

Accepted Manuscript

Structural Mechanisms of Peptide Recognition and Allosteric Modulation of Gene Regulation by the RRNPP Family of Quorum-Sensing Regulators

Hackwon Do, Muthiah Kumaraswami

PII: S0022-2836(16)30205-4
DOI: doi: [10.1016/j.jmb.2016.05.026](https://doi.org/10.1016/j.jmb.2016.05.026)
Reference: YJMBI 65112

To appear in: *Journal of Molecular Biology*

Received date: 18 February 2016
Revised date: 19 May 2016
Accepted date: 29 May 2016



Please cite this article as: Do, H. & Kumaraswami, M., Structural Mechanisms of Peptide Recognition and Allosteric Modulation of Gene Regulation by the RRNPP Family of Quorum-Sensing Regulators, *Journal of Molecular Biology* (2016), doi: [10.1016/j.jmb.2016.05.026](https://doi.org/10.1016/j.jmb.2016.05.026)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Structural Mechanisms of Peptide Recognition and Allosteric Modulation of Gene Regulation by the RRNPP-Family of Quorum-Sensing Regulators*

Hackwon Do¹ and Muthiah Kumaraswami^{1,2}

¹From the Center for Molecular and Translational Human Infectious Diseases Research, Houston Methodist Hospital Research Institute, and Department of Pathology and Genomic Medicine, Houston Methodist Hospital System, Houston, TX USA

*Running title: Structural insights into RRNPP-family regulators

²To whom correspondence should be addressed: Muthiah Kumaraswami, Department of Pathology and Genomic Medicine, Houston Methodist Hospital System, Houston, TX 77030, USA

Tel.: (713)-441-5252; Fax: (713)-441-3447; Email: mkumaraswami@houstonmethodist.org

Keywords: RRNPP family; Peptide signaling; Allostery; Gene regulation; Structural mechanisms

Abstract

The members of RRNPP family of bacterial regulators sense population density-specific secreted oligopeptides and modulate the expression of genes involved in cellular processes, such as sporulation, competence, virulence, biofilm formation, conjugative plasmid transfer and antibiotic resistance. Signaling by RRNPP regulators include several steps: generation and secretion of the signaling oligopeptides, re-internalization of the signaling molecules into the cytoplasm, signal sensing by the cytosolic RRNPP regulators, signal-specific allosteric structural changes in the regulators, and interaction of the regulators with their respective regulatory target and gene regulation. The recently determined structures of the RRNPP regulators provide insight into the mechanistic aspects for several steps in this signaling circuit. In this review, we discuss the structural principles underlying peptide specificity, regulatory target recognition, and ligand-induced allostery in RRNPP regulators and its impact on gene regulation. Despite the conserved tertiary structure of these regulators, structural analyses revealed unexpected diversity in the mechanism of activation and molecular strategies that couple the peptide-induced allostery to gene regulation. Although these structural studies provide a sophisticated understanding of gene regulation by RRNPP regulators, much needs to be learned regarding the target DNA binding by yet-to-be characterized RNPP regulators as well as several aspects of signaling by Rgg regulators.

Introduction

Bacteria communicate with each other using secreted signaling molecules and coordinate population-wide behavior in concert with cell density^{1; 2}. Bacterial gene regulation by intercellular signaling pathways occurs in the host during infection and is critical for bacterial survival and virulence *in vivo*^{1; 3; 4; 5; 6; 7}. Gram-positive bacteria predominantly use bacterially

encoded secreted oligopeptides to monitor population density in the immediate environment and mediate gene regulation^{2; 8}. The secreted oligopeptides are sensed either extracellularly by the membrane-bound histidine kinases of two-component signaling systems or intracellularly by the cytosolic transcription regulators^{2; 8; 9; 10}. The latter group of transcription regulators belongs to a newly characterized family known as RRNPP family of cytosolic peptide-sensing regulators^{9; 11}. The **R**ap phosphatases from *Bacillus subtilis*, **R**gg from *Streptococcus* species, **N**prR from *B. cereus* group, **P**lcR from *B. cereus* group, and **P**rgX from *Enterococcus faecalis* comprise the founding members of the RRNPP family of regulators^{9; 11}. The members of the RRNPP family regulators control an array of important cellular processes, which include sporulation, competence, virulence, biofilm formation, necrotrophic lifestyle, conjugative plasmid transfer and antibiotic resistance (Table 1)^{3; 12; 13; 14; 15}.

Rap proteins, originally identified as response regulator aspartyl phosphatases, regulate the two-component signaling systems by blocking the activity of response regulators^{16; 17; 18}. Rap proteins control diverse cellular processes including sporulation, genetic competence, biofilm formation, secreted enzyme production, and conjugative transposon mobility^{18; 19; 20}. In *B. subtilis*, a total of 11 Rap proteins are encoded in the genome and an additional 5 are present in plasmids^{21; 22}. Eight of the 11 Rap proteins in the chromosome form cassettes with the cognate inhibitory peptides, Phr, while others either cross-react with non-cognate Phr peptides or remain as orphan Rap proteins^{23; 24; 25; 26; 27}. RapA, RapB, RapE, RapH, and RapJ proteins function by dephosphorylation of the sporulation factor, Spo0F, whereas RapC, RapD, RapF, RapG, RapH, and RapK act as anti-activators by blocking response regulator-DNA interactions^{16; 17}.

Rgg regulators are prevalent in firmicutes, and peptide-mediated Rgg signaling pathways have been identified in several streptococci^{12; 28; 29; 30; 31}. The genome of *S. pyogenes* encodes

four *rgg* paralogs, each having regulatory roles in different cellular processes. The regulator of protease B (RopB) is a transcription activator of virulence genes including the major virulence factor SpeB, and gene regulation by RopB is critical for bacterial pathogenesis³². RopB shares significant structural homology with other RRNPP regulators and exhibits cell density-dependent transcription regulation, but the identities of the regulatory peptide signals remain to be elucidated^{4; 15}. Rgg2 and Rgg3 control biofilm formation and modulate expression of the target genes in an opposing fashion²⁹. Both Rgg2 and Rgg3 have their cognate peptide signals, shp2 and shp3, respectively, encoded in their immediate genetic vicinities, and expression of *rgg* and *shp* genes is autoregulated. Apo Rgg3 acts as a repressor by directly interacting with promoter sequences, and shp3 binding to Rgg3 induces its dissociation from DNA and causes derepression of target genes^{29; 30}. On the other hand, Rgg2 binds to DNA in the presence of shp2, and the peptide-bound Rgg2 activates target gene expression³⁰. ComR is a Rgg regulator that controls the cryptic competence pathway in *S. pyogenes* in the presence of its peptide ligand, SigX-inducing peptide (XIP) Rgg¹².

Both NprR and PlcR have been extensively studied in the *B. cereus* group and they control the expression of genes required for necrotrophic and virulence lifestyles of the bacteria, respectively³. NprR is active only in the sporulation-specific growth medium and NprR-mediated gene regulation is required for bacterial survival in necrotic host tissue^{33; 34}. It activates expression of ~41 genes encoding various tissue degradation enzymes, including metalloproteases, chitinases, lipases, and a lipopeptide³⁴. The gene encoding the activation peptide NprX is located immediately downstream of NprR, and transcription of NprR-NprX is modulated by several transcription factors including PlcR¹⁴. The mature NprX binds directly to NprR and promotes its DNA binding by inducing NprR tetramerization³⁵. Although NprR-

binding elements in the target promoters have been identified, the binding motif has not been heretofore elucidated. Based on the amino acid sequences of NprR and NprX, 7 different strain-specific NprR-NprX pairs were identified and NprX-dependent activation of NprR is highly specific among the 7 identified groups³. The transcription regulator of phosphatidylinositol-specific phospholipase C, PlcR, controls expression of ~45 genes during early stationary phase of growth, and gene regulation by PlcR is critical for bacterial virulence^{36; 37; 38}. Genes controlled by PlcR include several secreted proteases, lipases, hemolysins, and toxins that contribute to bacterial pathogenesis^{37; 39}. A gene encoding small, secreted peptide, *papR*, is located immediately downstream of *plcR* gene, and is required for the regulatory activity of PlcR⁴⁰. PapR binds to PlcR and promotes interaction between PlcR and the operator sequences, the *plcR* box, in the target promoters⁴⁰. Similarly to NprR, four strain-specific PlcR-PapR groups have been identified and PlcR from each group is activated by group-specific peptides^{41; 42}.

