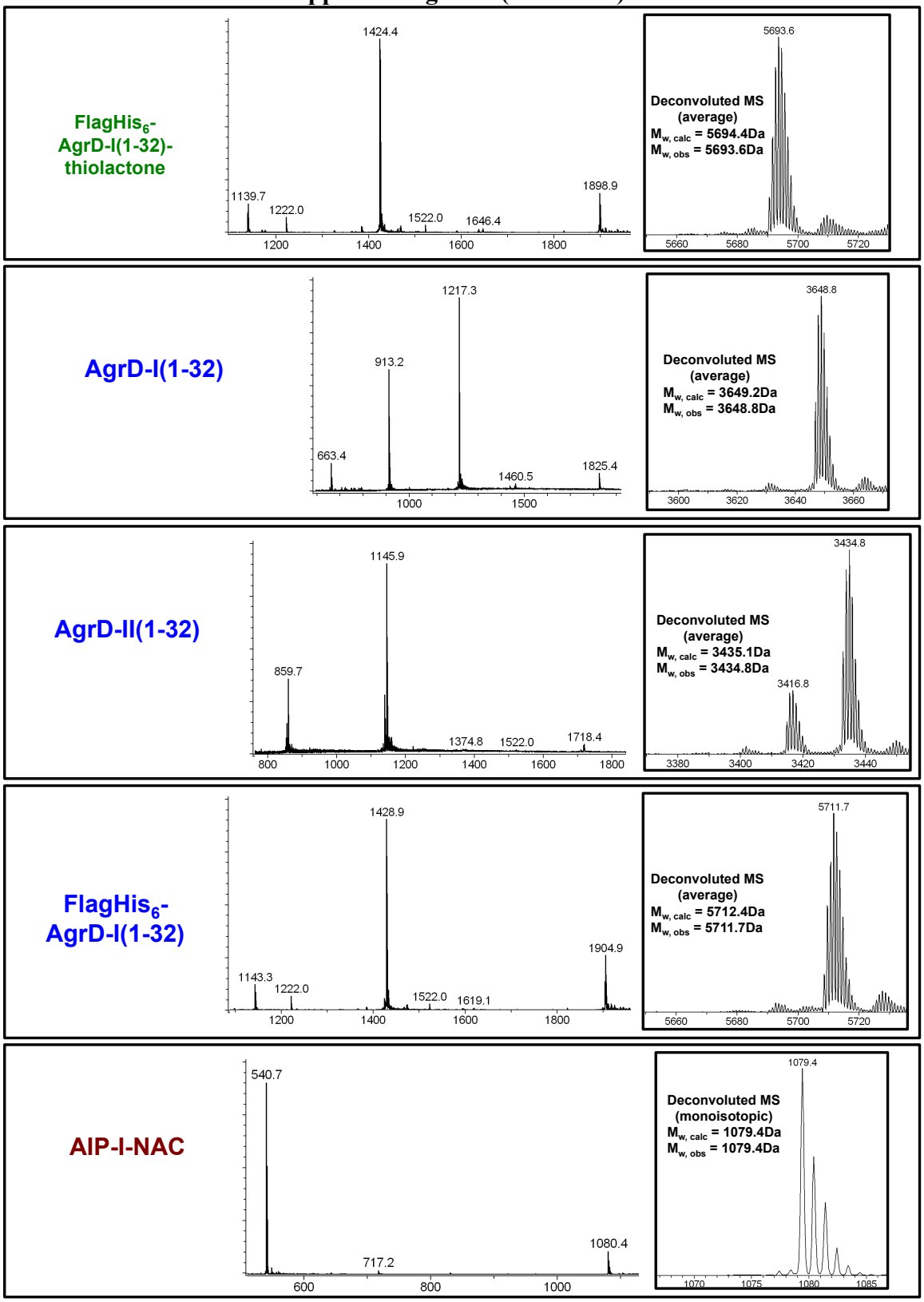
Appendix Figure 1 (continued)



