1. INTRODUCTION & REVIEW OF LITERATURE

All forms of life have the ability to sense a change in the environment and respond by modulating the gene expression. The regulation of gene expression is crucial for all living organisms. Bacteria thrive in complex environments such as temporal shifts in nutrient availability, adverse conditions such as temperature, pH, and osmotic stress by exploiting the genetic information to rapidly respond towards the change. These rapid responses and adaptation to discrete growth conditions rely on the intricate mechanism of gene regulation, which is mediated by a complex interlinked network of multiple regulators.

1.1 Fine tuning the gene expression in bacteria: An introduction

Bacteria have evolved mechanisms to adapt constantly to the distinct niche they encounter during colonization by regulating the expression of multiple virulence factors. Bacteria use diverse and multi-layered gene regulatory mechanisms that act at different levels in the flow of genetic information and control the expression of target genes. The expression of these virulence genes is controlled not by one but by a combination of diverse mechanisms at transcriptional or translational levels.

1.1.1 Modulation of transcription

The pioneering and groundbreaking investigation of Jacob and Monod on the regulation of *lac* operon in *E. coli* gives the first insight into bacterial gene regulation (Jacob and Monod, 1959; Jacob and Monod, 1961). Following this discovery, several distinguished mechanisms of gene regulation have been identified. Bacterial gene expression primarily depends on the recognition of promoter by the DNA-dependent RNA polymerase to initiates transcription. RNA polymerase consists of a catalytic core enzyme with subunit composition $\alpha_2\beta\beta'\omega$ and associates sigma factor (σ) (Reviewed in Browning and Busby, 2004). Bacteria fine-tune their transcriptional programs in response to changing environments, through diverse mechanisms that regulate the association of RNA polymerase subunits, activity of RNA polymerase and promoter recognition mediated by various factors including transcription factors. Transcription factors (TF) are key regulatory molecules with DNA binding domains that regulate the transcription of a gene in response to environmental cues. Transcription factors regulate the promoter activity by several mechanisms that activate or repress the transcription.

The TF may exerts a negative effect on the transcription through various mechanisms including alteration of DNA structure such as DNA looping which prevents the binding of RNA polymerase at the promoter region, steric hindrance by binding of TF at the region overlapping to –10 or –35 elements of the promoter which in turn block the access of RNA polymerase to the promoter and also acting as an anti-activator that repress the transcription (reviewed in Bervoets and Charlier, 2019). One of the well-documented examples of transcription repression by DNA looping includes the galactose operon's repressor protein GalR in *E. coli*. GalR is a dimeric protein that binds to two spatially separated flanking regions in the operator element and forms the DNA looping between these sites mediated by protein-protein interaction (Semsey *et al.*, 2004).

The TFs can stimulate the rate of transcription by acting as the transcriptional activator. Activator TF binds upstream the -35 promoter, or between the -10 and -35 promoter region, in class I and class II activation respectively. This binding stimulates the RNA polymerase recruitment at the promoter region and enhances the transcription (reviewed in Lee, Minchin and Busby, 2012).

Regulation of transcription also includes alternative sigma factors and phosphorylation-dependent response regulators proteins such as PhyR which upon phosphorylation modulates the gene expression by mimicking sigma factor and control the stress response in *Methylobacterium extorquens* (Francez-Charlot *et al*., 2009). Sequestration of the RNA polymerase by 6S RNA is another well documented example of global regulator of gene expression in bacteria that expressed under different stress condition and involved in the pathogenesis in many pathogenic bacteria (Wassarman, 2018). 6S RNA of *E. coli* is a global transcription regulatory non-coding RNA that mimics promoter and sequester RNA polymerase holoenzyme (Wassarman, 2018; Wassarman and Storz, 2000).

1.1.2 Modulation of translation

Adaptation to change in the environment does not only depend on transcriptional regulation, but regulation at the level of translation also plays a crucial role in rapid responses to challenges. The translation, which decodes genetic information, is a dynamic process partitioned into three sequential phases namely initiation, elongation, and termination.

Translation initiation is a key regulatory step in protein synthesis, demands significant energy pools in the cell, and a tightly controlled process including sequestration of the ribosome-binding site and control of the number of available ribosomes (Condon *et al.*,

1993; Sherwood and Henkin, 2016). Sequestration of the ribosome-binding site (RBS) through mRNA folding or binding of small RNA reduces the accessibility of ribosome and represses the translation of protein (Meyer, 2017; Nakagawa *et al*., 2017). Binding of small RNA also positively regulates the translation initiation by unmasking the loop structure preventing ribosome accessibility to RBS through base pairing in the upstream region of RBS such as DsrA and RNAIII (Majdalani *et al*., 1998; Morfeldt *et al*., 1995).

The folding of mRNA is the most important property of RNA that controls the gene expression response to multiple stimuli, such as temperature, pH, and metabolites. The efficiency of translation initiation depends on the structure surrounding the RBS that controls the gene expression via *cis*-regulatory elements riboswitches. Riboswitches are RNA-based non-coding structural elements present in some mRNAs that regulate the gene expression in response to binding a wide range of small-molecule ligands. Riboswitches undergo conformational changes upon binding of small-molecules ligands at the aptamer domain and, in consequence, change the gene expression pattern (reviewed in Lotz and Suess, 2018). Riboswitches are an example of thermosensing RNA that regulates the expression of genes in response to temperature shifts (Sherwood and Henkin, 2016). The majority of riboswitches encoded from the 5'UTR of mRNA and forms the secondary structure sequestering the RBS by intrinsic base pairing at low temperature conditions. At high temperature conditions melting of secondary structure allows ribosome binding and initiates translation.

A translation pausing at the level of translation elongation can be induced by a different mechanism including RelA mediated synthesis of (p)ppGpp which induces the stringent response that controls the ribosome biosynthesis (Brown *et al*., 2016; Starosta *et al*., 2014; Traxler *et al*., 2008). Regulation of translation during termination requires the involvement of release factors RF1 -RF3, for peptide hydrolysis and ribosome recycling (Ii and Ibba, 2020). Regulation of translation by interacting with other RNAs plays significant role in gene regulation in bacteria. Many *cis* or *trans* encoded small RNAs interact with the different mRNAs and control the translation of protein in response to diverse environmental condition.

1.2 Small regulatory RNAs

In the last two decades, small regulatory RNAs (sRNAs) are established as a significant element of gene regulation in the control of bacterial adaptation in the different environments. Bacterial sRNAs are 50-500 nucleotides long regulatory molecules that

fine-tune various physiological processes and control the gene expression (Chakravarty and Massé, 2019). The sRNAs exert their regulatory role by base pairing with *cis*- or *trans*- encoded target mRNAs, to influence their stability and translation (Caldelari *et al*., 2013).

1.2.1 Classification of small regulatory RNAs

sRNAs are usually classified into three major classes: *cis*-encoded sRNAs, *trans*-encoded sRNAs, and sRNA modulating the proteins.

1.2.1.1 *Cis*-encoded sRNAs

Cis-encoded sRNAs are transcribed in the opposite direction of their target mRNAs and share around 75 nucleotides extended regions of full complementarity with their target mRNA (Figure 1) (Brantl, 2007; Waters and Storz, 2009). These *cis*-encoded small RNAs are also known as antisense RNAs and are involved in a plethora of physiological effects by duplex formation with target mRNAs which leads to transcription termination or mRNA cleavage (Georg and Hess, 2011).