The sex pheromone receptor, PrgX, of *E. faecalis*, participates in transcription regulation of genes involved in conjugative transfer of the antibiotic resistance plasmid pCF10⁴³. Conjugative exchange of genetic material occurs between two cell types, a donor cell having the plasmid, and a recipient cell lacking it⁴³. PrgX is encoded in the pCF10 plasmid, and its activity is modulated by two antagonistic peptides, the inhibition and activation peptides, iCF10 and cCF10, respectively. The gene encoding iCF10 is located in the plasmid pCF10^{44; 45}, whereas the cCF10 is located in the secretion signal sequence of lipoprotein ccfA in the genome^{46; 47}. Although both donor and recipient cells produce cCF10, expression of iCF10 in the donor cells neutralizes endogenous cCF10 and prevents self-induction by inhibiting PrgX-mediated upregulation of conjugative transfer genes⁴⁵. However, when the population density of recipient

cells increases, cCF10 at higher concentrations outcompetes iCF10 for PrgX binding and promotes PrgX-dependent transcription activation of target genes ⁴⁵.

Despite the differences in their downstream regulatory targets, the underlying mechanism of signaling is conserved among the members. Typically, the signaling process includes production and secretion of the oligopeptide signals, extracellular maturation of the peptide signals from precursor form to mature form, internalization of the peptide signals into the cytoplasm through oligopeptide permeases, recognition of the cognate signaling peptides by the respective RRNPP regulator in the cytoplasm, and modulation of the regulatory properties of the corresponding RRNPP regulator by peptide-induced allosteric structural changes ^{3; 27; 40; 48; 49} (Supplementary Fig. S1). With the exception of PrgX from *E. faecalis* and RopB from *S. pyogenes*, genes encoding the signaling peptide and receptor pairs form a cassette, and in some cases, the transcription of the peptide signals is controlled by the corresponding regulator ^{15; 46}. The activation peptide for PrgX, cCF10, is not encoded in the immediate vicinity of PrgX coding region; instead it is located as part of the secretion signal sequence of a lipoprotein ⁴⁶. However, PrgX forms a cassette with its inhibition peptide signal, iCF10. The genetic identity of cognate peptide signals for RopB has not been elucidated yet ^{4; 15}.

Further, peptide sensing by RRNPP family regulators results in varying downstream consequences. The regulators that belong to the Rgg-NprR-PlcR-PrgX (RNPP) subgroup interact with their cognate peptide signals and peptide binding modulate the interactions between the regulator and their respective operator sequences in the target promoter and influence gene regulation ^{14; 29; 35; 40; 50; 51}. On the other hand, Rap proteins mediate their gene regulation by their interactions with transcription regulators. Based on the mechanism of gene regulation, Rap proteins can be further categorized into two subgroups. The interactions between the first group

of Rap proteins and their regulatory target sequester the transcription activator and influence gene expression by sterically blocking the regulator from accessing its target promoter^{25; 27; 52; 53; 54}. The second group of Rap proteins possesses phosphatase activity and affects gene regulation by dephosphorylating the response regulator^{27; 52; 55}. Regardless of the mechanism, in both cases binding of the regulatory peptide allosterically inhibits the interactions between Rap proteins and regulatory targets, which results in modulation of the respective response regulator function^{25; 27; 52; 56}. For the purpose of clarity, when discussing the specific aspects pertaining to the DNA-binding subgroup of RRNPP members, we used the term RNPP (Rgg-NprR-PlcR-PrgX) family in the sections below. Recently, the crystal structures of each founding member of the RRNPP family in different conformational states were determined. In light of the recent structural studies of the RRNPP family regulators, this review highlights the structural principles and mechanisms that contribute to peptide discrimination, regulatory target recognition, and peptide-induced allosteric changes that contribute to the modulation of gene regulation by the RRNPP family of regulators.

Overall architecture of the RRNPP regulators:

The members of the RRNPP regulators share a two-domain structure with an N-terminal regulatory target-binding domain, a C-terminal domain that has the structural elements for regulatory peptide binding and oligomerization, and a short linker helix connecting the two domains (Fig. 1 and Supplementary Fig. S2). With the exception of Rap proteins, members of the RRNPP family have the DNA-binding domain with the classical helix-turn-helix (HTH) motif at their N-terminus (Fig. 1C, F and G)^{11; 35; 51; 57; 58; 59}. However, the 3-helix bundle at the N-terminal domain of Rap proteins directly interacts with the transcription regulators and

modulates gene regulation by blocking the regulator-DNA interactions and/or by influencing the phosphorylation state of the regulator (Fig. 1A and B)^{26; 27; 49}. The C-terminal domain of RRNPP regulators is characterized by the presence of a tetratricopeptide repeat (TPR)-motif, a hallmark structural motif involved in protein-peptide or protein-protein interactions in bacterial and eukaryotic proteins^{60; 61}. Each TPR motif is ~ 34 amino acids long and constitutes a pair of antiparallel helices⁶¹. The individual TPR motifs are arranged in tandem to form a right-handed superhelical structure, which forms the TPR domain. Most of the RRNPP regulators have a total of 5 TPR motifs in the TPR domain, whereas Rap protein and NprR have 7 and 9 TPR motifs, respectively. Nevertheless, the TPR domain has a convex exterior and a concave internal surface that contains the binding pocket for the regulatory oligopeptides (Fig. 1 and Supplementary Fig. S2). The DNA-binding RNPP regulators are dimeric proteins and the dimerization interactions span through the entire molecule with predominant contributions arising from the α -helices from the C-terminal domain. Additional intersubunit interactions between the linker helices and the DNA-binding domains of opposing subunits further stabilize the dimer (Fig. 1C-G). In this arrangement, the individual domains of RNPP regulators are swapped in a way that positions the N- and C-terminal domains of the same subunit on the opposite sides of a dimer molecule (Fig. 1C-G). Contrary to the RNPP regulators, the structurally-characterized Rap proteins exist as monomers in solution and remain functional in their monomeric state (Fig. 1A and B)^{27; 49}. In PlcR and PrgX, a small fragment C-terminal to the last TPR motif, known as the “cap subdomain”, constitutes the allosteric switch. Structural changes occurring in this subdomain upon cognate peptide binding in the TPR domain trigger the conformational changes in PlcR and PrgX that transition the regulators from their inactive to active functional state^{11; 57; 58}.

Peptide recognition by RRNPP regulators:

The signaling oligopeptides for the RRNPP family regulators are 5-8 amino acids long and are predominantly composed of hydrophobic amino acids with one or two polar or charged residues interspersed among the apolar residues^{9; 13; 14; 40; 46} (Fig. 2). Despite the similarities in the amino acid composition and length of the peptide signals, the regulators exhibit a high degree of specificity for their cognate peptide signals with minimal to no cross reactivity between the subfamilies of regulators. In this regard, the recently determined crystal structures of RRNPP regulators in their peptide-bound states reveal the structural basis for peptide recognition and specificity exhibited by the members of this family^{11; 27; 35; 49; 51; 57; 58}. In each case, the signaling peptide binds to the concave surface of the TPR domain and the peptide is oriented with its C-terminus facing the interior of the pocket, whereas the N-terminus of the peptide is located at the entry point (Fig. 1 and 2).

Several determinants dictate peptide selectivity of the regulators that include anchoring contacts, hydrophobic contacts conferring shape complementarity between the two molecules, stabilizing peptide backbone interactions, and side chain-specific H-bonding interactions. Invariably, the carboxylate oxygen of the C-terminus of the peptide signals interacts with the polar and charged residues located at the deep end of the binding pocket; these interactions play an anchoring role for proper positioning of the peptides within the binding pocket⁶². A second layer of interactions occurs between the aromatic or hydrophobic side chains of the peptide signals and the apolar pockets of the regulators (Fig. 2; top panels). These interactions confer shape complementarity between the two molecules and favor a snugged fit of the cognate signaling peptides in their respective binding pocket (Fig. 2; top panels). Additional stabilizing contacts occur between the peptide backbone and the polar or charged residues in the TPR domain (Fig. 2; bottom panels). All the RRNPP regulators contain conserved surface-exposed

asparagine residues within the concave surface of the TPR domain and the side chains of the asparagines are involved in peptide backbone contacts in all the structurally characterized RRNPP regulators (Fig. 2; bottom panels)^{15; 35; 49; 51; 57; 59}. Interestingly, fewer side chain-specific contacts exist between the RRNPP regulators and their cognate signaling peptides. Sequence-specific H-bonding interactions occur between the side chains of the polar or charged residues of the signaling peptides and the polar and charged residues in the binding pocket (Fig.2; bottom panels).