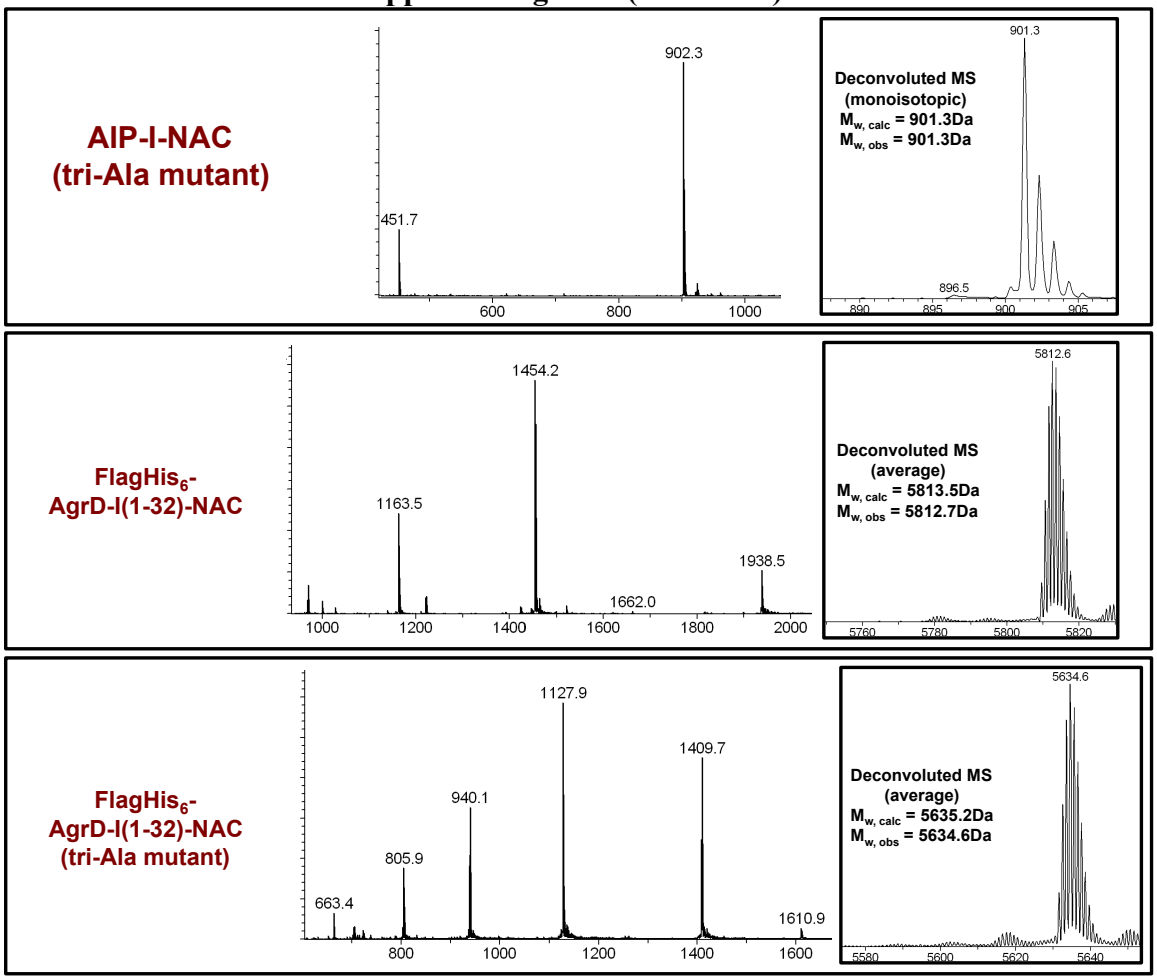
Appendix Figure 2. MS results of all AgrD products from the biochemical reactions described in Chapter 5.



Appendix Figure 2 (continued)



Appendix Figure 2 (continued)



References

1. Nealson KH & Hastings JW (1979) Bacterial Bioluminescence - Its Control and Ecological Significance. *Microbiol Rev* 43(4):496-518.
2. Eberhard A*, et al.* (1981) Structural Identification of Autoinducer of Photobacterium-Fischeri Luciferase. *Biochemistry-Us* 20(9):2444-2449.
3. Engebrecht J, Nealson K, & Silverman M (1983) Bacterial Bioluminescence - Isolation and Genetic-Analysis of Functions from Vibrio Fischeri. *Cell* 32(3):773-781.
4. Engebrecht J & Silverman M (1984) Identification of Genes and Gene-Products Necessary for Bacterial Bioluminescence. *P Natl Acad Sci-Biol* 81(13):4154-4158.
5. Fuqua WC, Winans SC, & Greenberg EP (1994) Quorum Sensing in Bacteria - the Luxr-Luxi Family of Cell Density-Responsive Transcriptional Regulators. *J Bacteriol* 176(2):269-275.
6. Waters CM & Bassler BL (2005) Quorum sensing: Cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology* 21:319-346.
7. Miller MB & Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* 55:165-199.
8. Shapiro JA (1998) Thinking about bacterial populations as multicellular organisms. *Annu Rev Microbiol* 52:81-104.
9. Parsek MR & Greenberg EP (2005) Sociomicrobiology: the connections between quorum sensing and biofilms. *Trends Microbiol* 13(1):27-33.
10. Kuipers OP, de Ruyter PGGA, Kleerebezem M, & de Vos WM (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *J Biotechnol* 64(1):15-21.
11. Zhu J*, et al.* (2002) Quorum-sensing regulators control virulence gene expression in Vibrio cholerae. *P Natl Acad Sci USA* 99(5):3129-3134.
12. Novick RP (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol Microbiol* 48(6):1429-1449.
13. Dubnau D & Losick R (2006) Bistability in bacteria. *Mol Microbiol* 61(3):564-572.
14. Hammerschmidt K, Rose CJ, Kerr B, & Rainey PB (2014) Life cycles, fitness decoupling and the evolution of multicellularity. *Nature* 515(7525):75-+.
15. Dufour P*, et al.* (2002) High genetic variability of the agr locus in Staphylococcus species. *J Bacteriol* 184(4):1180-1186.
16. Schuster M, Sexton DJ, Diggle SP, & Greenberg EP (2013) Acyl-Homoserine Lactone Quorum Sensing: From Evolution to Application. *Annual Review of Microbiology, Vol 67* 67:43-63.
17. Lyon GJ & Novick RP (2004) Peptide signaling in Staphylococcus aureus and other Gram-positive bacteria. *Peptides* 25(9):1389-1403.
18. Okada M*, et al.* (2005) Structure of the Bacillus subtilis quorum-sensing peptide pheromone ComX. *Nat Chem Biol* 1(1):23-24.
19. Pottathil M & Lazazzera BA (2003) The extracellular Phr peptide-rap phosphatase signaling circuit of Bacillus subtilis. *Front Biosci* 8:D32-D45.