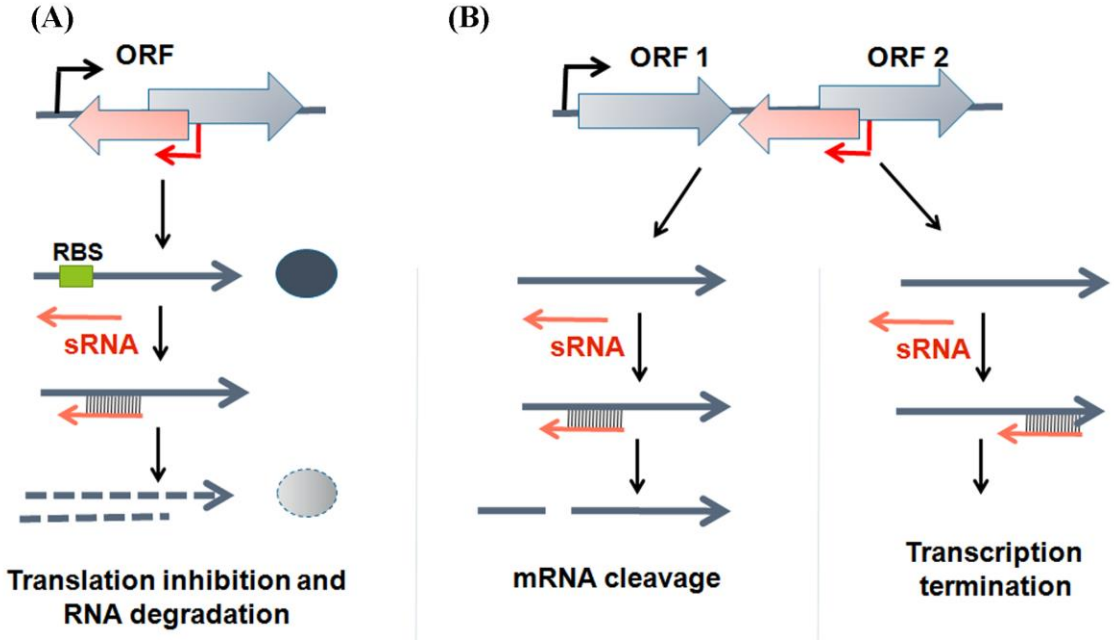

Figure 1: Regulatory role of *cis*-encoded sRNA (Diagrammatic representation)
(A) *cis*-encoded sRNAs (orange), expressed from the antisense strand at the 5' UTR of its target mRNAs (grey) share extensive complementarity. Base pair interaction between sRNA and mRNA inhibits ribosome binding and often leads to mRNA degradation. (B) sRNAs expressed from the antisense strand in the region separating two genes in an operon. This antisense sRNA base pairs with target mRNAs and controls its expression by recruiting RNases to induce mRNA cleavage or terminating the transcription.

The antisense transcript RNAI, encoded from plasmid ColE1, is a well-studied example of *cis*-encoded sRNA (Brantl, 2007). The *cis*-encoded antisense RNAs also act as RNA antitoxins to regulate the expression of toxic proteins (Fozo *et al.*, 2008; Harms *et al.*, 2018). IsrA is an example of *cis*-encoded sRNA that overlaps with the 5' end of *iasE* mRNA encoding SOS-induced endonuclease toxin in *S. typhimurium* (Acuña *et al.*, 2020).

1.2.1.2 *Trans*-encoded sRNAs

Trans-encoded sRNAs in contrast to the *cis*-encoded sRNAs, share very short complementarity with their target mRNAs and regulates the translation and stability of target mRNAs (Aiba, 2007; Gottesman, 2005). *Trans*-encoded sRNAs interact at 5' UTR by often occluding the ribosome binding site and negatively regulates the expression of target RNAs (Sharma *et al.*, 2007). Besides, *trans*-encoded sRNAs can also stimulate the expression of target mRNAs through an anti-antisense mechanism by disrupting the secondary structure that occludes the ribosome binding site (Figure 2) (Gottesman, 2005; Hammer and Bassler, 2007; Urban and Vogel, 2008).

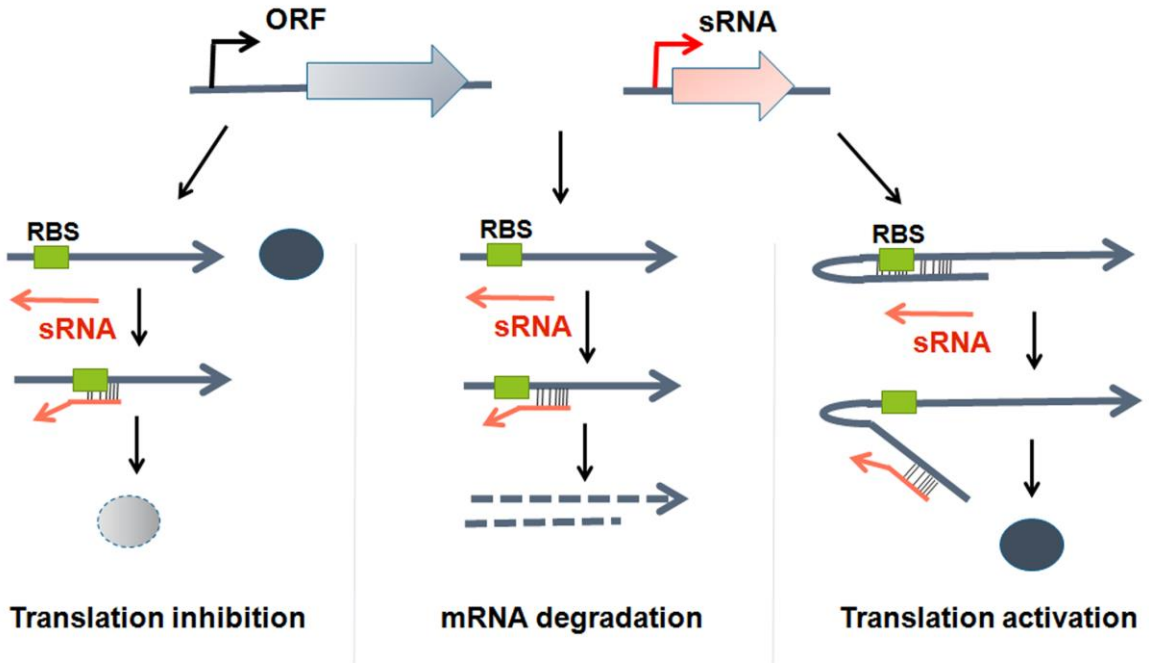

Figure 2: Depiction of regulatory role of *trans*-encoded sRNA

Trans-encoded sRNAs base pair with target mRNA at 5' UTR and negatively regulates its expression through translation inhibition and/or mRNA degradation. *Trans*-encoded sRNAs also act as a positive regulator by activating the translation of target mRNA using anti-antisense mechanism.

Trans-encoded sRNAs expressing in gram-negative bacteria often require RNA-binding proteins such as Hfq, ProQ, and CsrA, that promote sRNA-mRNA base-pairing interaction and stability (Jørgensen *et al.*, 2020). However, in gram-positive bacteria, sRNAs exert their regulatory functions, by base pairing with target mRNAs without the assistance of RNA chaperone.

1.2.1.2.1 The RNA-binding proteins

a) Hfq: Hfq is a well-characterized RNA-binding protein and was initially recognized as a host factor for replication of phage Qβ. Hfq interacts with the sRNAs and plays important role in bacterial physiology and virulence. Hfq is a homo-hexameric protein containing two RNA binding regions, called proximal and distal face (Sauter *et al*., 2003; Schumacher *et al*., 2002). Hfq is essential for the function of many small RNAs; it facilitates the binding between sRNA and target mRNA and also stabilizes the sRNA-mRNA duplex (Fender *et al*., 2010; Soper and Woodson, 2008).

b) CsrA: CsrA (carbon storage regulator A) is a conserved RNA-binding protein originally discovered in *E. coli* and present in a broad range of bacterial species encoding sRNAs. CsrA plays multiple roles in bacterial physiology and represents a global post-transcriptional regulator that control motility, quorum sensing, and biofilm formation (Durica-mitic *et al*., 2018; Romeo *et al*., 1993; Sabnis *et al*., 1995). CsrA, a sequence-specific RNA-binding homo-dimer protein that recognizes and binds at consensus binding GGA motif and GGA sequences (Potts *et al*., 2017; Titration, 2018) that leads to positive and negative regulation (Baker *et al*., 2002; Patterson-Fortin *et al*., 2013).

c) ProQ: ProQ is a novel RNA-binding protein that belongs to the member of FinO-domain proteins and is highly conserved in Gram-negative β- and γ-proteobacteria but is absent in Gram-positive species (Olejniczak and Storz, 2017). ProQ protects the RNA heteroduplexes from RNase mediated cleavage, limits the 3'-dependent RNA degradation, and coordinates the sRNA dependent translation in bacteria. ProQ mainly interacts with target mRNAs at 3' UTRs based on secondary structure, without the need of sequence-specific motifs (Holmqvist *et al*., 2018). *Salmonella* sRNA RaiZ is a ProQ dependent sRNA that represses the translation of histone-like protein HU-α via base-pairing with *hupA* mRNA at the RBS, facilitated by ProQ (Smirnov *et al*., 2017).

1.2.1.3 sRNAs modulating the proteins

Other small RNAs exert their functions by directly binding proteins, and without base-pairing to target mRNAs. The 6S RNA is the first example of sRNA that regulates sigma factor availability by mimicking a σ70 promoter complex. 6S RNA forms a complex with RNA polymerase and modulates the transcription activity (Figure 3) (Wassarman, 2018; Wassarman and Storz, 2000). Additional well-characterized sRNAs include CsrB and CsrC that sequester the regulatory protein CsrA, a key regulatory protein that binds and regulates multiple target mRNAs. CsrB and CsrC sRNAs contain multiple GGA binding sites, 22 and 13, respectively, for CsrA and indirectly regulate many target mRNAs involved in various physiological responses by reducing the concentration of CsrA protein (Suzuki *et al.*, 2006).