Two structures of different Rap proteins bound to their cognate inhibitory peptides, namely RapF bound to PhrF, and RapJ bound to PhrC, were recently determined^{26; 49}. In both cases, the peptides make contacts with similar surfaces within the TPR domain of the respective Rap regulators (Fig. 1A and B). Unlike other RRNPP regulators, the binding pocket of the Rap proteins forms a closed tunnel with both faces of the channel open²⁶. Peptide binding induces narrowing of the binding pocket, which is reduced from 8588Å³ in RapF to 4954 Å³ in PhrF-bound RapF. The mode of peptide recognition and discrimination by Rap proteins follow the general principles of peptide binding for the RRNPP family of regulators. The oxygen moieties of the carboxyl terminus in both structures are anchored inside the pocket (Fig. 2A). However, the coordinating residues from the respective Rap proteins vary; Y66 of the N-terminal domain, R223 of TPR4, and Q183 of TPR3 coordinate the carboxylate oxygen atoms in RapF, whereas Q181 of TPR3 that is analogous to Q183 in RapF, and Y217 of TPR4 are engaged in similar interactions in RapJ. Extensive interactions between the peptide backbones and the non-polar residues lining the peptide binding pockets of both RapF and RapJ stabilize the protein-peptide interactions (Fig. 2A). With one exception, the side chains of analogous residues in both structures are involved in these interactions. Both PhrF and PhrC are 5 amino acids long and they

differ in their amino acid composition at positions 1 and 5. As expected, the recognition of the dissimilar residues by their respective Rap proteins confers specificity to the interactions between the cognate pairs of Phr peptides and Rap proteins. In particular, the PhrF peptide has a glutamine at position 1, whereas PhrC contains a glutamate at the same position (Fig. 2A, bottom panel and Supplementary Fig. S3). Consistent with their electrostatic potential, the amine group of glutamine at position 1 of PhrF is coordinated by the negatively charged side chain of E303 of RapF, whereas the carboxyl group of glutamate at position 1 of PhrC is engaged in electrostatic interactions with the positively charged side chain of K300 of RapJ (Fig. 2A, bottom panel and Supplementary Fig. S3). Consistent with this structural observation, replacement of E303 of RapF with the positively charged lysine caused a change in specificity as the E303K mutant of RapF binds to non-cognate PhrC peptide and displays PhrC-dependent regulatory activity. Similarly, the chemical nature of the amino acid at position 5 vary between the PhrF and PhrC peptides as PhrF has a hydrophobic isoleucine and PhrC contains a polar threonine at position 5. As a result, the side chain of threonine in PhrC is engaged in H-bonding interactions with E147, whereas isoleucine in PhrF is involved in hydrophobic interactions with L187 and T190 (Fig. 2A, bottom panel and Supplementary Fig. S3). However, RapF has a glutamate (E149) at the position analogous to E147 of RapJ and this might participate in the H-bonding interactions between E303K mutant of RapF and T5 of PhrC, thus making RapF responsive to PhrC.

Crystal structure of NprR bound to 8-mer form of NprX provided insights into peptide discrimination by NprR³⁵. The side chain of R126 from TPR1 anchors the carboxyl oxygen of the C-terminus of NprX in the deeper end of the binding pocket (Fig. 2B). Synthetic NprX peptide variants containing substitutions of the carboxyl oxygen with amine group drastically affected the binding affinity of NprR-NprX interactions. Similarly, alanine substitutions at R126

impaired the peptide binding and regulatory activity of NprR, emphasizing the critical role of these anchoring interactions. Several peptide backbones spanning the entire length of NprX are coordinated by the surface-exposed asparagines, N275, N306, and D309, and stabilize NprX in the binding pocket (Fig. 2B, bottom panel). Specificity is conferred by the H-bonding interactions between S2 of NprX and S274 of NprR as well as D5 of NprX and R126 of NprR (Fig. 2B, bottom panel).

PlcR was crystallized with two active forms of PapR, namely the 5-mer and 7-mer, and peptide recognition by PlcR in both structures is identical^{11; 51}. However, additional protein-peptide contacts exist between the two extra residues in PapR7 and PlcR, which explains the slightly increased affinity for PapR7 as compared to PapR5. Anchoring contacts between the carboxyl oxygen of the peptide and Lys197 orient the peptide within the TPR domain, whereas van der Waals interactions between L3, F5, and F7 of PapR and the hydrophobic pocket of PlcR aligns PapR along the binding surface (Fig. 2C). The side chains of asparagines in the peptide-binding pocket of PlcR, namely N159, N163, and N201, make stabilizing peptide backbone contacts along the longitudinal axis of PapR. The sequence-specific H-bonding interactions occur with the side chain of charged residue, E6, of PapR and the side chains of Y275 of capping helix, K89 of the linker helix, and Q237 of helix 12 (Fig. 2C, bottom panel).

In the case of PrgX, both the inhibition and activation peptides bind to a similar concave surface of the TPR domain but they differ in their contacts with the cap subdomain. In both cases, the carboxyl oxygen is coordinated by the side chain of Lys79 from TPR1 motif located in the interior of the binding pocket (Fig. 2D and Supplementary Fig. S3). The regulatory peptides for PrgX contain single polar residue, a threonine at position 3, and the H-bonding interactions between the Thr3 of the peptide signals and PrgX dictate the specific interactions (Fig. 2D,

bottom panel). Despite the conservation of threonine at the third position in both oligopeptides, different amino acid side chains of PrgX interact with the analogous threonine from inhibition and activation peptides. In the crystal structure of PrgX bound to inhibition peptide iCF10, Thr3 is involved in H-bonding interactions with Glu312 and K314 of PrgX (Fig. 2D, bottom panel and Supplementary Fig. S3), whereas the analogous residue in the activation peptide is engaged in interactions with the side chains of Thr296, Y298, and Tyr302 of PrgX. Interestingly, such differences in the interactions between PrgX and its regulatory peptides constitute the allosteric switch for the peptide-specific PrgX regulatory activity (discussed below).

Although the crystal structures for the peptide-bound Rgg regulators have yet to be determined, the structure determination of a small molecule inhibitor bound Rgg structure from *S. dysgalactiae* and the unliganded structure of C-terminal domain of RopB from *S. pyogenes* indicated that Rgg regulators likely conform to the principles of peptide recognition encountered in the structurally characterized members of RRNPP family regulators^{15; 59}. Both structures contain a TPR domain that has the characteristic concave peptide-binding pocket and the putative peptide-binding surface has surface-exposed conserved asparagines that are involved in protein-peptide interactions^{15; 59}.

Structural basis for the interactions between the regulatory targets and RRNPP regulators

Members of the RRNPP family regulators employ their N-terminal domains (NTD) to bind the respective regulatory targets and mediate gene regulation. The crystal structures of the Rap proteins from both groups bound to their regulatory targets, namely anti-activator RapF bound to ComA and aspartyl phosphatase RapH complexed with Spo0F, provided mechanistic insights into the biological function. The 3-helix bundle of RapF NTD directly binds to the

transcription activator ComA and negatively influences transcription regulation by inhibiting ComA-DNA interactions as well as ComA dimerization²⁵ (Fig. 1A). The ComA binding pocket in RapF is formed by the C-terminal one third of helices $\alpha 2$ and $\alpha 3$, the intervening loop between helices $\alpha 3$ and $\alpha 4$, and the N-terminal end of linker helix $\alpha 4$. The ComA binding surface of RapF mimics the DNA major groove both in shape and electrostatic potential²⁵. As a result, the recognition helix of the ComA DNA-binding domain that is rich in basic residues interacts with the acidic patch in RapF binding pocket; these electrostatic interactions bury 6 out of 7 DNA-contacting residues of ComA at the RapF-ComA interface²⁵ (Supplementary Movie M1). Additionally, RapF influences the DNA-binding function of ComA by interfering with its dimerization allosterically as dimer is the active oligomeric form of ComA²⁵. However, the precise mechanism by which RapF controls ComA dimerization remains unknown. Although aspartyl phosphatase RapF inhibits the sporulation factor Spo0F by direct interaction through its N-terminal domain, structural studies of RapF-Spo0F revealed significant mechanistic differences between the two subgroups of Rap proteins. First, the Spo0F binding site is located on the opposite face of the N-terminal domain as compared to the ComA binding surface and involves helix $\alpha 2$, the N-terminal end of helix $\alpha 3$, and the loop connecting helices $\alpha 4$ and $\alpha 5$ (Fig. 1A). Second, in addition to the N-terminal domain, RapH also utilizes the TPR domain to interact with Spo0F, which includes the loop between TPR1 and TPR2 motifs²⁷ (Fig. 1A). Collectively, RapH-Spo0F interactions inhibit the phosphorelay signaling by Spo0F by two mechanisms: 1) dephosphorylation of Spo0F, and 2) partial occlusion of the binding surface for the interaction partners of Spo0F and inhibition of phosphotransfer from the upstream as well as to the downstream regulators²⁷ (Supplementary Movie M2).