20. Ji GY, Beavis R, & Novick RP (1997) Bacterial interference caused by autoinducing peptide variants. *Science* 276(5321):2027-2030.
21. Chen X*, et al.* (2002) Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* 415(6871):545-549.
22. Pesci EC*, et al.* (1999) Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa. *P Natl Acad Sci USA* 96(20):11229-11234.
23. Takano E (2006) gamma-Butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol* 9(3):287-294.
24. Manefield M & Turner SL (2002) Quorum sensing in context: out of molecular biology and into microbial ecology. *Microbiol-Sgm* 148:3762-3764.
25. Mayville P*, et al.* (1999) Structure-activity analysis of synthetic autoinducing thiolactone peptides from Staphylococcus aureus responsible for virulence. *P Natl Acad Sci USA* 96(4):1218-1223.
26. Ng WL & Bassler BL (2009) Bacterial Quorum-Sensing Network Architectures. *Annu Rev Genet* 43:197-222.
27. Lyon GJ & Muir TW (2003) Chemical signaling among bacteria and its inhibition. *Chem Biol* 10(11):1007-1021.
28. Hibbing ME, Fuqua C, Parsek MR, & Peterson SB (2010) Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* 8(1):15-25.
29. Ziebandt AK*, et al.* (2004) The influence of agr and sigma(B) in growth phase dependent regulation of virulence factors in Staphylococcus aureus. *Proteomics* 4(10):3034-3047.
30. Wright JS, Jin R, & Novick RP (2005) Transient interference with staphylococcal quorum sensing blocks abscess formation. *P Natl Acad Sci USA* 102(5):1691-1696.
31. Coleman G, Jakeman CM, & Martin N (1978) Patterns of Total Extracellular Protein Secretion by a Number of Clinically Isolated Strains of Staphylococcus-Aureus. *J Gen Microbiol* 107(Jul):189-192.
32. Mallonee DH, Glatz BA, & Pattee PA (1982) Chromosomal Mapping of a Gene Affecting Enterotoxin-a Production in Staphylococcus-Aureus. *Appl Environ Microb* 43(2):397-402.
33. Recsei P*, et al.* (1986) Regulation of Exoprotein Gene-Expression in Staphylococcus-Aureus by Agr. *Mol Gen Genet* 202(1):58-61.
34. Peng HL, Novick RP, Kreiswirth B, Kornblum J, & Schlievert P (1988) Cloning, Characterization, and Sequencing of an Accessory Gene Regulator (Agr) in Staphylococcus-Aureus. *J Bacteriol* 170(9):4365-4372.
35. Morfeldt E, Janzon L, Arvidson S, & Lofdahl S (1988) Cloning of a Chromosomal Locus (Exp) Which Regulates the Expression of Several Exoprotein Genes in Staphylococcus-Aureus. *Mol Gen Genet* 211(3):435-440.
36. Janzon L & Arvidson S (1990) The Role of the Delta-Lysin Gene (Hld) in the Regulation of Virulence Genes by the Accessory Gene Regulator (Agr) in Staphylococcus-Aureus. *Embo J* 9(5):1391-1399.
37. Novick RP*, et al.* (1993) Synthesis of Staphylococcal Virulence Factors Is Controlled by a Regulatory Rna Molecule. *Embo J* 12(10):3967-3975.

38. Novick RP*, et al.* (1995) The Agr P2 Operon - an Autocatalytic Sensory Transduction System in Staphylococcus-Aureus. *Mol Gen Genet* 248(4):446-458.
39. Ji GY, Beavis RC, & Novick RP (1995) Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *P Natl Acad Sci USA* 92(26):12055-12059.
40. Huntzinger E*, et al.* (2005) Staphylococcus aureus RNAIII and the endoribonuclease III coordinately regulate spa gene expression. *Embo J* 24(4):824-835.
41. Geisinger E, Adhikari RP, Jin RZ, Ross HF, & Novick RP (2006) Inhibition of rot translation by RNAIII, a key feature of agr function. *Mol Microbiol* 61(4):1038-1048.
42. Boisset S*, et al.* (2007) Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Gene Dev* 21(11):1353-1366.
43. Said-Salim B*, et al.* (2003) Global regulation of Staphylococcus aureus genes by rot. *J Bacteriol* 185(2):610-619.
44. Cheung GYC, Wang R, Khan BA, Sturdevant DE, & Otto M (2011) Role of the Accessory Gene Regulator agr in Community-Associated Methicillin-Resistant Staphylococcus aureus Pathogenesis. *Infect Immun* 79(5):1927-1935.
45. Queck SY*, et al.* (2008) RNAIII-Independent Target Gene Control by the agr Quorum-Sensing System: Insight into the Evolution of Virulence Regulation in Staphylococcus aureus. *Mol Cell* 32(1):150-158.
46. Chatterjee SS*, et al.* (2013) Essential Staphylococcus aureus toxin export system. *Nat Med* 19(3):364-367.
47. Jarraud S*, et al.* (2000) Exfoliatin-producing strains define a fourth agr specificity group in Staphylococcus aureus. *J Bacteriol* 182(22):6517-6522.
48. Lyon GJ, Wright JS, Muir TW, & Novick RP (2002) Key determinants of receptor activation in the agr autoinducing peptides of Staphylococcus aureus. *Biochemistry-Us* 41(31):10095-10104.
49. Traber KE*, et al.* (2008) agr function in clinical Staphylococcus aureus isolates. *Microbiol-Sgm* 154:2265-2274.
50. Geisinger E, Chen J, & Novick RP (2012) Allele-Dependent Differences in Quorum-Sensing Dynamics Result in Variant Expression of Virulence Genes in Staphylococcus aureus. *J Bacteriol* 194(11):2854-2864.
51. Shine J & Dalgarno L (1975) Determinant of Cistron Specificity in Bacterial Ribosomes. *Nature* 254(5495):34-38.
52. Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361:13-37.
53. Libby EA, Roggiani M, & Goulian M (2012) Membrane protein expression triggers chromosomal locus repositioning in bacteria. *P Natl Acad Sci USA* 109(19):7445-7450.
54. Qiu RD, Pei WH, Zhang LS, Lin JQ, & Ji GY (2005) Identification of the putative staphylococcal AgrB catalytic residues involving the proteolytic cleavage of AgrD to generate autoinducing peptide. *J Biol Chem* 280(17):16695-16704.