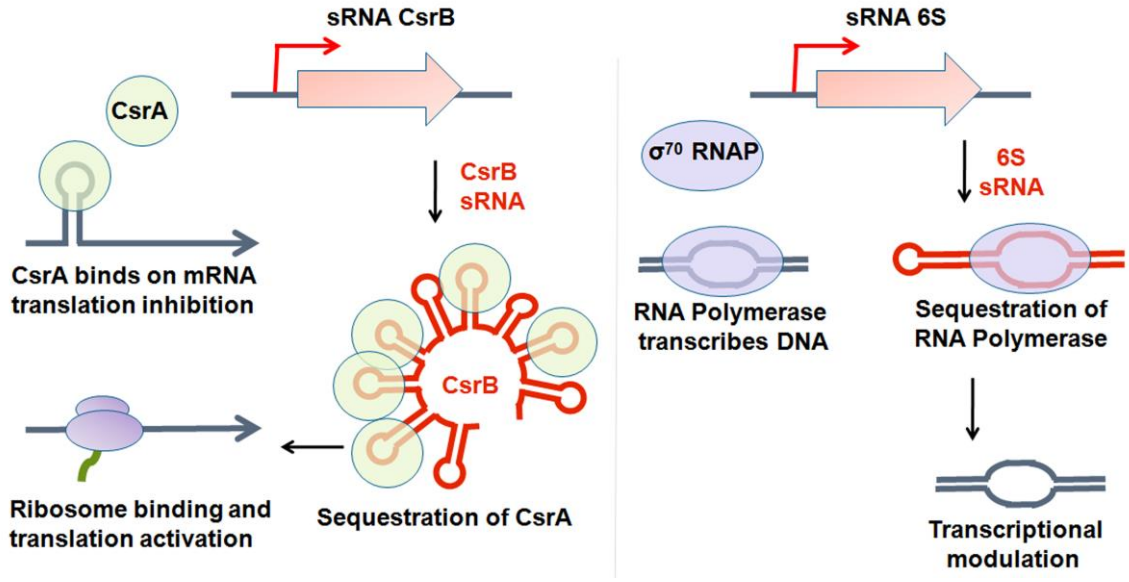

Figure 3: sRNAs modulating protein activity

sRNA CsrB interacts with the CsrA protein. CsrA binds with the target mRNA through a specific GGA motif present in the hairpins structure of mRNA and acts as a transcriptional regulator. When the concentration of CsrB increases during the stationary phase it sequesters the CsrA protein and inhibits its regulatory function (left panel). 6S sRNA also functions as a mimic that sequesters the sigma-70 RNA polymerase and controls its transcriptional function (right panel).

1.2.1.4 Dual Function small RNAs

From the large repository of sRNAs, a few sRNAs have been reported to contain small ORFs that encode functional peptides. These sRNAs are subcategorized as “dual-function sRNAs” that on one hand act as a base-pairing regulatory RNAs and on other hand

encodes a peptide that functions in the same or separate metabolic pathway. The *S. aureus* sRNA RNAIII is an example of dual-function sRNAs that controls many virulence factors through base-pair interaction and also encodes a small 26 amino acid protein, delta hemolysin (Boisset *et al*., 2007). Another sRNA, *psm-mec*, is expressed from the staphylococcal cassette chromosome, and suppresses the translation of *agrA* transcript by base-pairing within the coding region. Psm-mec is a 22 amino acid small toxic protein encoded by sRNA *psm-mec* and involved in the *S. aureus* virulence (Kaito *et al*., 2011). More examples of dual-function-sRNAs include the SgrS and AzuCR in *E. coli* (Raina *et al*., 2021; Wadler and Vanderpool, 2007), and, SR1 and SR7 in *Bacillus subtilis* (Gimpel *et al*., 2010; Haq *et al*., 2020). A recent study showed that *E. coli* sRNA AzuCR encodes the 28 amino acid small protein AzuC which impacts the glycerol metabolism by enhancing the activity of GlpD, glycerol-3-phosphate dehydrogenase protein, required in the catabolism of glycerol (Raina *et al.*, 2021).

1.3 Mechanism of sRNA mediated gene regulation

The sRNA-mRNA base-pairing interactions and consequences have been analyzed in detail for many sRNAs. These interactions involve a continuous stretch of around 8 nucleotides known as seed region, involved in base pairing interaction which positively or negatively influences the translation or stability of mRNA (Bronner *et al*., 2020; Sharma *et al*., 2007; Vanderpool and Gottesman, 2004).

1.3.1 Positive and negative regulation of translation

The base pairing between sRNA and mRNA in the upstream region of the ribosome binding site activates the translation through an anti-antisense mechanism by removing the secondary structure which blocks the ribosome accessibility at the RBS (Figure 4A). The sRNA DsrA is the well-characterized example of sRNA-mediated translation activation of stress σ factor, RpoS. The presence of secondary stem-loop structure in the upstream region of the start codon in *rpoS* mRNA inhibits the ribosome binding. DsrA interacts with the upstream region in the 5' UTR of *rpoS* mRNA that uncovers the RBS by disrupting the stem-loop interaction and activates the translation of *rpoS* mRNA by an anti-antisense mechanism (Kim *et al*., 2019; Majdalani *et al*., 1998; Soper and Woodson, 2008). Additional examples of sRNAs include the GlmZ, which activates the translation of *glmS* mRNA which codes for an essential enzyme in amino-sugar metabolism by preventing secondary structure formation (Urban and Vogel, 2008) and RNAIII that

activates the translation of *hla* mRNA via anti-antisense mechanism by unmasking the RBS (Morfeldt *et al*., 1995).

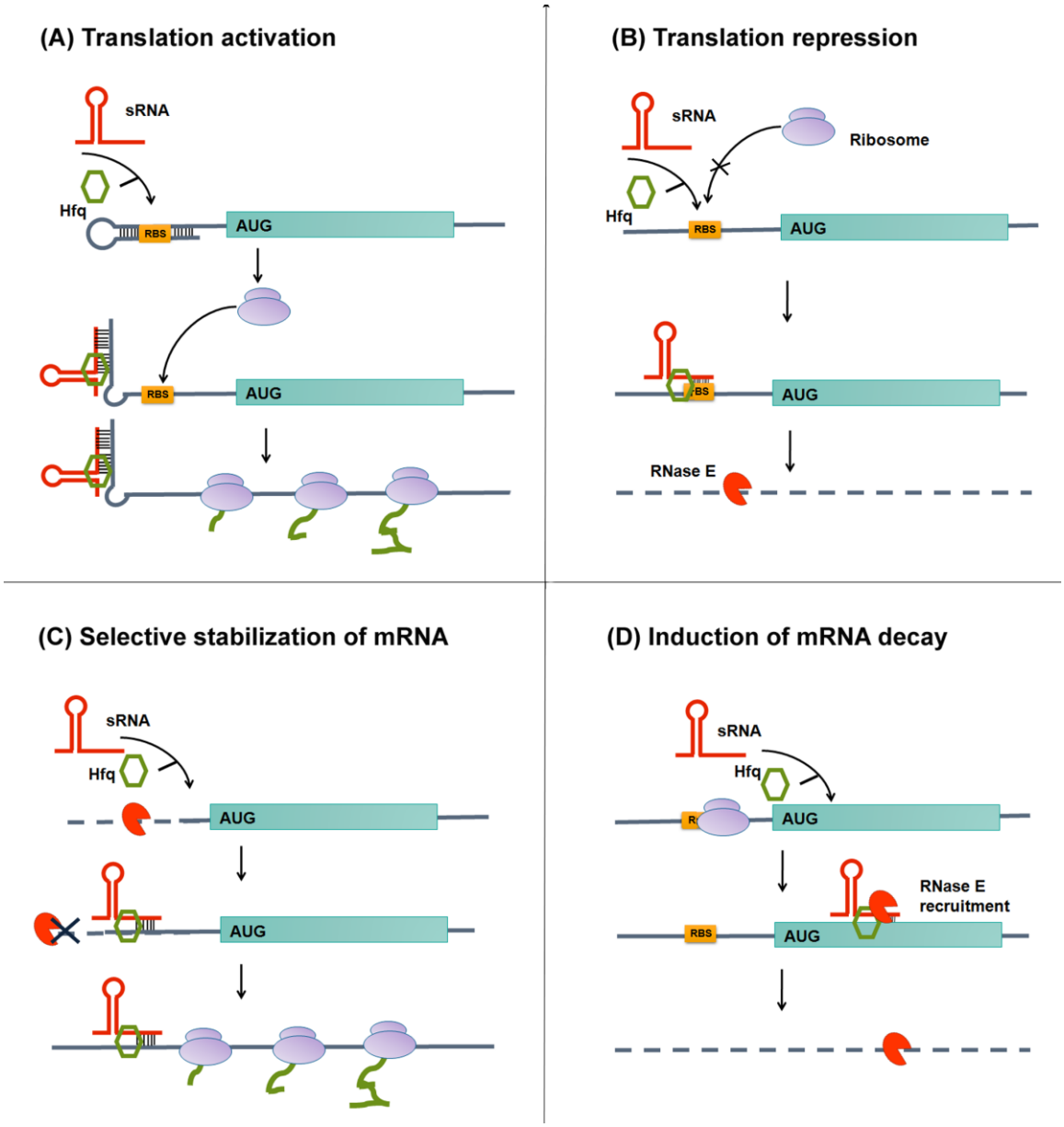

Figure 4: Mechanisms of sRNA regulation

(A) Translation activation: sRNA positively regulates the mRNA translation, by the anti-antisense mechanism, via disrupting the native secondary structure of target mRNA that blocks the ribosome binding site. Interaction of sRNA at the 5'UTR, upstream to the RBS, unblocks the ribosome binding, leading to translation activation. (B) Translation repression: sRNA base pairs in the vicinity of ribosome binding site on target mRNA that restricts the ribosome binding and indirectly leads to mRNA degradation. (C) mRNA stabilization: sRNA interferes with mRNA degradation mediated by RNases through base pairing, leading to stabilization of mRNA by hindering accessibility of RNase. (D) Induction of mRNA decay: sRNA-mRNA duplex formation can recruits the RNase E, resulting in mRNA degradation.