The crystal structure of PlcR-DNA complex is the only DNA-bound structure available for the DNA-binding subgroup of RRNPP family regulators⁵¹. In this structure, the 18-base pair long DNA-binding motif for PlcR, *plcR box*, contains an imperfect palindrome with only the bases at positions 2,3,4,5, and 7 of the 9 bp repeat are highly conserved. As observed in HTH-containing bacterial transcription regulators, PlcR binds to DNA as a dimer and inserts the recognition helix $\alpha 3$ of the HTH motif of each subunit into two adjacent major grooves on the same face of DNA (Fig. 1F). PlcR employs the side chains of Q31, S32, and R36 of helix $\alpha 3$ to make base-specific contacts with the conserved bases at positions 2,3,4,5, and 7⁵¹. Additional non-specific contacts between the phosphate backbone of DNA and HTH motifs of PlcR further stabilize the PlcR-DNA complex⁵¹. Mechanistic understanding of the mode of DNA binding and mechanism of promoter recognition by other RRNPP regulators will require the structure determination of protein-DNA complexes of PrgX, NprR, and Rgg proteins bound to their respective operator sequences.

Peptide-induced allosteric changes in RRNPP regulators

Although the RRNPP regulators share common modes of signaling and peptide sensing, each of the structurally characterized members displays a unique mechanism of peptide-dependent allostery and gene regulation (Fig. 2). Specifically, the regulators vary in the fundamental allosteric changes occurring in their structures that lead to altered gene regulatory properties. Three-dimensional structures for both groups, anti-activator RapF in three different conformational states, and Rap phosphatases, RapH, RapJ, and RapI, in multiple functional states were determined^{25; 26; 27; 49}. In both cases, peptide and target proteins bind to distinct sites in the Rap proteins and the cognate Phr peptides control the regulatory activity of Rap proteins

allosterically. Specifically, peptide occupancy at the TPR domain induces drastic structural changes in the N-terminal 3-helix bundle, which results in burial of the target protein-binding surface of Rap proteins (Fig. 3A)(Supplementary Movie M1 and M2). Such burial of the binding surface prevents the Rap proteins from interacting with its regulatory targets, thereby inhibiting its regulatory activity (Fig. 3A)(Supplementary Movie M1 and M2). Comparison of the apo- and ComA-bound RapF structures indicated the lack of global structural changes in RapF with only local differences between the two states noted in the N-terminal 3-helix bundle^{25; 26}. Importantly, in both cases, the 3-helix bundle is detached from the C-terminal TPR domain and exists as a distinct, exposed domain, with the binding surface remaining accessible for interaction with ComA (Fig. 3A)(Supplementary Movie M1 and M2). However, upon PhrF binding, conformational changes originating from the TPR domain results in drastic relocation of the N-terminal domain. Interestingly, the C-terminal half of TPR domain that has the TPR4-TPR7 motifs does not undergo significant structural perturbations upon PhrF binding, while major structural changes occur in the N-terminal half of TPR domain²⁶)(Supplementary Movie M1 and M2). Structural changes propagate from the TPR3 to the TPR1 motif with more pronounced transitions occurring in the TPR1 motif. As a result, the linker helix along with the 3-helix bundle of RapF complexed to PhrF undergoes a 155° rotation movement relative to the apo-RapF structure, which results in the upward swing movement of the NTD towards the TPR domain²⁶ (Fig. 3A)(Supplementary Movie M1 and M2). Further, the unfolding of the last two turns of the helix α 3 along with the linker region forms a long unstructured loop, and this region is buried in the interface between the N- and C-terminal domains. Since the helix α 3, the loop connecting α 2 and α 3, and the linker helix form the ComA binding surface, burial of this region results in

partial occlusion of the ComA binding pocket and the consequent blockade of RapF-ComA interactions²⁵.

Biochemical analysis of NprR-NprX interactions revealed that the transition from a dimer to tetramer constitute the allosteric switch in NprR³⁵ (Fig. 3B). NprR exists as an inactive dimer in its peptide-free state and NprX binding promotes tetramerization, which represents the transcriptionally active state of NprR³⁵. Subsequent structural analysis of the C-terminal domain of NprR bound to NprX identified two oligomerization interfaces, a dimer interface that is formed by the interactions between the TPR9 motifs of the opposing subunits and a tetramer interface that is generated by the contacts between the loops connecting the TPR4 and TPR5 motifs of subunits that are not involved in dimerization³⁵. Both interfaces are highly conserved among different NprR-NprX pairs in *B. cereus* group and the integrity of the two interfaces is critical for the regulatory function of NprR. Although a speculative model for the NprX-dependent activation of NprR has been proposed, structure determination of the full-length NprR in different functional states namely, apo, peptide-bound, and DNA-bound, are required to fully understand the NprX-induced allosteric changes in NprR that modulate its DNA binding and regulatory activity.

The crystal structures of PlcR in 3 different conformational states, namely the unliganded, peptide-bound, and peptide-DNA-bound, have been determined^{11; 51}. Since DNA binding by PlcR depends on its interactions with PapR, the structures of PlcR bound to PapR and PapR-DNA provide invaluable insights into the structural changes occurring in PlcR that leads to altered gene regulation. Conformational changes triggered by PapR binding in PlcR involve subtle changes in the C-terminal domain of the molecule that transduces through the dimer interface to the N-terminal domain (Supplementary Movie M3). Specifically, the PlcR-peptide

interactions cause the movement of capping helix that propagates to the dimer interface, which is further transduced to the N-terminal linker helix⁵¹ (Fig. 3C)(Supplementary Movie M3). In the apo PlcR, key intersubunit interactions occur between the side chains of Y64 and I68 with their respective counterparts of the opposing subunit. Such stacking interactions lock the N-terminal DNA-binding domain in an inactive conformation that is incompatible for DNA binding. PapR-induced structural changes occurring at the capping helix moves the linker helices of the dimer farther apart from each other, which results in the disruption of Y64-Y64' (where ' - indicates the amino acid from the second subunit of a dimer) interactions. The breakage of Y64-Y64' stacking interactions unlocks the DNA-binding domain and confers conformational flexibility⁵¹ (Fig. 3C). Additional structural changes occur in PlcR in the presence of DNA, which include the breakage of linker helix into two helices as the C-terminal half interacts with the helix 6 to form classical TPR motif 1, whereas the N-terminal half of helix 5 constitutes the linker helix (Supplementary Movie M3). The structural findings were further corroborated by elegant biochemical experiments, which demonstrated that unlocking PlcR by introducing substitutions either at Y64 or I68 rendered it active independent of peptide binding at the TPR domain. As a result, the DNA-binding domain is pulled upwards toward the TPR domain of the opposing subunit and assumes a conformation that is compatible for DNA binding (Fig. 3C)(Supplementary Movie M3). In this conformation, the two recognition helices are only 32 Å apart, which is suitable for its interactions with two adjacent major grooves in a B-DNA (~ 34Å)⁵¹.