55. Thoendel M & Horswill AR (2009) Identification of Staphylococcus aureus AgrD Residues Required for Autoinducing Peptide Biosynthesis. *J Biol Chem* 284(33):21828-21838.
56. Zhang LS, Lin JQ, & Ji GY (2004) Membrane anchoring of the AgrD N-terminal amphipathic region is required for its processing to produce a quorum-sensing pheromone in Staphylococcus aureus. *J Biol Chem* 279(19):19448-19456.
57. Kavanaugh JS, Thoendel M, & Horswill AR (2007) A role for type I signal peptidase in Staphylococcus aureus quorum sensing. *Mol Microbiol* 65(3):780-798.
58. Thoendel M & Horswill AR (2013) Random mutagenesis and topology analysis of the autoinducing peptide biosynthesis proteins in Staphylococcus aureus. *Mol Microbiol* 87(2):318-337.
59. Zhang LS, Gray L, Novick RP, & Ji GY (2002) Transmembrane topology of AgrB, the protein involved in the post-translational modification of AgrD in Staphylococcus aureus. *J Biol Chem* 277(38):34736-34742.
60. White SH & von Heijne G (2008) How translocons select transmembrane helices. *Annu Rev Biophys* 37:23-42.
61. von Heijne G (2006) Membrane-protein topology. *Nat Rev Mol Cell Bio* 7(12):909-918.
62. Zhang LS & Ji GY (2004) Identification of a staphylococcal AgrB segment(s) responsible for group-specific processing of AgrD by gene swapping. *J Bacteriol* 186(20):6706-6713.
63. Gonzalez DJ*, et al.* (2014) N-Terminal ArgD Peptides from the Classical Staphylococcus aureus Agr System Have Cytotoxic and Proinflammatory Activities. *Chem Biol* 21(11):1457-1462.
64. Schwartz K*, et al.* (2014) The AgrD N-Terminal Leader Peptide of Staphylococcus aureus Has Cytolytic and Amyloidogenic Properties. *Infect Immun* 82(9):3837-3844.
65. Lina G*, et al.* (1998) Transmembrane topology and histidine protein kinase activity of AgrC, the agr signal receptor in Staphylococcus aureus. *Mol Microbiol* 28(3):655-662.
66. Wang LN*, et al.* (2014) New insight into transmembrane topology of Staphylococcus aureus histidine kinase AgrC. *Bba-Biomembranes* 1838(3):988-993.
67. Thoendel M, Kavanaugh JS, Flack CE, & Horswill AR (2011) Peptide Signaling in the Staphylococci. *Chem Rev* 111(1):117-151.
68. Beining PR, Huff E, Prescott B, & Theodore TS (1975) Characterization of Lipids of Mesosomal Vesicles and Plasma-Membranes from Staphylococcus-Aureus. *J Bacteriol* 121(1):137-143.
69. Gao R & Stock AM (2009) Biological Insights from Structures of Two-Component Proteins. *Annu Rev Microbiol* 63:133-154.
70. Dutta R & Inouye M (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences* 25(1):24-28.
71. Cisar EA, Geisinger E, Muir TW, & Novick RP (2009) Symmetric signalling within asymmetric dimers of the Staphylococcus aureus receptor histidine kinase AgrC. *Mol Microbiol* 74(1):44-57.