The most common mechanism to negatively regulate the translation is the interaction of sRNA alone or associated with Hfq at the region covering the ribosome binding site. This interaction occludes the ribosome entry and negatively regulates the translation of mRNA (Figure 4B). The negative regulation of outer membrane porin N by sRNA RybB is mediated by a long stretch of sRNA-mRNA interaction that occludes ribosome binding site and covers five codons (Bouvier *et al*., 2008). Negative regulation of translation is also mediated by the interaction of sRNA at the ribosome stand-by-site, disrupting the stem-loop structure, which then competes with ribosomes under initiation of translation. The sRNA IstR-1 negatively regulates the expression of SOS-induced toxin TisB by interacting with *tisB* mRNA at ribosome stand-by-site located in the ~100 nt upstream from the start codon (Darfeuille *et al*., 2007). Apart from the regulatory mechanism discussed above, a recent study illustrates the role of a sRNA, SgrS, in translation regulation by interacting with distant regions of the translation initiation window containing RBS and the initial five codons. The sRNA SgrS regulates the sugar transport by repressing the translation of *manY* mRNA via base-pairing with the 5' coding region. This interaction further recruits the Hfq at the upstream 5'UTR region of *manY* mRNA and inhibits the S1-dependent translation initiation (Azam and Vanderpool, 2020).

1.3.2 Positive and negative regulation of mRNA stability

The sRNA-mRNA interaction interferes with the RNase-mediated cleavage and positively regulates the gene expression by stabilization of target mRNA (Figure 4C). The sRNA SgrS control the glucose homeostasis by base pairing with *pldB-yigL* mRNA, hinders the RNase E mediated mRNA decay and increases the stability of mRNA (Papenfort *et al*., 2013). Another example of mRNA stabilization includes the base pairing between *Salmonella enterica* sRNA RydC and *cfa* mRNA (encoding cyclopropane fatty acid synthase) that block RNase E mediated mRNA degradation in the 5'UTR (Fröhlich *et al*., 2013). RNase E is a major regulator of RNA degradation in gram-negative bacteria; its homologs are absent in firmicutes such as *S. aureus* and *B. subtilis*. Added examples include RNAIII of *S. aureus* and RoxS sRNA in *B. subtilis*. RNAIII of *S. aureus* interacts with *mgrA* transcript at two distinct regions in 5' UTR and positively regulates its expression by blocking RNase mediated degradation leading to mRNA stabilization (Gupta *et al*., 2015). The RoxS sRNA expressed in *B. subtilis* pairs at the 5' end of *yflS* mRNA encoding malate transporter protein and blocks the RNase J1 mediated mRNA decay leading to increased mRNA stability and translation (Durand *et al*., 2017).

The base-pairing interaction between sRNA-mRNA can also reduce the ribosome binding and lead to destabilization of the mRNA. The duplex formation between sRNA-mRNA promotes mRNA degradation by recruiting the RNase E (Figure 4D). The sRNA MicC interacts with *ompD* mRNA (outer membrane protein D) in the coding region and promotes the RNase E mediated degradation of RNA in *Salmonella* (Pfeiffer *et al.*, 2009). Interaction of RNAIII with target mRNAs (*sbi*, *coa*, *rot*) form the RNA-RNA duplex that induces RNA degradation by RNase III and negatively regulates target expression in *S. aureus* (Chabelskaya *et al*., 2014; Chevalier *et al*., 2010; Geisinger *et al*., 2006). RNase III is a double-stranded specific endoribonuclease that regulates the turnover of sRNA and mRNA and is further involved in the maturation of rRNA and tRNA in *S. aureus* (Lioliou *et al.*, 2012).

1.4 Identification and characterization of small RNAs

The initial approach for the sRNA identification included the global bio-computational prediction, shotgun cloning or RNomics, and microarrays by examining the following characteristics of sRNAs: (1) Expression from the intergenic region, (2) small in size, (3) presence of Rho-independent terminator and (4) lack of ORFs (Chen *et al*., 2002; Saito *et al*., 2009). However, not all of the properties specified earlier are necessary as recent studies suggest that small RNAs could also be transcribed from 5' and 3' UTR of mRNAs, and that few sRNAs encode small ORFs (Figure 5). Advancement in genome-wide RNA sequencing approaches such as dRNA-seq (Bischler *et al*., 2015; Sharma and Vogel, 2014), CLIP-seq and Grad-seq (Smirnov *et al*., 2016) provides a platform for the global RNA landscape analysis.

1.4.1 Biocomputational approach

The computational tools can be used for the prediction of the protein-coding region from the available genomic data. This method primarily depends on the transcriptional features, conserved IGRs, the genomic location of putative sRNAs, and the presence of the Rho-independent transcription terminator. This strategy provides the rapid identification of many sRNAs in diverse bacteria. Using the phylogenic approach and availability of genomic data several potential sRNAs have been identified in related bacteria (Chen *et al.*, 2002; Saito *et al*., 2009).

1.4.2 RNomics approach

This strategy involves shotgun cloning of RNA, in which total RNAs were initially subjected to size fractionation, followed by directional cDNA cloning and sequencing. This method allows the detection of all RNAs in a defined size range without the requirement of prior knowledge concerning RNA features and led to the identification of many novel species-specific sRNAs as well as processed transcripts (Hüttenhofer, 2005; Vogel *et al*., 2003). RNomics was used to identify several non-coding RNAs in many bacteria (Bohn *et al*., 2010; Hüttenhofer, 2005; Vogel *et al*., 2003).

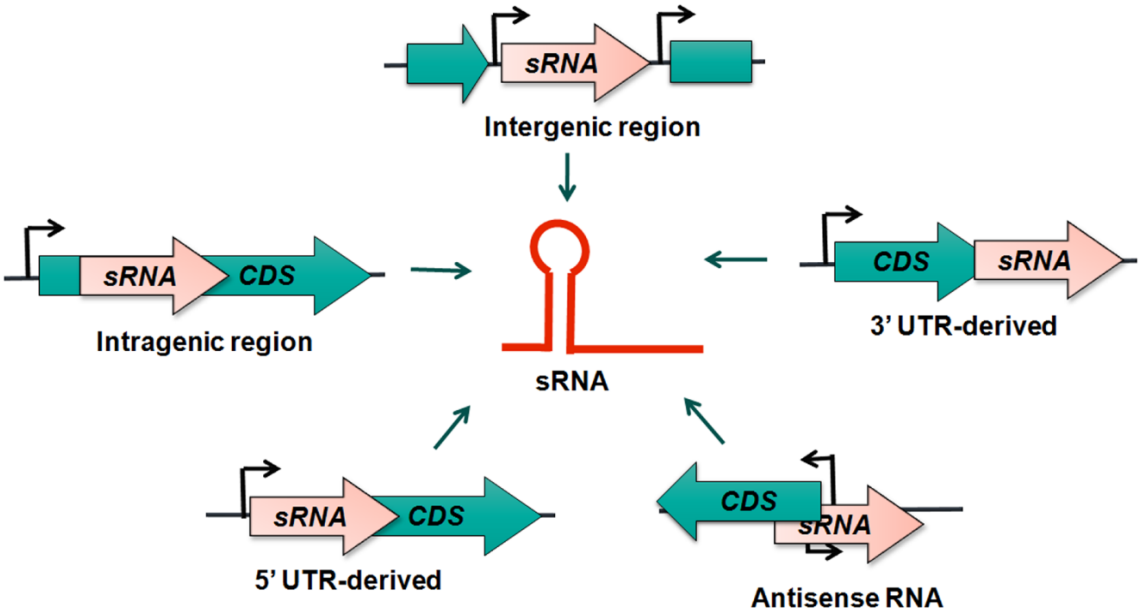

Figure 5: Genomic location of sRNA

sRNAs discovered recently, transcribed from distinct genomic loci including 5' and 3' UTRs, inter and intragenic region and antisense elements

1.4.3 Differential RNA-sequencing

Differential RNA-sequencing (dRNA-seq) is a deep sequencing approach based on selective sequencing of the primary transcript. The primary transcripts in bacteria contain 5' triphosphate (5' PPP), whereas processed transcripts carry a 5' monophosphate (5' P). This distinguishing feature is employed to specifically removing the processed RNAs by 5' P-dependent terminator exonuclease (TEX). This depletion selectively enriches the primary transcripts with transcriptional start site (TSSs) and provides evidence of specific novel transcripts at the single-nucleotide level (Bischler *et al.*, 2015; Sharma and Vogel, 2014). Additionally, dRNA-seq uncovers another emerging class of sRNAs that are

derived from the 3' UTR of mRNAs through cleavage or independent promoters (Chao *et al*., 2012; Miyakoshi *et al*., 2015). This method was initially used in the global mapping of transcripts in *Helicobacter pylori* and identified around 60 novel sRNA transcripts (Sharma *et al*., 2010).

