Enterococcal Sex Pheromones: Evolutionary Pathways to Complex, Two-Signal Systems

Gary M. Dunny,a Ronnie Per-Arne Berntssonb
Department of Microbiology and Immunology, University of Minnesota, Minneapolis, Minnesota, USAa; Department of Medical Biochemistry and Biophysics, Umeå University, Umeå, Swedenb

Gram-positive bacteria carry out intercellular communication using secreted peptides. Important examples of this type of communication are the enterococcal sex pheromone systems, in which the transfer of conjugative plasmids is controlled by intercellular signaling among populations of donors and recipients. This review focuses on the pheromone response system of the conjugative plasmid pCF10. The peptide pheromones regulating pCF10 transfer act by modulating the ability of the PrgX transcription factor to repress the transcription of an operon encoding conjugation functions. Many Gram-positive bacteria regulate important processes, including the production of virulence factors, biofilm formation, sporulation, and genetic exchange using peptide-mediated signaling systems. The key master regulators of these systems comprise the RRNPP (RegRap/NprR/PlcR/PrgX) family of intracellular peptide receptors; these regulators show conserved structures. While many RRNPP systems include a core module of two linked genes encoding the regulatory protein and its cognate signaling peptide, the enterococcal sex pheromone plasmids have evolved to a complex system that also recognizes a second host-encoded signaling peptide. Additional regulatory genes not found in most RRNPP systems also modulate signal production and signal import in the enterococcal pheromone plasmids. This review summarizes several structural studies that cumulatively demonstrate that the ability of three pCF10 regulatory proteins to recognize the same 7-amino-acid pheromone peptide arose by convergent evolution of unrelated proteins from different families. We also focus on the selective pressures and structure/function constraints that have driven the evolution of pCF10 from a simple, single-peptide system resembling current RRNPPs in other bacteria to the current complex inducible plasmid transfer system.

BACKGROUND AND SIGNIFICANCE

In 1965, Tomasz (1) described “a new type of regulatory mechanism in bacteria,” in which the control of competent cell genetic transformation in pneumococci was expressed in a density-dependent fashion (1). He reported that the culture medium of cells grown to the optimal density for maximum competence contained a soluble factor capable of inducing competence expression when added to low-density noncompetent cultures. Conceptually, the phenomenon of density-dependent pneumococcal competence expression mediated by intercellular signaling molecules is very similar to the autoinduction of light production in marine Vibrio species described a few years later by Nealson (2). These seminal studies initiated a paradigm shift in microbial research, changing the concept of normal bacterial behavior from single cells acting independently to coordinated behaviors of microbial populations via communication between individuals. Quorum sensing, in which a single cell type monitors its population density to coordinate activity (3), is perhaps the best studied mechanism for the modulation of multicellular behaviors by intercellular signaling, which is more broadly termed sociomicrobiology (4).

Enterococcus faecalis is a major cause of opportunistic infections of hospital patients, and E. faecalis clinical isolates are notorious for their carriage of antibiotic resistance genes (5, 6). These are frequently disseminated by conjugation. In 1978, Dunny et al. (7) reported that donor/recipient clumping and conjugative transfer of plasmids in Enterococcus (formerly Streptococcus) faecalis could be induced by low-molecular-weight signaling molecules excreted by recipient cells and sensed by plasmid-containing donor cells; it was suggested that these signals served as bacterial sex pheromones. A few years later, the Clewell et al. (8) and Suzuki et al. (9) research groups reported the identification of several different molecules that mediated signaling for various plasmids; these signals were unmodified hydrophobic peptides 7 to 8 amino acid residues in length. These studies were the first demonstrations that the prevalent extracellular signaling molecules of Gram-positive bacteria were oligopeptides, in contrast to the acyl-homoserine-lactone signals that frequently mediate quorum sensing in Gram-negative microbes (10). Both the peptide-mediated signaling mechanisms and the peptide signals themselves fall into two categories. Some signals are secreted as unmodified peptides processed from longer precursors, while others are both processed and posttranslationally modified (11–13). Likewise, sensing of peptide signals can involve either signal transduction across the membrane or signal import, followed by binding to a cytoplasmic receptor protein, which is often a transcription factor (14).