72. Srivastava SK, Rajasree K, Fasim A, Arakere G, & Gopal B (2014) Influence of the AgrC-AgrA Complex on the Response Time of Staphylococcus aureus Quorum Sensing. *J Bacteriol* 196(15):2876-2888.
73. Lyon GJ, Wright JS, Christopoulos A, Novick RP, & Muir TW (2002) Reversible and specific extracellular antagonism of receptor-histidine kinase signaling. *J Biol Chem* 277(8):6247-6253.
74. Wright JS, Lyon GJ, George EA, Muir TW, & Novick RP (2004) Hydrophobic interactions drive ligand-receptor recognition for activation and inhibition of staphylococcal quorum sensing. *P Natl Acad Sci USA* 101(46):16168-16173.
75. Geisinger E, George EA, Muir TW, & Novick RP (2008) Identification of ligand specificity determinants in AgrC, the Staphylococcus aureus quorum-sensing receptor. *J Biol Chem* 283(14):8930-8938.
76. Malone CL, Boles BR, & Horswill AR (2007) Biosynthesis of Staphylococcus aureus the Synechocystis autoinducing peptides by using DnaB mini-intein. *Appl Environ Microb* 73(19):6036-6044.
77. Lyon GJ, Mayville P, Muir TW, & Novick RP (2000) Rational design of a global inhibitor of the virulence response in Staphylococcus aureus, based in part on localization of the site of inhibition to the receptor-histidine kinase, AgrC. *P Natl Acad Sci USA* 97(24):13330-13335.
78. McDowell P*, et al.* (2001) Structure, activity and evolution of the group I thiolactone peptide quorum-sensing system of Staphylococcus aureus. *Mol Microbiol* 41(2):503-512.
79. Johnson JG, Wang BY, Debelouchina GT, Novick RP, & Muir TW (2015) Increasing AIP Macrocycle Size Reveals Key Features of agr Activation in Staphylococcus aureus. *Chembiochem* 16(7):1093-1100.
80. George EA, Novick RP, & Muir TW (2008) Cyclic peptide inhibitors of staphylococcal virulence prepared by Fmoc-based thiolactone peptide synthesis. *J Am Chem Soc* 130(14):4914-4924.
81. Fowler SA, Stacy DM, & Blackwell HE (2008) Design and synthesis of macrocyclic peptomers as mimics of a quorum sensing signal from Staphylococcus aureus. *Org Lett* 10(12):2329-2332.
82. Tal-Gan Y, Stacy DM, & Blackwell HE (2014) N-Methyl and peptoid scans of an autoinducing peptide reveal new structural features required for inhibition and activation of AgrC quorum sensing receptors in Staphylococcus aureus. *Chem Commun* 50(23):3000-3003.
83. Tal-Gan Y, Stacy DM, Foegen MK, Koenig DW, & Blackwell HE (2013) Highly Potent Inhibitors of Quorum Sensing in Staphylococcus aureus Revealed Through a Systematic Synthetic Study of the Group-III Autoinducing Peptide. *J Am Chem Soc* 135(21):7869-7882.
84. Stock AM, Robinson VL, & Goudreau PN (2000) Two-component signal transduction. *Annu Rev Biochem* 69:183-215.
85. Nikolskaya AN & Galperin MY (2002) A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res* 30(11):2453-2459.

86. Sidote DJ, Barbieri CM, Wu T, & Stock AM (2008) Structure of the Staphylococcus aureus AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure* 16(5):727-735.
87. Nicod SS*, et al.* (2014) Systematic mutational analysis of the LytTR DNA binding domain of Staphylococcus aureus virulence gene transcription factor AgrA. *Nucleic Acids Res* 42(20):12523-12536.
88. Reynolds J & Wigneshweraraj S (2011) Molecular Insights into the Control of Transcription Initiation at the Staphylococcus aureus agr operon. *J Mol Biol* 412(5):862-881.
89. Koenig RL, Ray JL, Maleki SJ, Smeltzer MS, & Hurlburt BK (2004) Staphylococcus aureus AgrA binding to the RNAIII-agr regulatory region. *J Bacteriol* 186(22):7549-7555.
90. Warne SE & Dehaseth PL (1993) Promoter Recognition by Escherichia-Coli Rna-Polymerase - Effects of Single-Base Pair Deletions and Insertions in the Spacer DNA Separating the -10 and -35 Regions Are Dependent on Spacer DNA-Sequence. *Biochemistry-Us* 32(24):6134-6140.
91. Morfeldt E, Taylor D, Vongabain A, & Arvidson S (1995) Activation of Alpha-Toxin Translation in Staphylococcus-Aureus by the Trans-Encoded Antisense Rna, Rnaiii. *Embo J* 14(18):4569-4577.
92. Sun F*, et al.* (2012) Quorum-sensing agr mediates bacterial oxidation response via an intramolecular disulfide redox switch in the response regulator AgrA. *P Natl Acad Sci USA* 109(23):9095-9100.
93. Reyes D*, et al.* (2011) Coordinated Regulation by AgrA, SarA, and SarR To Control agr Expression in Staphylococcus aureus. *J Bacteriol* 193(21):6020-6031.
94. Cheung AL, Bayer AS, Zhang GY, Gresham H, & Xiong YQ (2004) Regulation of virulence determinants in vitro and in vivo in Staphylococcus aureus. *Fems Immunol Med Mic* 40(1):1-9.
95. Chien YT, Manna AC, & Cheung AL (1998) SarA level is a determinant of agr activation in Staphylococcus aureus. *Mol Microbiol* 30(5):991-1001.
96. Manna A & Cheung AL (2001) Characterization of sarR, a modulator of sar expression in Staphylococcus aureus. *Infect Immun* 69(2):885-896.
97. Bronner S, Monteil H, & Prevost G (2004) Regulation of virulence determinants in Staphylococcus aureus: complexity and applications. *Fems Microbiol Rev* 28(2):183-200.
98. Majerczyk CD*, et al.* (2008) Staphylococcus aureus CodY negatively regulates virulence gene expression. *J Bacteriol* 190(7):2257-2265.
99. Liu KQ, Bittner AN, & Wang JD (2015) Diversity in (p)ppGpp metabolism and effectors. *Curr Opin Microbiol* 24:72-79.
100. Roux A, Todd DA, Velazquez JV, Cech NB, & Sonenshein AL (2014) CodY-Mediated Regulation of the Staphylococcus aureus Agr System Integrates Nutritional and Population Density Signals. *J Bacteriol* 196(6):1184-1196.
101. Cegelski L, Marshall GR, Eldridge GR, & Hultgren SJ (2009) The biology and future prospects of antivirulence therapies (vol 6, pg 17, 2008). *Nat Rev Microbiol* 7(11):836-836.
102. Kong KF, Vuong C, & Otto M (2006) Staphylococcus quorum sensing in biofilm formation and infection. *Int J Med Microbiol* 296(2-3):133-139.