1.4.4 RIP-seq

In RIP-seq (RNA Immunoprecipitation-seq), an RNA-binding protein (RBP) such as Hfq, is immunoprecipitated on streptavidin beads after cell lysis. RNA associated with protein in weak RNA: protein interactions were recovered and analyzed by deep sequencing. RIP-seq allows the rapid characterization of RNA regulon and many Hfq associated small RNAs have been characterized in *E. coli* and *Salmonella* (Sittka *et al*., 2008; Zhang *et al*., 2003).

1.4.5 CLIP-seq

Cross-linking Immuno-precipitation is a modification of the RIP-seq approach. In CLIP-seq, cells are subjected to UV cross-linking before cell lysis. In comparison to RIP-seq, protein and associated RNAs are covalently crosslinked which increases the specificity between RBPs and associated RNAs *in vivo* (Barquist and Vogel, 2015). This strategy yields high-resolution data on the RNA binding sites of the RBPs and was employed for the determination of transcriptome-wide binding sites of Hfq and CsrA in *Salmonella* (Holmqvist *et al*., 2016).

1.4.6 Grad-seq

Another advanced approach includes gradient profiling by sequencing (Grad-seq) which uses RNA–protein interaction analyses to provide the global RNA landscape by partitioning all cellular transcripts. These transcripts and associated proteins were identified by transcriptomic and proteomic analyses (Smirnov *et al*., 2016). This approach was used for mapping the regulatory role of novel RNA binding protein ProQ and identified its role in sRNA regulation.

1.5 Identification of sRNA targets for discovery of sRNA function

Functional characterization of sRNA requires the full understanding of its interacting partners. Various computational and experimental tools are available for the prediction and validation of target mRNAs at transcript levels and protein levels. The sRNA–mRNA

interaction depends on Watson–Crick and G–U base pairings between complementary sequences stretch in the two RNAs. This interaction involves the noncontiguous base pairing between 8-50 nucleotides, alternative secondary structure, and presence of interacting protein such as Hfq (Dos Santos *et al*., 2019) and ProQ (Smirnov *et al*., 2016).

1.5.1 Computational Approaches

Computational programs use for the prediction of sRNA targets, mainly depend on the thermodynamic model of base pairing between two RNAs.

1.5.1.1 IntaRNA

IntaRNA (Interacting RNAs) is *in silico* modeling tool for efficient prediction of RNA–RNA interactions. The IntaRNA algorithm calculates the combined energy score of the RNA: RNA interaction as the sum of free energy required for making the interaction sites accessible and the free energy required for hybridization. It also enables the details of minimum energy profiles of interacting RNAs, which helps in the study of alternative RNA-RNA interaction (Mann *et al*., 2017). This program was also used as the guide for mutational analysis between the interacting RNAs for wet-lab studies and used previously for target validation (Beisel *et al.*, 2012; Beisel and Storz, 2011; Robledo *et al*., 2015).

1.5.1.2 CopraRNA

CopraRNA (Comparative prediction algorithm for small RNA targets) is a tool that combined the sRNA target prediction in a set of organisms. CopraRNA computes the whole-genome target prediction for conserved sRNA sequence by integrating the phylogenetic information from the diverse organisms and utilizes the IntaRNA prediction algorithm (Wright *et al*., 2014). CopraRNA has been applied for the target prediction in many enterobacterial sRNAs, non enterobacterial systems and serve as the comprehensive benchmark tool for target identification (Durand *et al.*, 2015; Lokstein *et al.*, 2014; Overlöper *et al*., 2014; Pain *et al.*, 2015; Robledo *et al*., 2015).

1.5.2 Experimental approaches

Diverse experimental approaches including proteomics and transcriptomics have been employed previously to explore the interacting targets of small RNA. Construction of sRNA overexpression and deletion strains to modify the copy number of sRNA was used in the early discovery of sRNA targets. Effect of differential expression of sRNA on

whole-cell proteins was analyzed in single dimension SDS-PAGE and two dimensional SDS-PAGE. Such approaches unveiled the down-regulation of many ABC transporters by GcvB sRNA in *E. coli* (Urbanowski *et al*., 2000), OmpA by VrrA in *Vibrio* (Song *et al*., 2008) and LamB expression by MicA in *Salmonella* (Bossi and Figueroa-Bossi, 2007).

1.5.2.1 Pulse-expression

Pulse-expression of sRNA in combination with transcriptomic approach was more significant in sRNA identification, where constitutive expression of sRNA becomes toxic to the cells. In this method, sRNAs were overexpressed for a shorter period and altered expression of mRNAs subjected to microarray analysis. The pulse-expression strategy revealed the regulatory role of *E. coli* sRNA RyhB in iron homeostasis (Massé *et al*., 2005) and OmrAB in the regulation of outer membrane protein-coding mRNAs (Guillier and Gottesman, 2006).

1.5.2.2 RIL Seq

RIL-seq (RNA interaction by ligation and sequencing) is a hybrid of an experimental and computational methodology for identifying Hfq-bound sRNA-mRNA targets interaction in bacteria. The methodology involves the co-immuno-precipitation of Hfq bound RNAs followed by ligation and deep sequencing. This approach was used for the identification of Hfq associated targets under different stress conditions in *E. coli* which uncovered the hundreds of new Hfq associated sRNA-mRNA targets as well as novel small RNAs involved in global cellular processes (Melamed *et al*., 2018, 2016).

1.5.2.3 MAPS

RNA binding protein Hfq is used as bait for the identification of sRNA targets in many previously described approaches including the most recent RIL-seq approach (Melamed *et al*., 2016). However, the role of Hfq is limited in sRNA-mediated gene regulation in *S. aureus* as compared to *E. coli* and *Salmonella* (Bohn *et al*., 2007). This limitation is overcome by the discovery of the MAPS approach for target identification without the requirement of RNA chaperone protein. MS2-affinity purification coupled with RNA-Seq (MAPS) approach is a recent alternative, in which MS2 RNA aptamer is used to tag the sRNA. Due to high affinity towards the MS2 coat proteins; sRNA associated targets RNAs can be co-purified, sequenced and identified (Lalaouna *et al*., 2018, 2017). MAPS

approach has already been used for targetome analysis of several sRNAs in *Staphylococcus aureus* (Bronesky *et al*., 2019; Lalaouna *et al*., 2019).

1.6 *Staphylococcus aureus*, an opportunistic pathogen

Staphylococcus aureus is an exceptionally adaptable versatile gram-positive pathogen that causes many serious nosocomial and community-acquired infections such as skin infections, sepsis, endocarditis, pneumonia and osteomyelitis (Gnanamani *et al*., 2017). *S. aureus* expresses a battery of virulence factors including cell wall associated adhesins and secreted toxins that promote its adaptation and colonization under adverse stress conditions and intracellular survival, during pathogeneses (reviewed in Novick, 2003; Gordon and Lowy, 2008).

1.7 Virulence factors of *S. aureus*

Bacterial adherence and interaction with the host cell is the primary step in colonization and *S. aureus* utilizes an extensive range of strategies to adhere to human extracellular matrix components and serum proteins. These abilities to colonization are attributed to the regulated expression of multifactorial surface-associated adhesins that facilitates its interaction and dissemination into the host cell and subsequent secretion of various exotoxins that damage the host cell and helps in evading the immune system (Figure 6) (Simon R Clarke *et al*., 2006).

1.7.1 Cell wall associated virulence factors

Surface proteins expressed in *S. aureus* are covalently attached to the peptidoglycan and known as the microbial surface component recognizing adhesive matrix molecules (MSCRAMMs). MSCRAMMs are important virulent determinants that help *S. aureus* to colonize diverse host matrices and are also involved in biofilm formation (Foster *et al*., 2014).