The enterococcal sex pheromone systems function by import of a signaling pheromone peptide encoded by the chromosome. For simplicity, we use “C” as an abbreviation for all conjugation/clumping-inducing peptide pheromones, where cCF10 is the peptide that specifically induces cells carrying pCF10, cAD1 induces
those carrying pAD1, etc. Mature C is processed by host-encoded proteins, and all known members of the sex pheromone family are processed from the cleaved signal peptides of secreted lipoproteins (15, 16). Binding of imported C by its cytoplasmic receptor initiates the pheromone response in the donor; the presence of C in the growth medium of donor cells thus serves as a cue for the presence of recipients (Fig. 1). Peptide binding modulates the ability of the C receptor (PrgX in the case of pCF10) to regulate the transcription of an operon containing conjugation genes (17). However, the enterococcal sex pheromone systems have several additional layers of complexity, including a second plasmid-encoded peptide (inhibitor [I]) that competes directly with C for binding to the same receptor (17, 18). In addition, several layers of posttranscriptional regulation greatly amplify the direct effects of the peptides on the expression of conjugation genes (17). The remainder of this essay will focus on the tetracycline-resistant pheromone-responsive plasmid pCF10 to illustrate the salient features of many sex pheromone plasmids (19) and to explore how the current complex systems may have evolved from simpler progenitor systems similar to the peptide-regulated RRNPP signaling systems that have now been implicated in the control of virulence, developmental processes, and horizontal gene transfer in numerous Gram-positive pathogens (20, 21).

OVERVIEW OF THE PEPTIDE-MEDIATED REGULATION OF pCF10 CONJUGATION

Figure 2 depicts a simplified map of the pheromone-inducible conjugation genes found on enterococcal plasmids (approximate size of the entire region indicated at the top). This map depicts the prg genes of pCF10 with single-letter designations, but similar gene content and organization are found on other well-studied plasmids, such as pAD1 and pPD1 (17). The left portion of the map shows conserved genes involved in pheromone sensing, and the relative locations of the genes of the pheromone-inducible prgQ operon encoding the I peptide, surface adhesin gene module (ABUC), downstream type IV secretion system (T4SS) genes, and conjugative DNA transfer genes (Dtr) are shown. The prgQ gene encodes the production of I, whereas an ~1-kb segment between prgQ and prgA encodes two small open reading frames (ORFs) and small RNAs (sRNAs) that regulate the expression of downstream genes posttranscriptionally (65). The sizes of the individual genes are not drawn to scale. I, the putative origin of the system as a surface protein module negatively regulated by quorum sensing through the X/Q cassette; this gene pair resembles RRNPP systems recently identified in numerous Gram-positive pathogens (21, 31). II shows how the system became more complex as it acquired the ability to enable its host cell to recognize C as an indicator of close proximity of plasmid-free recipients (mate sensing). At the mechanistic level, the C peptide competes with I, which functions as a classic quorum-sensing signal of donor density (self-sensing) (64). Evolution of the ability to differentially respond to these two antagonistic peptides was accompanied by the acquisition of genes encoding an oligopeptide binding protein, PrgZ, which binds both C and I with high affinity and increases their import via the Opp ABC transporter (37, 38), and PrgY, a predicted membrane peptidase that reduces the production of endogenous C by the host cell (36). III depicts the acquisition of T4SS and Dtr genes conferring conjugative transfer ability. There is high conservation of the regions indicated by I and II among many pheromone plasmids, suggesting that they all arose from a common ancestor, but step III likely occurred multiple times to link different conjugation gene cassettes to the pheromone-inducible aggregation module.
PrgX controls the initiation of transcription of this long operon from the prgQ promoter; the interaction of I with PrgX reduces transcription, whereas the interaction of C with PrgX allows for increased transcription. It is important to note that the direct effects of the peptides on control of the prgQ promoter by PrgX are actually quite modest, but they are greatly amplified by several posttranscriptional mechanisms, which are described elsewhere (23–27). Determination of the structures of Apo–PrgX and of PrgX bound to I or C, along with extensive genetic and biochemical analyses, indicates that Apo–PrgX and PrgX–I complexes repress transcription from the prgQ promoter, while PrgX–C complexes are impaired in repression (28, 29). It was originally suggested that the replacement of I by C in PrgX–DNA complexes disrupts PrgX tetramers within repressing complexes, allowing RNA polymerase to access the prgQ promoter (28–30). Very recent data (Y. Chen, A. Bandyopadhyay, B. K. Kozlowicz, H. A. H. Haemig, A. Tai, W.-S. Hu, and G. Dunny, unpublished data) suggest that PrgX forms tetramers when complexed with either peptide, but conformational differences of the DNA-bound tetramers account for differential repression. In both models, the ultimate induction state of a donor cell is dependent on the relative intra-cellular levels of I and C in donor cells. Interestingly, all of the peptide-controlled transcription factors of the RRNPP family appear to have a structure that is very similar to that of PrgX (20, 21, 31), and in most cases, the gene organization of the determinants for regulatory protein and the cognate regulatory peptide is similar to that of prgX and prgQ. Below, we focus on the evolutionary processes that likely shaped the emergence of the dual-peptide-controlled pCF10 system and how it may have evolved from a simple RRNPP-like system to its present complex state.