Although the peptide-induced structural changes in PrgX involve the cap subdomain, the downstream consequences of these structural changes are quite different. Unlike PlcR, PrgX binds DNA in its apo, inhibitory peptide iCF10-bound and activation peptide cCF10-bound states. Consistent with this, the distances between the recognition helices of HTH motifs of PrgX

dimer in its three conformational states are similar (~ 29 Å), suggesting that it exists in DNA-binding competent conformation in all functional states. On the other hand, the signaling peptide-specific oligomeric state of PrgX determines the regulatory outcome. The dimeric activation peptide-bound PrgX positively influences target gene transcription whereas the tetrameric inhibition peptide-bound PrgX represses gene expression (Fig. 3D). The tetramer interface is formed by end-to-end interactions between two PrgX dimers with their C-terminal tail engaged in intersubunit interactions^{57; 58} (Fig. 3D). The C-terminal tail involved in tetramerization is composed of helix $\alpha 17$ (residues 293-306), and the preceding loop between $\alpha 16$ and $\alpha 17$ (residues 287-292) (Fig. 3D)(Supplementary Movie M4). The side chain-specific interactions between T3 of cCF10 and PrgX involve the residues T296, Y298, and Y302 from $\alpha 17$, whereas similar interactions between T3 of iCF10 and PrgX include residues E312 and K314 from the fragment C-terminal to the tetramer interface^{57; 58}. As a result, the burial or exposure of the C-terminal tail in response to the activation peptide or inhibition peptide occupancy at the TPR domain constitutes the allosteric switch of PrgX^{57; 58} (Fig. 3D)(Supplementary Movie M4). Specifically, the activation peptide binding induces the unfolding of $\alpha 17$ and places the C-terminal tail on top of the activation peptide. Such interactions between the C-terminal half of the cap subdomain of PrgX and activation peptide pull the C-terminal tail away from the tetramer interface and prevent it from participating in tetramerization interactions (Fig. 3D)(Supplementary movie M4). Conversely, the interactions between the inhibition peptide and PrgX result in exposure of the C-terminal tail and promote tetramerization^{57; 58} (Fig. 3D)(Supplementary movie M4).

Conclusions

Recent years have brought about significant progress in the structural understanding of the various mechanistic steps of intercellular signaling in Gram-positive bacteria, namely, signal sensing, signal-induced allostery, and gene regulation. These notable advances notwithstanding, critical aspects of these mechanisms still remain elusive. Several speculative models for DNA binding and promoter recognition by RNPP regulators have been proposed, which include DNA looping by the inhibitor peptide-bound PrgX, and recognition of non-adjacent major grooves by NprR. With the exception of PlcR, the crystal structures for DNA-bound RNPP regulators have not been determined and the mode of promoter recognition by RNPP regulators has yet to be elucidated. Further, the structures of liganded state of the largest group of RRNPP regulators, Rgg, have not been determined; the lack of this structural data hampers the understanding of the mechanistic details of peptide sensing, and allostery in this subgroup of regulators. Finally, structure-function studies also indicated that the peptide-binding surface and the dimer interface of RNPP regulators are critical for their regulatory function^{15; 59}.

Acknowledgments

This work was supported by the National Institute of Health grants (1R01AI109096-01A1 to M.K). The authors declare no conflicts of interest.

References

1. Rutherford, S. T. & Bassler, B. L. (2012). Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. *Cold Spring Harb. Perspect. Med.* **2**.
2. Dunny, G. M. & Leonard, B., A.B. (1997). Cell-Cell communication in Gram-positive bacteria. *Annu. Rev. Microbiol.* **51**, 527-564
3. Slamti, L., Perchat, S., Huillet, E. & Lereclus, D. (2014). Quorum Sensing in *Bacillus thuringiensis* Is Required for Completion of a Full Infectious Cycle in the Insect. *Toxins* **6**, 2239.

4. Shelburne III, S. A., Olsen, R. J., Makthal, N., Brown, N. G., Sahasrabhojane, P., Watkins, E. M., Palzkill, T., Musser, J. M. & Kumaraswami, M. (2011). An amino-terminal signal peptide of Vfr protein negatively influences RopB-dependent SpeB expression and attenuates virulence in *Streptococcus pyogenes*. *Mol. Microbiol.* **82**, 1481-1495.
5. Middleton, B., Rodgers, H. C., Cámara, M., Knox, A. J., Williams, P. & Hardman, A. (2002). Direct detection of N-acylhomoserine lactones in cystic fibrosis sputum. *FEMS Microbiol. Lett.* **207**, 1-7.
6. Marks, L. R., Mashburn-Warren, L., Federle, M. J. & Hakansson, A. P. (2014). *Streptococcus pyogenes* Biofilm Growth *In Vitro* and *In Vivo* and Its Role in Colonization, Virulence, and Genetic Exchange. *J. Infect. Dis.* **210**, 25-34.
7. Antunes, L. C. M., Ferreira, R. B. R., Buckner, M. M. C. & Finlay, B. B. (2010). Quorum sensing in bacterial virulence. *Microbiology* **156**, 2271-2282.
8. Thoendel, M., Kavanaugh, J. S., Flack, C. E. & Horswill, A. R. (2011). Peptide Signaling in the Staphylococci. *Chem. Rev.* **111**, 117-151.
9. Rocha-Estrada, J., Aceves-Diez, A. E., Guarneros, G. & de la Torre, M. (2010). The RNPP family of quorum-sensing proteins in Gram-positive bacteria. *Appl. Microbiol. & Biotech.* 1-11.
10. Novick, R. P. & Geisinger, E. (2008). Quorum Sensing in Staphylococci. *Annu. Rev. Genet.* **42**, 541-564.
11. Declerck, N., Bouillaut, L., Chaix, D., Rugani, N., Slamti, L., Hoh, F., Lereclus, D. & Arold, S. T. (2007). Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. *Proc. Natl. Acad. Sci. USA* **104**, 18490-18495.
12. Mashburn-Warren, L., Morrison, D. A. & Federle, M. J. (2010). A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol. Microbiol.* **78**, 589-606.
13. Perego, M. & Brannigan, J. A. (2001). Pentapeptide regulation of aspartyl-phosphate phosphatases. *Peptides* **22**, 1541-1547.
14. Perchat, S., Dubois, T., Zouhir, S., Gominet, M., Poncet, S., Lemy, C., Aumont-Nicaise, M., Deutscher, J., Gohar, M., Nessler, S. & Lereclus, D. (2011). A cell-cell communication system regulates protease production during sporulation in bacteria of the *Bacillus cereus* group. *Mol. Microbiol.* **82**, 619-633.
15. Makthal, N., Gavagan, M., Do, H., Olsen, R. J., Musser, J. M. & Kumaraswami, M. (2015). Structural and functional analysis of RopB - a major virulence regulator in *Streptococcus pyogenes*. *Mol. Microbiol.* **epub**.
16. Smits, W. K., Bongiorno, C., Veening, J.-W., Hamoen, L. W., Kuipers, O. P. & Perego, M. (2007). Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. *Mol. Microbiol.* **65**, 103-120.
17. Bongiorno, C., Ishikawa, S., Stephenson, S., Ogasawara, N. & Perego, M. (2005). Synergistic Regulation of Competence Development in *Bacillus subtilis* by Two Rap-Phr Systems. *J. Bacteriol.* **187**, 4353-4361.
18. Perego, M., Glaser, P. & Hoch, J. A. (1996). Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. *Mol. Microbiol.* **19**, 1151-1157.

19. Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P. & Grossman, A. D. (2005). Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. USA* **102**, 12554-12559.
20. López, D. & Kolter, R. (2010). Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. *FEMS. Microbiol. Rev.* **34**, 134-149.
21. McLoon, A. L., Guttenplan, S. B., Kearns, D. B., Kolter, R. & Losick, R. (2011). Tracing the Domestication of a Biofilm-Forming Bacterium. *J. Bacteriol.* **193**, 2027-2034.
22. Meijer, W. J. J., Wisman, G. B. A., Terpstra, P., Thorsted, P. B., Thomas, C. M., Holsappel, S., Venema, G. & Bron, S. (1998). Rolling-circle plasmids from *Bacillus subtilis*: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from Gram-positive bacteria. *FEMS Microbiol. Rev.* **21**, 337-368.
23. Pottathil, M. & Lazazzera, B. (2003). The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Front. Biosci.* **8**, 32-45.
24. Perego, M. (1997). A peptide export-import control circuit modulating bacterial development regulates protein phosphatases of the phosphorelay. *Proc. Natl. Acad. Sci. USA* **94**, 8612-8617.
25. Baker, M. D. & Neiditch, M. B. (2011). Structural Basis of Response Regulator Inhibition by a Bacterial Anti-Activator Protein. *PLoS Biol.* **9**, e1001226.
26. Gallego del Sol, F. & Marina, A. (2013). Structural Basis of Rap Phosphatase Inhibition by Phr Peptides. *PLoS Biol.* **11**, e1001511.
27. Parashar, V., Mirouze, N., Dubnau, D. A. & Neiditch, M. B. (2011). Structural Basis of Response Regulator Dephosphorylation by Rap Phosphatases. *PLoS Biol.* **9**, e1000589.
28. Fleuchot, B., Gitton, C., Guillot, A., Vidic, J., Nicolas, P., Besset, C., Fontaine, L., Hols, P., Leblond-Bourget, N., Monnet, V. & Gardan, R. (2011). Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Mol. Microbiol.* **80**, 1102-1119.
29. Chang, J. C., LaSarre, B., Jimenez, J. C., Aggarwal, C. & Federle, M. J. (2011). Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. *PLoS Pathog.* **7**, e1002190.
30. LaSarre, B., Aggarwal, C. & Federle, M. J. (2012). Antagonistic Rgg Regulators Mediate Quorum Sensing via Competitive DNA Binding in *Streptococcus pyogenes*. *mBio* **3**.
31. Cook, L. C., LaSarre, B. & Federle, M. J. (2013). Interspecies Communication among Commensal and Pathogenic Streptococci. *mBio* **4**.
32. Carroll, R. K., Shelburne III, S. A., Olsen, R. J., Suber, B., Sahasrabhojane, P., Kumaraswami, M., Beres, S. B., Shea, P. R., Flores, A. R. & Musser, J. M. (2011). Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J. Clin. Invest.* **121**, 1956-1968.
33. Dubois, T., Perchat, S., Verplaetse, E., Gominet, M., Lemy, C., Aumont-Nicaise, M., Grenha, R., Nessler, S. & Lereclus, D. (2013). Activity of the *Bacillus thuringiensis* NprR–NprX cell–cell communication system is co-ordinated to the physiological stage through a complex transcriptional regulation. *Mol. Microbiol.* **88**, 48-63.
34. Dubois, T., Faegri, K., Perchat, S., Lemy, C., Buisson, C., Nielsen-LeRoux, C., Gohar, M., Jacques, P., Ramarao, N., Kolstø, A.-B. & Lereclus, D. (2012). Necrotrophism Is a Quorum-Sensing-Regulated Lifestyle in *Bacillus thuringiensis*. *PLoS Pathog.* **8**, e1002629.

35. Zouhir, S., Perchat, S., Nicaise, M., Perez, J., Guimaraes, B., Lereclus, D. & Nessler, S. (2013). Peptide-binding dependent conformational changes regulate the transcriptional activity of the quorum-sensor NprR. *Nucl. Acids Res.* **41**, 7920-7933.
36. Lereclus, D., Agaisse, H., Gominet, M., Salamiou, S. & Sanchis, V. (1996). Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase. *J. Bacteriol.* **178**, 2749-56.
37. Agaisse, H., Gominet, M., Økstad, O. A., Kolstø, A.-B. & Lereclus, D. (1999). PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol. Microbiol.* **32**, 1043-1053.
38. Gohar, M., Faegri, K., Perchat, S., Ravnum, S., Økstad, O. A., Gominet, M., Kolstø, A.-B. & Lereclus, D. (2008). The PlcR Virulence Regulon of *Bacillus cereus*. *PLoS ONE* **3**, e2793.
39. Callegan, M. C., Kane, S. T., Cochran, D. C., Gilmore, M. S., Gominet, M. & Lereclus, D. (2003). Relationship of *plcR*-Regulated Factors to *Bacillus* Endophthalmitis Virulence. *Infect. Immun.* **71**, 3116-3124.
40. Slamti, L. & Lereclus, D. (2002). A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. *EMBO J.* **21**, 4550-4559.
41. Slamti, L. & Lereclus, D. (2005). Specificity and Polymorphism of the PlcR-PapR Quorum-Sensing System in the *Bacillus cereus* Group. *J. Bacteriol.* **187**, 1182-1187.
42. Bouillaut, L., Perchat, S., Arold, S., Zorrilla, S., Slamti, L., Henry, C., Gohar, M., Declerck, N. & Lereclus, D. (2008). Molecular basis for group-specific activation of the virulence regulator PlcR by PapR heptapeptides. *Nucl. Acids Res.* **36**, 3791-3801.
43. Chandler, J. R. & Dunny, G. M. (2004). Enterococcal peptide sex pheromones: synthesis and control of biological activity. *Peptides* **25**, 1377-1388.
44. Nakayama, J., Ruhfel, R. E., Dunny, G. M., Isogai, A. & Suzuki, A. (1994). The *prgQ* gene of the *Enterococcus faecalis* tetracycline resistance plasmid pCF10 encodes a peptide inhibitor, iCF10. *J. Bacteriol.* **176**, 7405-7408.
45. Chandler, J. R., Flynn, A. R., Bryan, E. M. & Dunny, G. M. (2005). Specific Control of Endogenous cCF10 Pheromone by a Conserved Domain of the pCF10-Encoded Regulatory Protein PrgY in *Enterococcus faecalis*. *J. Bacteriol.* **187**, 4830-4843.
46. Antiporta, M. H. & Dunny, G. M. (2002). *ccfA*, the Genetic Determinant for the cCF10 Peptide Pheromone in *Enterococcus faecalis* OG1RF. *J. Bacteriol.* **184**, 1155-1162.
47. Chandler, J. R. & Dunny, G. M. (2008). Characterization of the Sequence Specificity Determinants Required for Processing and Control of Sex Pheromone by the Intramembrane Protease Eep and the Plasmid-Encoded Protein PrgY. *J. Bacteriol.* **190**, 1172-1183.
48. Lazazzera, B. A. (2001). The intracellular function of extracellular signaling peptides. *Peptides* **22**, 1519-1527.
49. Parashar, V., Jeffrey, P. D. & Neiditch, M. B. (2013). Conformational Change-Induced Repeat Domain Expansion Regulates Rap Phosphatase Quorum-Sensing Signal Receptors. *PLoS Biol.* **11**, e1001512.
50. Caserta, E., Haemig, H. A. H., Manias, D. A., Tomsic, J., Grundy, F. J., Henkin, T. M. & Dunny, G. M. (2012). *In Vivo* and *In Vitro* Analyses of Regulation of the Pheromone-Responsive *prgQ* Promoter by the PrgX Pheromone Receptor Protein. *J. Bacteriol.* **194**, 3386-3394.

51. Grenha, R., Slamti, L., Nicaise, M., Refes, Y., Lereclus, D. & Nessler, S. (2013). Structural basis for the activation mechanism of the PlcR virulence regulator by the quorum-sensing signal peptide PapR. *Proc. Natl. Acad. Sci. USA* **110**, 1047-1052.
52. Diaz, A. R., Core, L. J., Jiang, M., Morelli, M., Chiang, C. H., Szurmant, H. & Perego, M. (2012). *Bacillus subtilis* RapA Phosphatase Domain Interaction with Its Substrate, Phosphorylated Spo0F, and Its Inhibitor, the PhrA Peptide. *J. Bacteriol.* **194**, 1378-1388.
53. Core, L. & Perego, M. (2003). TPR-mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in *Bacillus subtilis*. *Mol. Microbiol.* **49**, 1509-1522.
54. Ogura, M., Shimane, K., Asai, K., Ogasawara, N. & Tanaka, T. (2003). Binding of response regulator DegU to the *aprE* promoter is inhibited by RapG, which is counteracted by extracellular PhrG in *Bacillus subtilis*. *Mol. Microbiol.* **49**, 1685-1697.
55. Perego, M. & Hoch, J. A. (1996). Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **93**, 1549-1553.
56. Ishikawa, S., Core, L. & Perego, M. (2002). Biochemical Characterization of Aspartyl Phosphate Phosphatase Interaction with a Phosphorylated Response Regulator and Its Inhibition by a Pentapeptide. *J. Biol. Chem.* **277**, 20483-20489.
57. Shi, K., Brown, C. K., Gu, Z.-Y., Kozlowicz, B. K., Dunny, G. M., Ohlendorf, D. H. & Earhart, C. A. (2005). Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in *Enterococcus faecalis*. *Proc. Natl. Acad. Sci. USA* **102**, 18596-18601.
58. Kozlowicz, B. K., Shi, K., Gu, Z.-Y., Ohlendorf, D. H., Earhart, C. A. & Dunny, G. M. (2006). Molecular basis for control of conjugation by bacterial pheromone and inhibitor peptides. *Mol. Microbiol.* **62**, 958-969.
59. Parashar, V., Aggarwal, C., Federle, M. J. & Neiditch, M. B. (2015). Rgg protein structure, function and inhibition by cyclic peptide compounds. *Proc. Natl. Acad. Sci. USA* **112**, 5177-5182.
60. Cervený, L., Strásková, A., Danková, V., Hartlová, A., Cecková, M., Staud, F. & Stulík, J. (2013). Tetratricopeptide Repeat Motifs in the World of Bacterial Pathogens: Role in Virulence Mechanisms. *Infect. Immun.* **81**, 629-635.
61. D'Andrea, L. D. & Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* **28**, 655-662.
62. Core, L. J., Ishikawa, S. & Perego, M. (2001). A free terminal carboxylate group is required for PhrA pentapeptide inhibition of RapA phosphatase. *Peptides* **22**, 1549-1553.
63. DeLano, W. L. (2002). The PyMol Molecular Graphics System. *DeLano Scientific, Palo Alto, CA, USA*.