103. Otto M (2004) Quorum-sensing control in Staphylococci - a target for antimicrobial drug therapy? *Fems Microbiol Lett* 241(2):135-141.
104. Rothfork JM*, et al.* (2004) Inactivation of a bacterial virulence pheromone by phagocyte-derived oxidants: New role for the NADPH oxidase in host defense. *P Natl Acad Sci USA* 101(38):13867-13872.
105. Peterson MM*, et al.* (2008) Apolipoprotein B Is an Innate Barrier against Invasive Staphylococcus aureus Infection. *Cell Host Microbe* 4(6):555-566.
106. Hall PR*, et al.* (2013) Nox2 Modification of LDL Is Essential for Optimal Apolipoprotein B-mediated Control of agr Type III Staphylococcus aureus Quorum-sensing. *Plos Pathog* 9(2).
107. Mansson M*, et al.* (2011) Inhibition of Virulence Gene Expression in Staphylococcus aureus by Novel Depsipeptides from a Marine Photobacterium. *Mar Drugs* 9(12):2537-2552.
108. Igarashi Y*, et al.* (2015) Avellanin C, an inhibitor of quorum-sensing signaling in Staphylococcus aureus, from Hamigera ingelheimensis. *The Journal of antibiotics*.
109. Desouky SE*, et al.* (2015) Cyclodepsipeptides produced by actinomycetes inhibit cyclic-peptide-mediated quorum sensing in Gram-positive bacteria. *Fems Microbiol Lett* 362(14).
110. Gordon CP, Williams P, & Chan WC (2013) Attenuating Staphylococcus aureus Virulence Gene Regulation: A Medicinal Chemistry Perspective. *J Med Chem* 56(4):1389-1404.
111. Gilligan PH (1991) Microbiology of Airway Disease in Patients with Cystic-Fibrosis. *Clin Microbiol Rev* 4(1):35-51.
112. Murray EJ*, et al.* (2014) Targeting Staphylococcus aureus Quorum Sensing with Nonpeptidic Small Molecule Inhibitors. *J Med Chem* 57(6):2813-2819.
113. Leonard PG, Bezar IF, Sidote DJ, & Stock AM (2012) Identification of a Hydrophobic Cleft in the LytTR Domain of AgrA as a Locus for Small Molecule Interactions That Inhibit DNA Binding. *Biochemistry-Us* 51(50):10035-10043.
114. Sully EK*, et al.* (2014) Selective Chemical Inhibition of agr Quorum Sensing in Staphylococcus aureus Promotes Host Defense with Minimal Impact on Resistance. *Plos Pathog* 10(6).
115. Park J*, et al.* (2007) Infection control by antibody disruption of bacterial quorum sensing signaling. *Chem Biol* 14(10):1119-1127.
116. Nakayama J*, et al.* (2006) Revised model for Enterococcus faecalis fsr quorum-sensing system: the small open reading frame fsrD encodes the gelatinase biosynthesis-activating pheromone propeptide corresponding to staphylococcal agrd. *J Bacteriol* 188(23):8321-8326.
117. Garmyn D, Gal L, Lemaitre JP, Hartmann A, & Piveteau P (2009) Communication and autoinduction in the species Listeria monocytogenes: A central role for the agr system. *Communicative & integrative biology* 2(4):371-374.
118. Cooksley CM*, et al.* (2010) Regulation of neurotoxin production and sporulation by a Putative agrBD signaling system in proteolytic Clostridium botulinum. *Appl Environ Microbiol* 76(13):4448-4460.

119. Vidal JE*, et al.* (2012) Evidence that the Agr-like quorum sensing system regulates the toxin production, cytotoxicity and pathogenicity of Clostridium perfringens type C isolate CN3685. *Mol Microbiol* 83(1):179-194.
120. Annis DA, Nickbarg E, Yang X, Ziebell MR, & Whitehurst CE (2007) Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Current opinion in chemical biology* 11(5):518-526.
121. Casino P, Rubio V, & Marina A (2009) Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* 139(2):325-336.
122. Mechaly AE, Sassoon N, Betton JM, & Alzari PM (2014) Segmental Helical Motions and Dynamical Asymmetry Modulate Histidine Kinase Autophosphorylation. *Plos Biol* 12(1).
123. Marina A, Waldburger CD, & Hendrickson WA (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *Embo J* 24(24):4247-4259.
124. Bilwes AM, Quezada CM, Croal LR, Crane BR, & Simon MI (2001) Nucleotide binding by the histidine kinase CheA. *Nature Structural Biology* 8(4):353-360.
125. Grebe TW & Stock JB (1999) The histidine protein kinase superfamily. *Advances in Microbial Physiology, Vol 41* 41:139-227.
126. Dago AE*, et al.* (2012) Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *P Natl Acad Sci USA* 109(26):E1733-E1742.
127. Zhang Z*, et al.* (1998) Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res* 26(17):3986-3990.
128. Huang Y, Niu BF, Gao Y, Fu LM, & Li WZ (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26(5):680-682.
129. Papadopoulos JS & Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073-1079.
130. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14(9):755-763.
131. Krogh A, Larsson B, von Heijne G, & Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305(3):567-580.
132. Cheung J & Hendrickson WA (2010) Sensor domains of two-component regulatory systems. *Curr Opin Microbiol* 13(2):116-123.
133. Martin B*, et al.* (2010) Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of Streptococcus pneumoniae. *Mol Microbiol* 75(6):1513-1528.
134. Havarstein LS, Diep DB, & Nes IF (1995) A Family of Bacteriocin Abc Transporters Carry out Proteolytic Processing of Their Substrates Concomitant with Export. *Mol Microbiol* 16(2):229-240.
135. Lin DYW, Huang S, & Chen J (2015) Crystal structures of a polypeptide processing and secretion transporter. *Nature* 523(7561):425-U492.
136. Wuster A & Babu MM (2008) Conservation and evolutionary dynamics of the agr cell-to-cell communication system across firmicutes. *J Bacteriol* 190(2):743-746.