**HOW AND WHY DID THE pCF10 SYSTEM BECOME SO COMPLEX?**

The key functional components of the pCF10 system, which are also found in other pheromone plasmids (17, 32), are illustrated in Fig. 2. It is likely that current pheromone-inducible conjugation systems originated from a system with a single I-regulated Q-X-like module. This module likely controlled the expression of adjacent genes for surface adhesins, as similar surface adhesion gene content and organization are conserved with other pheromone-controlled systems (17). Contemporary pheromone plasmids may have a common ancestor that includes contiguous genes corresponding to prgZ through prgQ and extending through the downstream cassette of LPXTG-anchored cell surface protein genes (prgABC) and the small regulatory prgU gene (22, 33); this gene cluster is indicated by the Roman numeral II in Fig. 2. Prior to the acquisition of the ability to recognize the peptide signal C, the I-autoregulated X–Q surface protein cluster may have functioned to increase the ability of the host bacterium to attach to other bacterial or metazoan host cells at low density while reducing these interactions at high bacterial density to enable escape from stagnant communities and recolonization of new niches.

The next major event in the evolution of the system was probably the ability to recognize the host-encoded C peptide as an indicator of the presence of plasmid-free enterococci in close proximity. In strict evolutionary biology parlance (34, 35), C would be classified as a cue rather than a signal, since the sensing system seems to have hijacked this molecule produced from a gene not linked to the sensing genes in physical proximity or in function. In the case of pCF10, the C peptide is produced by processing the cleaved signal peptide of a predicted secreted lipoprotein CcfA, whose function has not been demonstrated (15); likewise, all known pheromone-responsive plasmids analyzed to date encode a response to a specific peptide encoded by one of the >50 potential lipoprotein genes in the organism (16). As indicated in step II, the system acquired additional components that recognize C; PrgY prevents self-induction of donors by decreasing the amount of mature C released (36), and PrgZ binds both C and I and facilitates their import into the cell via a chromosomally encoded peptide transporter (37, 38). PrgX also needed to evolve to recognize C and I. These 3 proteins are all from different families and share only 9 to 13% sequence identity and no significant homology at the structural level. We have structural data on the interactions of PrgZ with C (37) and of PrgX with both C and I (28, 29), but to date, there are no structural data available on PrgY.

PrgZ belongs to the family of substrate-binding proteins found in ABC transporters, G protein-coupled receptors (GPCRs), and DNA binding proteins (39, 40). It is likely that PrgZ evolved from a chromosomal oligopeptide-binding protein. Previous experiments have shown that the oligopeptide-binding protein OppA of *E. faecalis* can facilitate the import of C, even though a higher concentration of C is required than is produced by recipients under normal physiological conditions (38). PrgZ can bind both C and I and has a typical Venus flytrap fold of a cluster C substrate-binding protein (37), with C bound within an internal cavity (Fig. 3a). C is firmly bound to PrgZ via 10 direct hydrogen bonds. These bonds are mostly formed with the peptide backbone, with one exception, an H-bond that is formed with the side chain of Thr3, giving an explanation for results from genetic screens that Thr3 of C was important for PrgZ binding (41). Further H-bonds between C and PrgZ are formed via bridging water molecules, and there is also a salt bridge that anchors the N terminus of C. Although no structure of PrgZ complexed with I is available, it is highly likely that I binds in the same way as C (37). This is expected due to the similarities of PrgZ to other oligopeptide-binding proteins (37, 40, 42, 43).

As noted previously, PrgX serves as the primary cytoplasmic receptor for both peptides and acts as the master transcription regulator of the prgQ operon. PrgX has a conserved helix-turn-helix (HTH) domain responsible for its interaction with DNA (28, 29). The binding site for C and I is located in the larger dimerization domain (Fig. 3b and c), where both peptides form β-sheet-like interactions with PrgX and have a similar amount of H-bonds to PrgX. There is no conservation in the binding site between PrgX and PrgZ at either the sequence level or structurally. The two peptides favor different conformational states of PrgX, with the consequence that one alpha-helix is preferentially stabilized (I bound) or unwound (C bound) (28, 29). These different conformations either favor transcriptional repression when inhibitor is bound, or they reduce the level of repression of PrgX when C is bound. While the basis for the differential effects of the two peptides on the conformation of PrgX is not completely understood, it probably relates to the N-terminal regions of the two peptides. The bulky leucine residue at the N terminus of C likely crowds the surrounding PrgX residues in the binding pocket to a greater extent than the alanine at the N terminus of I (for a partial illustration, compare the PrgX residues surrounding the bound peptide N termini in Fig. 3b and c). These differences may indirectly affect the conformation of the C terminus of PrgX.