1.7.1.1 Staphylococcal protein A (SpA)

SpA is a 42-kDa surface protein encoded by the *spa* gene is a major virulence factor expressed in *S. aureus* (Foster *et al*., 2014). SpA binds to the Fc region of the immunoglobulin that protects the *S. aureus* from opsono-phagocytic killing (Peterson *et al*., 1977). In addition to immune evasion, SpA acts as an important adhesin and superantigen that binds to the surface of B-lymphocytes. This complex stimulates

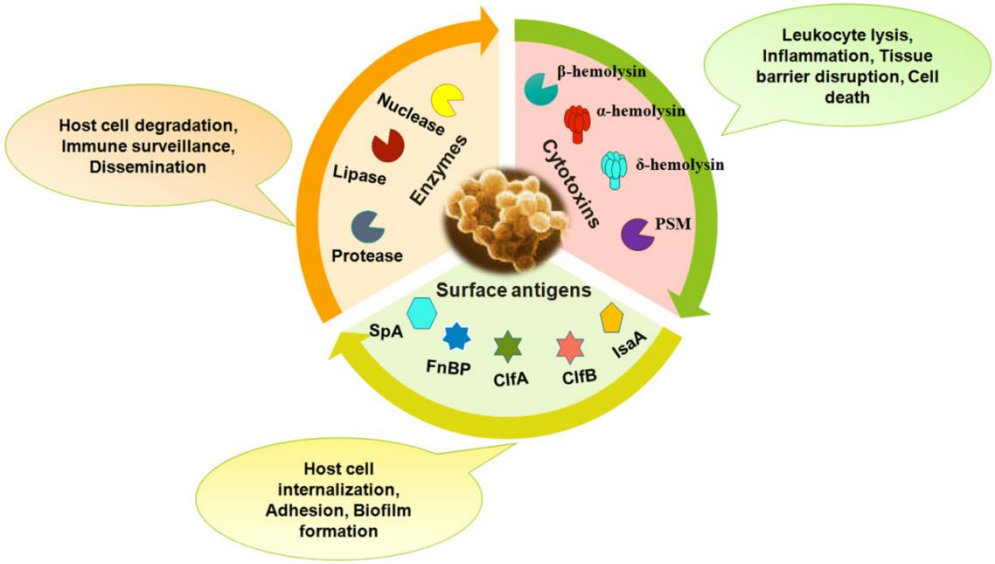

Figure 6: Virulence determinants of *Staphylococcus aureus*

S. aureus expressed many surface-associated virulence factors that function as an immunogen, promote its interaction with the diverse host matrix, and facilitates colonization (FnBPs, ClfA, ClfB, IsaA). Dissemination of host tissue, immune evasion, and cell lysis is mediated by numerous secreted virulence factors. Pore-forming toxins targeting specific surface receptors on host cells, cause inflammation, tissue disruption, and cell death (Hemolysins, PSMs, Leukocidin). Numerous exotoxins secreted by *S. aureus* include proteases, nucleases, and lipases that lead to host cell degradation, processing of many proteins and thwarting immune defense mechanism.

apoptosis, reduces antibody secreting B-cell and crippled the humoral immune response (Goodyear and Silverman, 2003).

1.7.1.2 Fibronectin binding proteins

Fibronectin is a multi-domain glycoprotein present in extracellular matrices of various cell types and plays a role in the regulation of cell migration, tissue repair, and cell adhesion. *S. aureus* expresses two fibronectin (Fn)-binding proteins, FnBPA and FnBPB (Fn-binding proteins A and B), which binds with fibrinogen, elastin, plasminogen and involves in adhesion and cell invasion (Testoni *et al.*, 2011). FnBPs are important virulence factors involved in skin abscess infection such as atopic dermatitis (Cho *et al.*, 2001; Kwiecinski *et al.*, 2014), endocarditis (Que *et al.*, 2005), bacteremia, sepsis (Shinji *et al.*, 2011) and play a crucial role in staphylococcal pneumonia by colonizing the airways epithelium (Mongodin *et al.*, 2002).

1.7.1.3 Clumping Factors A and B

ClfA is a fibrinogen-binding adhesin that interacts with fibrinogen and promotes its attachment under high physiological shear stress, and enhances the colonization ability of *S. aureus* (Herman-bausier *et al*., 2018). ClfA increases *S. aureus* adherence to plasma protein-coated biomaterials and facilitates the colonization, biofilm formation, inhibits phagocytosis and also involved in septic arthritis (Higgins *et al*., 2006; Josefsson *et al*., 2001; Vaudaux *et al*., 1995). ClfB promotes the clumping of cells, adherence to fibrinogen coated surface and facilitates its attachment to the anterior nares through a high-affinity interaction with cornified envelope, which promotes nasal colonization (Entenza *et al*., 2000; Lacey *et al*., 2019; Lopes *et al*., 2019). ClfB regulates biofilm formation, skin infection and emerged as a promising vaccine target against *S. aureus* (Abraham and Jefferson, 2012; Lacey *et al*., 2019).

1.7.1.4 Immunodominant staphylococcal antigen A (IsaA)

IsaA is a highly antigenic protein with a lytic trans-glycosylase activity that promotes autolysis and regulates clumping of cells (Sakata *et al*., 2005). IsaA is mainly localized at the septal region of dividing cells and also excreted in the extracellular environment (Simon R. Clarke *et al*., 2006; Lorenz *et al*., 2000; Sakata *et al*., 2005). IsaA is involved in the biofilm formation, β-lactam resistance, sepsis and monoclonal antibodies targeting IsaA were found effective against the staphylococcal infections, promoting IsaA as a promising vaccine candidate for immunotherapy (Lorenz *et al*., 2011; Oesterreich *et al*., 2014; van den Berg *et al*., 2015).

1.7.1.5 Autolysins

Autolysins are peptidoglycan hydrolases, required for the degradation, turnover, and maturation of peptidoglycan and play an essential role in cell division and daughter cell separation (Biswas *et al*., 2006; Vollmer *et al*., 2008). *S. aureus* expresses many cell wall-associated and secreted autolysins; yet the most prominent autolysin that it expresses is AtlA (138 kDa), a bifunctional protein with amidase (AM) and a glucosaminidase (GL) domain. Endo-β-*N*-acetylglucosaminidase (GL, 51 kDa) and *N*-acetyl muramyl-L-alanine amidase (AM, 62 kDa) are proteolytic products of bifunctional autolysin AtlA and play an important role in dispersing cell clusters of *S. aureus* (Bose *et al*., 2012; Sugai *et al*., 1995). Autolysins are reported to function as adhesins and immunomodulatory molecules that promote the interaction of *S. aureus* to the diverse host matrix, host cell

internalization, platelet activation and aggregation, and biofilm formation (Binsker *et al*., 2018; Hirschhausen *et al*., 2010; Porayath *et al*., 2018).

1.7.2 Secreted virulence factors

Staphylococcus aureus secretes various virulence factors including secreted toxins, exoenzymes that cause tissue damage, facilitate spreading, nutrients uptake, and manipulate the host's innate and adaptive immune responses. *S. aureus* encodes various pore-forming cytotoxins, such hemolysins, Panton valentine leukocidin and exoenzymes including protease and nucleases, which play important role in infection.

1.7.2.1 Hemolysins

S. aureus encodes four hemolysins including α, β, γ, and δ. α-hemolysin is the best-characterized pore-forming toxin of *S. aureus* encoded by the *hla* gene, secreted as a monomer which oligomerizes upon binding to the target cell membrane to form a heptameric pre-pore complex followed by the formation of a mature β-barrel transmembrane pore (Figure 7). The mature β-barrel transmembrane pore complex is a 14-A° diameter channel that enables the transport of K^+ and Ca^{2+} ions and causes cell lysis (Bhakdil and Tranum-Jensen2, 1991; Divyakolu *et al*., 2019).

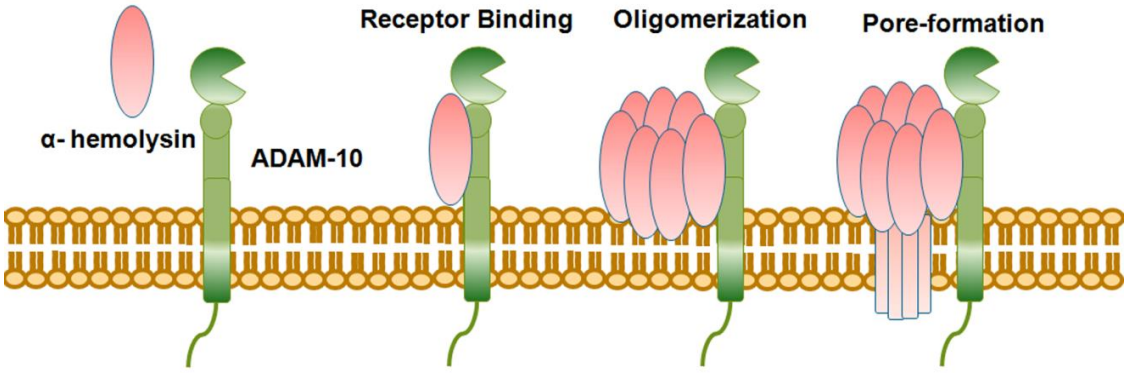

Figure 7: Pore formation of α-hemolysin

α-hemolysin expressed as monomer that binds to specific cell surface receptor ADAM-10 with high affinity. This interaction initiates the oligomerization of α-hemolysin and subsequent formation of active β-barrel pore structure. This cytolytic pore formation further triggers the sequence of events leading to cell lysis.