137. Kee JM, Oslund RC, Perlman DH, & Muir TW (2013) A pan-specific antibody for direct detection of protein histidine phosphorylation. *Nat Chem Biol* 9(7):416-U428.
138. Denisov IG, Grinkova YV, Lazarides AA, & Sligar SG (2004) Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *J Am Chem Soc* 126(11):3477-3487.
139. Ritchie TK*, et al.* (2009) Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in Enzymology; Liposomes, Pt F* 464:211-231.
140. Veesler D*, et al.* (2009) Production and biophysical characterization of the CorA transporter from Methanosarcina mazei. *Analytical Biochemistry* 388(1):115-121.
141. Kee JM & Muir TW (2012) Chasing Phosphohistidine, an Elusive Sibling in the Phosphoamino Acid Family. *Acs Chem Biol* 7(1):44-51.
142. Grimshaw CE*, et al.* (1998) Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in Bacillus subtilis. *Biochemistry-Us* 37(5):1365-1375.
143. Geisinger E, Muir TW, & Novick RP (2009) agr receptor mutants reveal distinct modes of inhibition by staphylococcal autoinducing peptides. *P Natl Acad Sci USA* 106(4):1216-1221.
144. Scaramozzino F, White A, Perego M, & Hoch JA (2009) A Unique GTP-Dependent Sporulation Sensor Histidine Kinase in Bacillus anthracis. *J Bacteriol* 191(3):687-692.
145. Kenney LJ (1997) Kinase activity of EnvZ, an osmoregulatory signal transducing protein of Escherichia coli. *Archives of Biochemistry and Biophysics* 346(2):303-311.
146. Tawa P & Stewart RC (1994) Kinetics of Chea Autophosphorylation and Dephosphorylation Reactions. *Biochemistry-Us* 33(25):7917-7924.
147. Thomas SA, Brewster JA, & Bourret RB (2008) Two variable active site residues modulate response regulator phosphoryl group stability. *Mol Microbiol* 69(2):453-465.
148. Tran QH & Unden G (1998) Changes in the proton potential and the cellular energetics of Escherichia coli during growth by aerobic and anaerobic respiration or by fermentation. *European Journal of Biochemistry* 251(1-2):538-543.
149. Gunstone FD, Harwood JL, & Padley FB (1994) *The Lipid handbook* (Chapman and Hall, London ; New York) 2nd Ed pp xiii, 722, 551 p.
150. Casino P, Rubio V, & Marina A (2010) The mechanism of signal transduction by two-component systems. *Current Opinion in Structural Biology* 20(6):763-771.
151. Albanesi D*, et al.* (2009) Structural plasticity and catalysis regulation of a thermosensor histidine kinase. *P Natl Acad Sci USA* 106(38):16185-16190.
152. Diensthuber RP, Bommer M, Gleichmann T, & Moglich A (2013) Full-Length Structure of a Sensor Histidine Kinase Pinpoints Coaxial Coiled Coils as Signal Transducers and Modulators. *Structure* 21(7):1127-1136.
153. Wang C*, et al.* (2013) Mechanistic Insights Revealed by the Crystal Structure of a Histidine Kinase with Signal Transducer and Sensor Domains. *Plos Biol* 11(2).
154. Mccaldon P & Argos P (1988) Oligopeptide Biases in Protein Sequences and Their Use in Predicting Protein Coding Regions in Nucleotide-Sequences. *Proteins-Structure Function and Genetics* 4(2):99-122.