PrgY is required to prevent cells carrying pCF10 from being
self-induced by their own endogenous pheromone (36). Its amino acid sequence, in conjunction with genetic and biochemical studies, suggest that a C-terminal subdomain anchors the protein in the membrane with the N-terminal region outside the cell; the external N-terminal subdomain confers the ability to specifically bind the mature C peptide (36, 44) and may contribute to its degradation. Initial studies of PrgY suggested that similar proteins, none with known functions, were present in organisms from all kingdoms, and that the protein phylogenies correlated with those of the host organisms (36). Recently, an important new study provided new insights into the structure/function relationships of these proteins. Zhang et al. (45) identified Tiki as a protease family playing a critical role in cell growth and development via specific cleavage of the Wnt protein (45). PrgY is homologous to the human Tiki metalloprotease, both having a pair of GXH motifs and a conserved glutamate residue, and it is predicted to have structural similarity to the so-called EraA/ChaN-like family of proteins (46). The structure of PrgY has not been determined, but structural modeling using Phyre2 gives a model with a 96% confidence level over most of the extracellular domain (Fig. 4). This model does not contain any structural motif that resembles the pheromone-binding site of either PrgZ or PrgX. From the homology to the Tiki metalloproteases, we can deduce which residues likely form the active site in PrgY, with some of those specific residues, like His21, having previously been verified to be important for function (36, 44). To date, only PrgY and Tiki are known to have specific interactions with polypeptide substrates.

The cumulative analysis suggests that the pCF10 system did not independently evolve these 3 different components from a single protein with a peptide-binding motif. More likely, an ancestral system, i.e., the inhibitor-regulated Q-X module, at some point acquired genes that coded for the early versions of PrgY and PrgZ, and those proteins then evolved specific binding affinity to the cognate C and I peptides, as illustrated in Fig. 2. The downstream genes predicted to encode the type 4 secretion systems (T4SSs) and DNA transfer (Dtr) machinery required for conjugation are located immediately downstream from the surface protein cassettes in other pheromone plasmids, but these T4SS loci

FIG 3 Comparison of the peptide binding of PrgZ (gray) complexed with C (a) and PrgX (dark red) complexed with C (b) or I (c). Each upper subfigure shows the full protein structure in a cartoon representation, with the bound ligand in spheres, C in teal, and I in green. The lower enlarged representations at the bottom show the ligand (as sticks) with interacting protein residues. As can be seen by comparing panels b and c, the C peptide has a very different structure when bound to PrgZ compared with its structure when bound to PrgX.

FIG 4 Predicted structure of PrgY. The extracellular part of PrgY, here shown as a cartoon representation, was modeled using Phyre2 and colored from the N terminus (blue) toward the C-terminal end of the model (yellow). The C-terminal domain, which could not be modeled, is predicted to contain 4 transmembrane helices, shown here as rectangles in a membrane. The predicted active site, based on the homology of PrgY to the Tiki metalloproteases (46), is shown within the dashed line.
show considerable divergence (22). This suggests that pheromone-inducible aggregation cassettes became linked to the additional components required for conjugation on multiple occasions (Fig. 2, step III). Interestingly, the available data suggest that the downstream conjugation functions for all known plasmids are transcriptionally regulated by the peptide signals even though they became linked to the upstream regions in multiple events (22).