Figure Legends

Figure 1. Overall architecture of the structures of the RRNPP-family regulators. Surface representation of the crystal structures of RapF bound to ComA (pdb code:3ULQ)(A), RapH

bound to SpoOF (pdb code:3Q15)(B), Rgg (pdb code:4YV6)(C), RopB-CTD (pdb code:5DL2)(D), NprR bound to signal peptide NprX (pdb code:4GPK)(E), PlcR bound to signal peptide PapR and DNA (pdb code:3U3W)(F), and PrgX bound to activation signal peptide cCF10 (pdb code:2AXZ)(G). In panels C-G, One subunit of the dimeric RRNPP regulators is depicted in surface representation and the second one is shown as ribbons. The C-terminal domain (CTD), the N-terminal domain (NTD) and the linker helix are colored in brown, green, and pink, respectively. The second subunit of dimer molecules in panels C-G is shown as ribbons and colored in grey. The peptide molecules in the binding pocket are shown as spheres. In panels A and B, the regulatory targets of Rap proteins, ComA and SpoOF, are shown in surface representation and colored blue. In panel F, the DNA duplex is shown as a ladder. The concave surfaces that have the peptide-binding pockets in the CTDs are boxed. For the purpose of clarity, the structural elements on top of the peptide-binding pockets of RapF, RapH, NprR, PlcR, and PrgX are removed to reveal the similarities between the binding surfaces of different RRNPP family members. In supplementary figure S2, the structural elements of the entire molecules are shown. All structure-related figures were generated using Pymol ⁶³.

Figure 2. Peptide recognition by RRNPP regulators. Structural determinants governing specific interactions between RapF-PhrF (A), NprR-NprX (B), PlcR-PapR (C), and PrgX-cCF10 (D) are shown. Top panels: Surface representation of molecules depicting shape complementarity between protein-peptide interactions; the signaling peptides are depicted as both surface and stick representation in yellow. The surfaces of the interacting residues in the respective binding pocket of the regulators are colored in accordance with their electrostatic potential. Bottom panels: Protein-peptide interactions depicted in stick representation. The

peptide molecule is shown as thick sticks and colored in pink. The interacting residues in the respective protein molecule is shown as thin sticks and colored in gray. Residues in the peptide are numbered and labeled in red, whereas the amino acids in the protein are labeled in black. The carbon, nitrogen, oxygen, and sulfur atoms in each residue are colored in black, blue, red, and yellow, respectively. The amino acid sequences of the corresponding mature signaling peptides are shown at the bottom of each panel. The anchoring interactions between the carboxyl terminus of each peptide and the respective regulators are shown as green sticks. The sticks representing the side chain specific H-bonding interactions are shaded in brown. The stabilizing contacts between the peptide backbone and the amino acid side chains from the peptide-binding pocket are depicted as light green sticks. All structure-related figures were generated using Pymol ⁶³.

Figure 3. Schematic representation of peptide-induced allosteric changes in RRNPP family regulators. **A)** In the absence of Phr peptides, Rap proteins bind to their respective regulatory targets, Spo0F or ComA, and mediate gene regulation by dephosphorylation and/or steric occlusion of regulatory target-DNA interactions. Phr occupancy at the C-terminal domain (CTD) induces an upward swinging displacement of the NTD towards CTD and buries the protein-protein interaction surface. **B)** NprR exists as a dimer in its unliganded state and binding of NprX at CTD promotes NprR tetramerization by unknown mechanisms and modulates gene regulation. **C)** PapR binding to the CTD causes structural changes in the linker helix and NTD, which results in upward rotation of NTD towards CTD and shortening of the distance between the recognition helices of the DNA-binding domain of a dimer. As a result, PlcR binds to operator sequences in the target promoter and activates gene expression. **D)** Gene regulation by PrgX is controlled by two antagonistic peptides, inhibitory peptide iCF10 and activation peptide cCF10.

The iCF10 binding to PrgX causes the exposure of the C-terminal tail, which promotes the tail-to-tail arrangement of two PrgX dimers and tetramerization. The PrgX tetramer occupies two distantly located operator sequences simultaneously, and represses gene expression by DNA looping. Conversely, PrgX-cCF10 interactions bury the C-terminal tail in the peptide-binding pocket and prevent tetramerization. As a result, the PrgX dimer binds to single operator sequences and activates transcription.

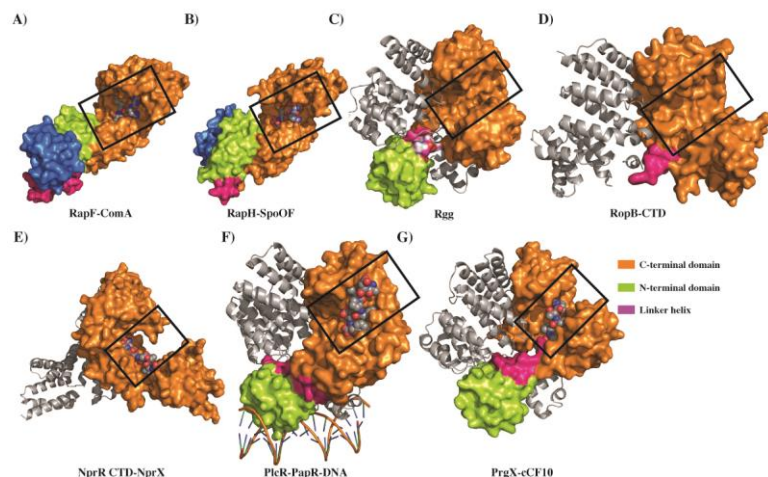


Figure 1. Overall architecture of the structures of the RRNPP-family regulators. Surface representation of the crystal structures of RapF bound to ComA (pdb code: 3ULQ) (A), RapH bound to SpoOF (pdb code: 3Q15) (B), Rgg (pdb code: 4YV6) (C), RopB-CTD (pdb code: 5DL2) (D), NprR bound to signal peptide NprX (pdb code: 4GPK) (E), PlcR bound to signal peptide PapR and DNA (pdb code: 3U3W) (F), PrgX bound to activation signal peptide cCF10 (pdb code: 2ANZ) (G). In panels C-G, one subunit of the dimeric RRNPP regulators is depicted in surface representation and the second one is shown as ribbons. The C-terminal domain (CTD), the N-terminal domain (NTD), and the linker helix are colored in brown, green, and pink, respectively. The second subunit of dimer molecule in panels C-G is shown as ribbons and colored in grey. The peptide molecules in the binding pocket are shown as spheres. In panels A and B, the regulatory targets of Rap proteins, ComA and SpoOF, are shown in surface representation and colored blue. In panel F, the DNA duplex is shown as a ladder. The concave surfaces that have the peptide-binding pockets in the CTDs are boxed. For the purpose of clarity, the structural elements on top of the peptide-binding pockets of RapF, RapH, NprR, PlcR, and PrgX are removed to reveal the similarities between the binding surfaces of different RRNPP family members. In supplementary figure S2, the structural elements of the entire molecules are shown. All structure-related figures were generated using PyMol ⁴⁰.