α-hemolysin is predominantly active against rabbit erythrocytes and the specificity of interaction is associated with its binding to the specific host cell receptor ADAM10

(Wilke and Wardenburg, 2010). α-Hemolysin is involved in various *S. aureus* infections including skin infection, septic arthritis, abscess, and pneumonia (O'Callaghan *et al.*, 1997; Tam and Torres, 2018).

β-hemolysin is a non-pore forming secreted toxin with sphingomyelinase activity that hydrolyzes sphingomyelin, a plasma membrane lipid and highly toxic for sheep erythrocytes (O'Callaghan *et al.*, 1997). It enhances the host cell susceptibility to other toxins such as α-hemolysin and plays a role in the recurrence of osteomyelitis, respiratory infections and additionally regulates the biofilm formation in *S. aureus* (Divyakolu *et al.*, 2019; Huseby *et al.*, 2007). γ-hemolysins (HlgAB, HlgCB), are pore-forming bi-component toxins that interact with CXCR1, CXCR2, CCR2, Duffy antigen receptor for chemokine and disrupts the human erythrocytes, neutrophils, monocytes, and macrophages (Cooney *et al.*, 1993; Spaan *et al.*, 2014). δ- hemolysin is a 26 amino acids small amphipathic peptide encoded by the *hld* gene located within the RNAIII locus of accessory gene regulator (*agr*) (Novick, 2003). δ-hemolysin disrupts the variety of host cells by three different mechanisms: first, it interacts with the cell surface and forms transmembrane pore; second, destabilizes the membrane at high concentrations; and third, it dissolves the membrane by acting as a surfactant. It causes a rapid influx of Ca^{2+}, stimulates free radicals production in human granulocytes and its activity is enhanced by β-hemolysin (Scheifele and Bjornson, 1988; Schmitz *et al.*, 1997). δ-hemolysin plays a significant role in colonization and skin infection, atopic dermatitis by causing mast cell degranulation and also induces the release of pro-inflammatory cytokines from keratinocytes (Böcher *et al.*, 2009; Nakamura *et al.*, 2013).

1.7.2.2 Phenol soluble modulins

Phenol soluble modulins (PSMs) are small peptides toxins that belong to a family of amphipathic peptides expresses in *S. aureus*. PSMs have a high affinity to lysing eukaryotic cell membrane and contribute to neutrophil lysis after phagocytosis (Wang *et al.*, 2007). It acts as an intracellular toxin that helps *S. aureus* invasion and escapes by killing osteoblasts (Rasigade *et al.*, 2013). PSMs are also involved in biofilm formation, cell lysis and immune modulation, a pathogenesis mechanism that is immensely important for the high virulence of *S. aureus* strains (Cheung *et al.*, 2012; Periasamy *et al.*, 2012; Rasigade *et al.*, 2013; Wang *et al.*, 2007).

1.7.2.3 Proteases

Staphylococcus aureus encodes around 12 different secreted proteases belongs to three families: metalloproteases, cysteine proteases, and serine proteases. Aurolysin is the example of metalloprotease that target diverse substrate including *S. aureus* proteins and host protein, involved in immune defense and phagocytosis (Burlak *et al*., 2007; Nickerson *et al*., 2007). SspA and SspB are serine protease and Staphopain A (ScpA) is an example of cysteine protease that modulates the biofilm formation by reducing bacterial adhesion and enhances bacterial dissemination (Karlsson *et al*., 2001; Loughran *et al*., 2014; McGavin *et al*., 1997).

1.8 Regulation of virulence factors in *S. aureus*

Staphylococcus aureus is an exceptionally versatile organism and causes a diverse spectrum of human diseases by coordinately expressing secreted and surface-associated virulence factors at different stages of infection. The expression of virulence factors in *S. aureus* is regulated at distinct levels by various two-component systems (TCS) (Bronner *et al*., 2004), sigma factors (Roberts *et al*., 2017), RNA binding proteins and small regulatory RNAs (Holmqvist and Vogel, 2018; Jørgensen *et al*., 2020).

1.8.1 Two Component System

Successful adaptation of bacteria in different environments is mediated by a two-component regulatory system that senses external stimuli and responds by changing gene expression. The acquisition of external signal is sensed by the surface-associated sensor histidine kinase (HK) and results in auto-phosphorylation that transfers the phosphoryl group to its cognate response regulator protein. Phosphorylation of response regulator protein triggers the conformational changes that increase its affinity towards target DNA domains and in turn regulates the expression of multiple genes (Casino *et al*., 2010). *S. aureus* encodes 16 different TCSs involved in sensing diverse environmental stimuli such as pH, stress, nutrient, and responses by fine-tuning the gene expression (Table 1) (Haag and Bagnoli, 2016). These TCS listed below play a significant role in the regulation of virulence gene expression (AgrCA, SaeRS), antibiotic resistance (VraSR, GraXSR, and BraRS), and cell wall metabolism (WalKR).

Table 1: Two component regulatory system of *S. aureus*

TCS	Major function[a]	Positively regulation	Negative regulation
WalKR	Cell wall metabolism	*alt, isaA, ssaA, lytM, sceD, hla, hlb, hlgA/hlgCB, lukSF, chp, sbi, scn, eap, efb, embp, fnbA, fnbB, atl, coa, vWfbp, splA, splB, splC, splD, splE, splF,* saePQRS	*spa, sarS, sarT*
AgrAC	Quorum sensing, cell wall and secreted protein synthesis	*eta, etb, hla, hlb, hld, hlgA/hlgCB, lukED, luk-PV, seb, sec, tst, capH, atl, aur, geh, hysA, nuc, sak, scpA, scpB, splA, splB, splC, splD, splE, splF, sspA, agrA, RNAIII, arlRS, saePQRS*	*spa, ssl5, ssl7, ssl8, ssl9, fnbA, fnbB, coa*
ArlRS	Adhesion, autolysis, multidrug resistance and virulence genes	*sdrC, sdrD, sdrE, rot, agrA, mgrA, capH*	*spa, coa, geh, splA, splB, splC, splD, splE, splF, sspA, isdA, isdB*
GraRS	CAMP sensing	*hlb, hld, hlgA/hlgCB, lukM, lukSF, capH, chp, sbi, clfB, efb, icaABCD, sdrC, sdrD, sdrE, atl, coa, geh, mntABC, agrA, lytSR, mgrA, perR, rot, sarA, sarS, sarX*	
SrrAB	Aerobic and anaerobic respiration	*tst, spa, and icaR*	*tst, srr*
SaeRS	Secreted factors involved in immune evasion	*eta, etb, hla, hlb, hld, hlgA/hlgCB, lukED, lukM, lukSF, seb, tst, chp, sbi, scn, spa, ssl5, ssl7, ssl8, ssl9, eap, efb, embp, fnbA, fnbB, coa, geh, nuc, sak, splA, splB, splC, splD, splE, splF, isdA, isdB, fur*	*capH, aur*

[a]TCS involved in gene regulation, reviewed in (Haag and Bagnoli, 2016).

1.8.1.1 AgrCA

S. aureus accessory gene regulator (agr) *agr* system is an example of the global gene regulators, controlling the expression of virulence genes (Figure 8). The *agr* locus contains two promoters P2 and P3 encoding, RNAII and RNAIII transcript respectively (Novick *et al.*, 1993). The RNAII transcript comprises an operon of the four quorum-sensing genes *agr*BDCA, and RNAIII is a major effector molecule that encodes the small peptide, Hld. AgrC is a dimeric, membrane-associated sensor kinase that upon interacting with quorum sensing stimulant AIP, undergoes autophosphorylation followed by the transfer of phosphate to the AgrA, a response regulator protein (Cisar *et al.*, 2009).

AgrA directly regulates the expression of RNAII and RNAIII transcripts by binding with the direct repeats within P2 and P3 promoters (Ji *et al*., 1995; Koenig *et al*., 2004; Srivastava *et al*., 2014). AgrA also interacts with the promoters of the alpha- and beta-PSM-encoding operons (Queck *et al*., 2008). The main effector of *agr* regulon is RNAIII, which upregulates the expression of various secreted toxins during high cell density and downregulates the expression of surface-associated proteins (Boisset *et al*., 2007; Singh and Ray, 2014).