155. Scholtz JM & Baldwin RL (1992) The Mechanism of Alpha-Helix Formation by Peptides. *Annu Rev Bioph Biom* 21:95-118.
156. Thompson KS, Vinson CR, & Freire E (1993) Thermodynamic Characterization of the Structural Stability of the Coiled-Coil Region of the Bzip Transcription Factor Gcn4. *Biochemistry-Us* 32(21):5491-5496.
157. Moglich A, Ayers RA, & Moffat K (2009) Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *J Mol Biol* 385(5):1433-1444.
158. Capra EJ & Laub MT (2012) Evolution of Two-Component Signal Transduction Systems. *Annual Review of Microbiology, Vol 66* 66:325-347.
159. Wang BY, Zhao AS, Novick RP, & Muir TW (2014) Activation and Inhibition of the Receptor Histidine Kinase AgrC Occurs through Opposite Helical Transduction Motions. *Mol Cell* 53(6):929-940.
160. Casino P, Miguel-Romero L, & Marina A (2014) Visualizing autophosphorylation in histidine kinases. *Nat Commun* 5.
161. Molnar KS*, et al.* (2014) Cys-Scanning Disulfide Cross linking and Bayesian Modeling Probe the Transmembrane Signaling Mechanism of the Histidine Kinase, PhoQ. *Structure* 22(9):1239-1251.
162. Matamouros S, Hager KR, & Miller SI (2015) HAMP Domain Rotation and Tilting Movements Associated with Signal Transduction in the PhoQ Sensor Kinase. *Mbio* 6(3).
163. Alm E, Huang K, & Arkin A (2006) The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *Plos Comput Biol* 2(11):1329-1342.
164. Stiffler MA, Hekstra DR, & Ranganathan R (2015) Evolvability as a Function of Purifying Selection in TEM-1 beta-Lactamase. *Cell* 160(5).
165. Moore JO & Hendrickson WA (2012) An Asymmetry-to-Symmetry Switch in Signal Transmission by the Histidine Kinase Receptor for TMAO. *Structure* 20(4):729-741.
166. Cheung J & Hendrickson WA (2009) Structural Analysis of Ligand Stimulation of the Histidine Kinase NarX. *Structure* 17(2):190-201.
167. Martin RB (1998) Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers* 45(5):351-353.
168. Lehninger AL, Nelson DL, & Cox MM (2005) *Lehninger principles of biochemistry* (W.H. Freeman, New York) 4th Ed.
169. Grimsley GR, Scholtz JM, & Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* 18(1):247-251.
170. Guthrie JP (1978) Hydrolysis of Esters of Oxy Acids - Pka Values for Strong Acids - Bronsted Relationship for Attack of Water at Methyl - Free-Energies of Hydrolysis of Esters of Oxy Acids - and a Linear Relationship between Free-Energy of Hydrolysis and Pka Holding over a Range of 20 Pk Units. *Can J Chem* 56(17):2342-2354.
171. Illuminati G & Mandolini L (1981) Ring-Closure Reactions of Bifunctional Chain Molecules. *Accounts Chem Res* 14(4):95-102.
172. MacKenzie KR (2006) Folding and stability of alpha-helical integral membrane proteins. *Chem Rev* 106(5):1931-1977.

173. Boldog T, Grimme S, Li MS, Sligar SG, & Hazelbauer GL (2006) Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties. *P Natl Acad Sci USA* 103(31):11509-11514.
174. Kirchdoerfer RN*, et al.* (2011) Structural Basis for Ligand Recognition and Discrimination of a Quorum-quenching Antibody. *J Biol Chem* 286(19):17351-17358.
175. Arnison PG*, et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(1):108-160.
176. Scheffner M, Nuber U, & Huibregtse JM (1995) Protein Ubiquitination Involving an E1-E2-E3 Enzyme Ubiquitin Thioester Cascade. *Nature* 373(6509):81-83.
177. Buczek O, Krowarsch D, & Otlewski J (2002) Thermodynamics of single peptide bond cleavage in bovine pancreatic trypsin inhibitor (BPTI). *Protein Sci* 11(4):924-932.
178. Junio HA*, et al.* (2013) Quantitative analysis of autoinducing peptide I (AIP-I) from Staphylococcus aureus cultures using ultrahigh performance liquid chromatography-high resolving power mass spectrometry. *J Chromatogr B* 930:7-12.
179. Maass S*, et al.* (2011) Efficient, Global-Scale Quantification of Absolute Protein Amounts by Integration of Targeted Mass Spectrometry and Two-Dimensional Gel-Based Proteomics. *Anal Chem* 83(7):2677-2684.
180. Bottcher T & Sieber SA (2008) beta-Lactones as Specific Inhibitors of ClpP Attenuate the Production of Extracellular Virulence Factors of Staphylococcus aureus. *J Am Chem Soc* 130(44):14400-14401.
181. Frees D, Qazi SNA, Hill PJ, & Ingmer H (2003) Alternative roles of ClpX and ClpP in Staphylococcus aureus stress tolerance and virulence. *Mol Microbiol* 48(6):1565-1578.
182. Brändén C-I & Tooze J (1999) *Introduction to protein structure* (Garland Pub., New York) 2nd Ed pp xiv, 410 p.
183. Joo C, Balci H, Ishitsuka Y, Buranachai C, & Ha T (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu Rev Biochem* 77:51-76.
184. Clemente JC, Ursell LK, Parfrey LW, & Knight R (2012) The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* 148(6):1258-1270.
185. Fischbach MA & Sonnenburg JL (2011) Eating For Two: How Metabolism Establishes lnterspecies Interactions in the Gut. *Cell Host Microbe* 10(4):336-347.
186. Grote A*, et al.* (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33:W526-W531.
187. Charpentier E*, et al.* (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl Environ Microb* 70(10):6076-6085.
188. Vergis JM & Wiener MC (2011) The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expres Purif* 78(2):139-142.
189. Swillens S (1995) Interpretation of Binding Curves Obtained with High Receptor Concentrations - Practical Aid for Computer-Analysis. *Mol Pharmacol* 47(6):1197-1203.

190. Scholtz JM, Qian H, York EJ, Stewart JM, & Baldwin RL (1991) Parameters of Helix-Coil Transition Theory for Alanine-Based Peptides of Varying Chain Lengths in Water. *Biopolymers* 31(13):1463-1470.
191. Ferrige AG, Seddon MJ, & Jarvis S (1991) Maximum-Entropy Deconvolution in Electrospray Mass-Spectrometry. *Rapid Commun Mass Sp* 5(8):374-377.
192. Ram G*, et al.* (2012) Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *P Natl Acad Sci USA* 109(40):16300-16305.
193. Davis JJ, Xia F, Overbeek RA, & Olsen GJ (2013) Genomes of the class Erysipelotrichia clarify the firmicute origin of the class Mollicutes. *Int J Syst Evol Microbiol* 63(Pt 7):2727-2741.