REMAINING QUESTIONS AND FUTURE DIRECTIONS

While significant questions about the molecular mechanisms of pheromone-mediated control of conjugation remain, the most compelling areas for future study may be the analysis of structure/function relationships of the key regulatory components and investigations of how these systems function in the natural environment, including their impacts on maintenance and dissemination of the plasmid itself, and on the fitness of the bacterial hosts. The significance of such studies is heightened by the fact that considerable experimental and theoretical investigations of the evolutionary aspects of the well-studied acyl-homoserine-lactone autoinducer systems in Gram-negative alphaproteobacteria have been carried out (reviewed in reference 47). For example, a recent study by Cornforth et al. (48) provided evidence that combining two separate quorum-sensing systems allowed for greater resolution of the local environment of a bacterium, both in terms of sensing social (microbial cell population density) and physical (diffusion, etc.) parameters. In contrast, much less attention has been directed toward the naturally occurring and more complex intercellular communication system represented by pCF10. While pCF10 was discovered because of its role in the transmission of antibiotic resistance (49), enterococcal pheromone-responsive plasmids frequently do not carry resistance genes (50, 51), suggesting that they may encode other traits that increase the fitness of their host cells, relative to the costs of plasmid maintenance. It is interesting to consider how the expression of the pheromone-inducible conjugation genes may impact host fitness and how this relates to the regulatory properties of the system. Published data from our group (52–57) and others (58–61) suggest that the expression of aggregation substance proteins, such as PrgB, can increase colonization and virulence by promoting biofilm formation and attachment to host tissues, and by increasing resistance to phagocytic killing. Notably, there are still no direct data on how PrgB or other inducible proteins might impact fitness in the gut. On the other hand, the overexpression of these genes likely has very high costs for the induced cell, including the energy required for synthesizing conjugation proteins, the likely inhibition of growth in cells trapped in large aggregates, and cell death and lysis due to toxic effects of overexpressed gene products on highly induced cells (62). Interestingly, clusters of genes related to the plasmid-encoded pheromone-inducible adhesins/transfer determinants have been identified within genomic islands in the chromosomes of some strains, but these chromosomal determinants are not capable of transfer unless a coresident plasmide encoded plasmid integrates and mobilizes them via an Hfr-like mechanism (63).

Numerous studies have documented the extremely tight regulation of the pheromone system (19). The system not only avoids spurious induction but also limits the duration of induction due to the fact that the induction process itself dramatically increases inhibitor production, leading to rapid shut off of the response after a short period of induction (64). Furthermore, the inhibitor can function as a classic quorum sensor of donor density; at high donor densities, donors are poorly induced even by high concentrations of C (64). These cumulative effects of the inhibitor apparently limit the extent of induction in mixed populations of donors and recipients. This raises the question of whether the system may have evolved to maintain mixed populations of donors and recipients in shared niches in the natural environment of the bacteria, e.g., the intestinal tract. The maintenance of recipient populations by limiting their conversion to donors should result in a steady supply of C within the niche, whose inducing capacity is limited by the inhibitor. In this scenario, basal levels of expression of the inducible genes could be maintained within the mixed population, providing the previously described benefits (note that induction of a few donors can coaggregate recipients and uninduced donors in close proximity) while minimizing costs of overexpression. The pheromone system has thus evolved under strong conflicting selective pressures for an extremely sensitive detection system to induce expression while simultaneously limiting the extent and duration of induction. This may have driven convergent evolution of the three unrelated proteins with vital, but distinct, regulatory functions to recognize the same peptide. Direct experimental testing of these ideas in the mammalian gastrointestinal (GI) tract, along with further mechanistic and structural studies of regulatory components, is in progress and might yield insights into more effective approaches to reduce the spread of antibiotic resistance and to impair the ability of resistant strains to overgrow and disrupt the gut microbiota of hospital patients.

ACKNOWLEDGMENTS

The research for this review was supported by U.S. PHS grant GM49530 to G.M.D. and by the Kempe Foundations grant JCK-1524 to R.P.-A.B.

FUNDING INFORMATION

This work, including the efforts of Gary M. Dunny, was funded by HHS | National Institutes of Health (NIH) (GM49530). This work, including the efforts of Bonnie Per-Arne Berntsson, was funded by Kempeförfattarna (Kempe Foundations) (JCK-1524).

REFERENCES

pheromone, CAD1, that induces plasmid transfer in Streptococcus faecalis. FEMS Lett 178:97–100. http://dx.doi.org/10.1016/0167-5260(91)874124-J.
Gary M. Dunny received his B.S. and Ph.D. from the University of Michigan and spent 11 years at Cornell University as a postdoctoral fellow and as a faculty member before moving to the University of Minnesota in 1989, where he is currently professor of microbiology. He has studied conjugation, cell signaling, and adaptation in enterococci using genetics, biochemistry, and microscopic imaging for his entire career.

Ronnie Per-Arne Berntsson studied biotechnology at Chalmers University in Gothenburg, Sweden. In 2010, he received his Ph.D. in biochemistry from the University of Groningen, the Netherlands, after working in the groups of Bert Poolman and Dirk-Jan Slotboom on studies of ABC transporters and their domains. After his Ph.D., he moved to Stockholm University, Sweden, where he received an EMBO fellowship to do postdoctoral research in the group of Pål Stenmark on botulinum neurotoxins and their receptors. In 2015, he became an assistant professor at the Department of Medical Biochemistry and Biophysics at Umeå University, Sweden. His laboratory studies the function, structure, and regulation of type 4 secretion systems in Gram-positive bacteria.