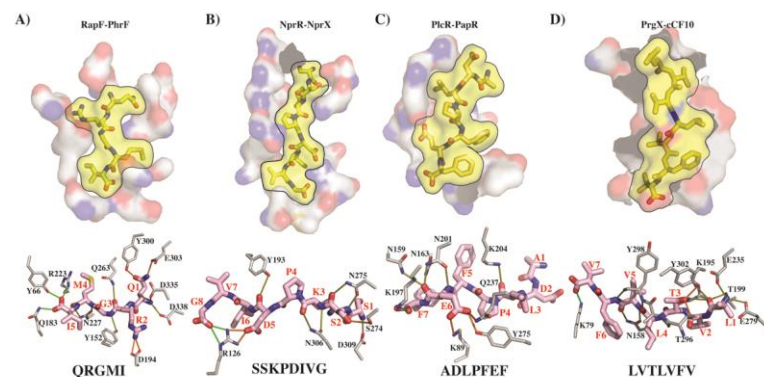


Figure 2. Peptide recognition by RRNTP regulators. Structural determinants governing specific interactions between RapF-PhrF (A), NprR-NprX (B), PicR-PapR (C), and PrgX-cCF10 (D) are shown. Top panels: Surface representation of molecules depicting shape complementarity between protein-peptide interactions; the signaling peptides are depicted as both surface and ball and stick representation in yellow. The surfaces of the interacting residues in the respective binding pocket of the regulators are colored in accordance with their electrostatic potential. Bottom panels: Protein-peptide interactions depicted in ball and stick representation. The peptide molecule is shown as thick ball and sticks and colored in pink. The interacting residues in the respective protein molecule are shown as thin ball and sticks and colored in gray. Residues in the peptide are numbered and labeled in red, whereas the amino acids in the protein are labeled in black. The carbon, nitrogen, oxygen, and sulfur atoms in each residue are colored in black, blue, red, and yellow, respectively. The amino acid sequences of the corresponding mature signaling peptides are shown at the bottom of each panel. The anchoring interactions between the carboxyl terminus of each peptide and the respective regulators are shown as green sticks. The sticks representing the side chain specific H-bonding interactions are shaded in brown. The stabilizing contacts between the peptide backbone and the amino acid side chains from the peptide-binding pocket are depicted as light green sticks. All structure-related figures were generated using PyMol[®].

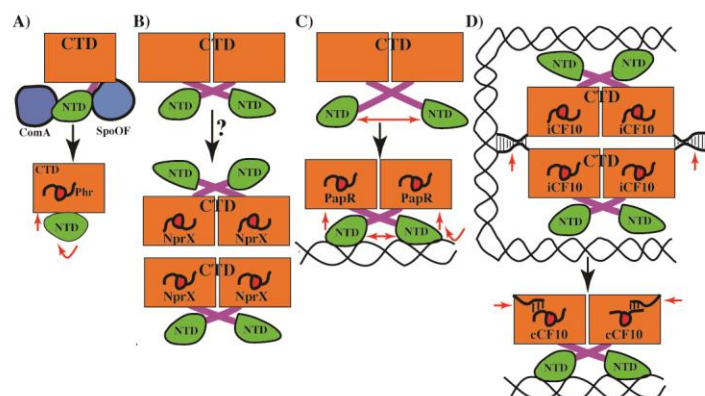


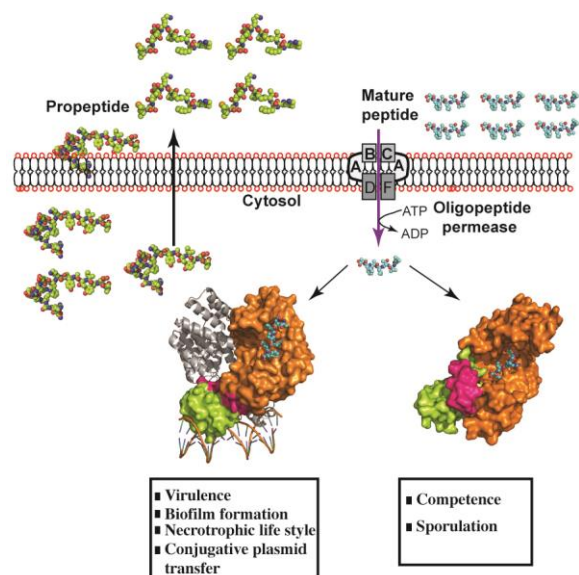
Figure 3. Schematic representation of peptide-induced allosteric changes in RRNPP family regulators. A) In the absence of Phr peptides, Rap proteins bind to their respective regulatory targets, SpoOF and ComA, and mediate gene regulation by dephosphorylation and/or steric occlusion of regulatory target-DNA interactions. Phr occupancy at the C-terminal domain (CTD) induces an upward swinging displacement of the NTD towards CTD and buries the protein-protein interaction surface. B) NprX exists as a dimer in its unliganded state and binding of NprX at CTD promotes NprX tetramerization by unknown mechanism and modulate gene regulation. C) PapR binding to the CTD causes structural changes in the linker helix and NTD, which results in upward rotation of NTD towards CTD and shortening of the distance between the recognition helices of the DNA-binding domain of a dimer. As a result, PicR binds to operator sequences in the target promoter and activates gene expression. D) Gene regulation by PrgX is controlled by two antagonistic peptides, inhibitory peptide iCF10 and activation peptide cCF10. The iCF10 binding to PrgX causes the exposure of the C-terminal tail, which promotes the tail-to-tail arrangement of two PrgX dimers and tetramerization. The PrgX tetramer occupies two distantly located operator sequences simultaneously, and represses gene expression by DNA looping. Conversely, PrgX-cCF10 interactions bury the C-terminal tail in the peptide-binding pocket and prevent tetramerization. As a result, the PrgX dimer binds to the single operator sequences and activates transcription.

Table 1. Summary of structurally characterized members of RRNPP family regulators

	Structure s available (<i>pdb ID</i>)	Organisms	Regulated function	Amino acid sequence of mature peptide	Mechanism of allostery
<u>Rap - Antiactivators</u> RapF	RapF (4I9E) RapF-PhrF (4I9C) RapF-ComA (3ULQ)	<i>Bacillus subtilis</i>	Competence induction	QRGMI	EXPOSURE /BURIAL OF REGULATORY TARGET BINDING
<u>Rap - Phosphatases</u> Rap	RapI (4I1A) RapJ-PhrC (4GYO) RapH-Spo0F (3Q15)	<i>Bacillus subtilis</i>	Sporulation initiation	ERGMT	EXPOSURE /BURIAL OF REGULATORY TARGET BINDING
<u>Rgg</u> Rgg2	Rgg2 (4YV6)	<i>Streptococcus dysgalactiae</i>	Biofilm formation	DILIIVGG	NOT KNOWN
RopB	RopB-CTD (5DL2)	<i>Streptococcus pyogenes</i>	Virulence factor production	NOT KNOWN	NOT KNOWN
<u>NprR</u>	NprR CTD- NprX8 (4GPK)	<i>Bacillus cereus group</i>	Necrotrophic lifestyle	SSKPDIVG	DIMERIZATION/ TETRAMERIZATI ON
<u>PrgX</u>	PrgX (2AXU) PrgX-iCF10 (2GRM) PrgX-cCF10 (2AXZ)	<i>Enterococcus faecalis</i>	Conjugative plasmid transfer and antibiotic resistance	iCF10 - AITLIFI cCF10 - LVTLVFV	DIMERIZATION/ TETRAMERIZATI ON

<u>PlcR</u>	PlcR (4FSC) PlcR-PapR5 (2QFC) PlcR-PapR7- DNA (3U3W)	<i>Bacillus cereus</i> <i>group</i>	Virulence factor production	ADLPFEF	DNA BINDING
--------------------	--	--	-----------------------------------	----------------	-------------

ACCEPTED MANUSCRIPT



Graphical abstract

Highlights

- The RRNPP family of bacterial regulators sense population density-specific secreted oligopeptides and modulate diverse cellular processes.
- Overall tertiary structure among the members of this family is conserved.
- However, they vary in the mechanism of activation and peptide-induced allostery.
- This review summarizes the structural advances in the understanding of peptide recognition and allosteric mechanisms in RRNPP family regulators.