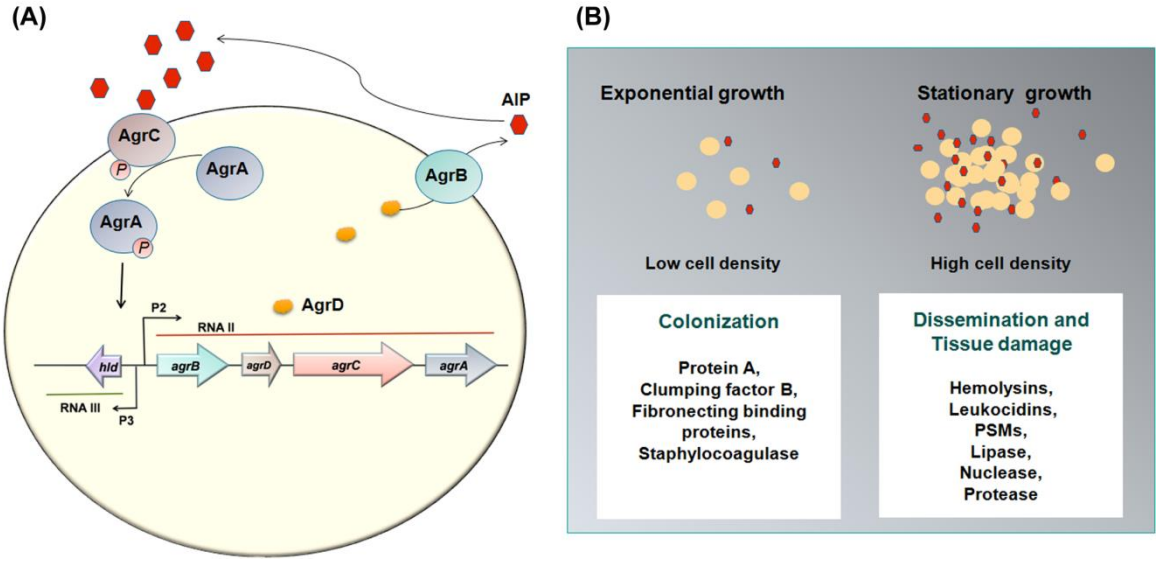

Figure 8**: AgrCA: The quorum sensing TCS regulating virulence gene expression in *S. aureus***
(A) The quorum-sensing mechanism in *S. aureus*. AgrD, a small peptide, is processed into an active pheromone called an autoinducing peptide (AIP) by AgrB, a transmembrane endopeptidase. At high cell density, AIP binds AgrC, a receptor histidine kinase, resulting in its autophosphorylation and further activation (phosphor transfer) of AgrA. AgrA induces the transcription of RNAII and RNAIII at P2 and P3 promoters respectively. (B) The quorum-sensing signaling is inactive at low cell densities, which favors the expression of surface virulence determinants that facilitates the colonization of *S. aureus*. At high cell density activation of the Agr system intensify the expression of RNAIII that represses the expression of colonization factors and induces the expression of secreted toxins to effectuate the host cell dissemination and tissue disruption.

1.8.1.2 WalKR

Bacterial TCSs that sense and respond according to change in environmental condition is an important factor for determining its evolutionary success. Among the 16 TCSs presents in *S. aureus*, WalKR (aka YycFG) is highly conserved between low GC gram-positive bacteria. WalKR is essential for cell viability during bacterial cell division but

non-essential during grown arrested conditions such as starvation and desiccation (Villanueva *et al.*, 2018). WalKR regulates various genes associated with cell wall metabolism and peptidoglycan biosynthesis, through cross-bridge hydrolysis and turnover (Dubrac *et al.*, 2007). Mutations in WalKR are reported to reduce the susceptibility of *S. aureus* to vancomycin and daptomycin antibiotics (Bayer *et al.*, 2013; Hu *et al.*, 2015). WalKR positively regulates the biofilm formation, promotes the host inflammatory response and plays an important role in *S. aureus* virulence (Delauné *et al.*, 2012). WalK a histidine kinase and WalR a response regulator protein differentially regulates the autolysis in *S. aureus* (Zheng *et al.*, 2015). Expression of genes encoding autolysins specifically *atlA, lytM, ssaA* and *isaA* are directly regulated by WalKR through binding of WalR to the promoter region (Figure 9) (Dubrac *et al.*, 2007; Dubrac and Msadek, 2004).

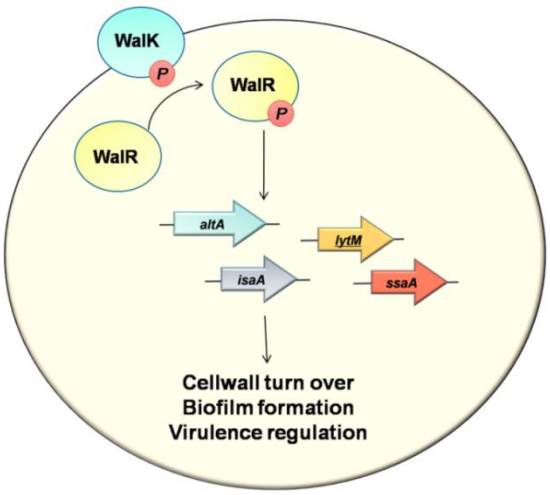

Figure 9: WalKR: The essential TCS control the cell wall metabolism in *S. aureus*

The activation of WalK histidine kinase through autophosphorylation, in turn, activates the response regulator protein WalR. WalR binds to a consensus recognition sequence in the upstream promoter region of many genes involved in cell wall metabolism (*atlA, isaA, lytM, ssaA*). WalKR positively influences autolytic activity and biofilm formation in *S. aureus*.

1.8.2 SarA

In addition to TCS, *S. aureus* employs several DNA-binding proteins that regulate the expression of virulence factors. SarA is one of the best-characterized DNA binding proteins expressed in *S. aureus*. SarA is a pleiotropic transcriptional regulator that directly binds to the promoter regions of secreted and surface virulence factors, such as *hla*, *spa* and also interacts with *agr* (Chien *et al.*, 1999; Chien and Cheung, 1998). SarA enhances

the expression of secreted toxins and also exerts its regulation post-transcriptionally, by modulating the mRNA turnover (Morrison *et al*., 2012). Besides regulating the expression of target proteins SarA also represses the expression of small RNA SprC and srn_9340 involved in *S. aureus* virulence (Le Pabic *et al*., 2015; Mauro *et al*., 2016).

1.8.3 Alternative sigma factors, SigB

Sigma factors required during transcription initiation confer the specificity of RNA polymerase and alternative sigma factors control the gene expression under different environmental conditions. Sigma B (SigB) is associated with numerous environmental stress conditions such as acid, ethanol, heat and modulates the gene expression in response to conditions encountered by bacteria (Chan *et al*., 1998; Ferreira *et al*., 2001; Rodriguez Ayala *et al*., 2020). SigB directly or indirectly regulates the around 200 genes in *S. aureus* that are involved in biofilm formation, virulence, antibiotic resistance, host cell internalization (reviewed in Guldimann *et al*., 2016). SigB enhanced bacterial adherence by positively modulating the expression of clumping factor A and fibrinogen binding protein A (Bischoff *et al*., 2004; Entenza *et al*., 2005). SigB regulates the expression of secreted toxin gamma hemolysin, and plays an important role in *S. aureus* adaptation to the lung environment (Ishii *et al*., 2014). SigB enhances the *S. aureus* adaptation during the long course of the infection by increasing the intracellular persistence (Tuchscherr *et al*., 2015).

1.9 sRNA discovery in *Staphylococcus aureus*

Over the last two decades, sRNAs have been reported to play a significant role in gene regulation under different stress conditions and adaptive processes in many bacteria (Jørgensen *et al*., 2020; Papenfort and Vogel, 2010). The first regulatory RNA discovered in *S. aureus* was RNAIII, a multifunctional small RNA, involved in the quorum-sensing system (Novick *et al*., 1993). Later on, various studies reported the diversity of sRNAs encoded from the *S. aureus* genome by employing a computational prediction approach coupled with expression analysis (Geissmann *et al*., 2009; Marchais *et al*., 2009; Pichon and Felden, 2005), microarrays (Anderson *et al*., 2006; Roberts *et al*., 2006), tilling arrays (Lasa *et al*., 2011), cDNA cloning and sequencing (Abu-Qatouseh *et al*., 2010), illumina sequencing (Beaume *et al*., 2010). The timeline of sRNAs discovered in *S. aureus* is depicted in Figure 10. The expression of many reported sRNAs has been validated at transcriptional levels using northern blot and RT-PCR. These studies reveal that *S. aureus*

encodes more than 500 sRNAs, located randomly in the core genome as well as on mobile genetic elements such as the pathogenicity island (Beaume *et al*., 2010; Bohn *et al*., 2010; Geissmann *et al*., 2009; Lasa *et al*., 2011; Roberts *et al*., 2006).

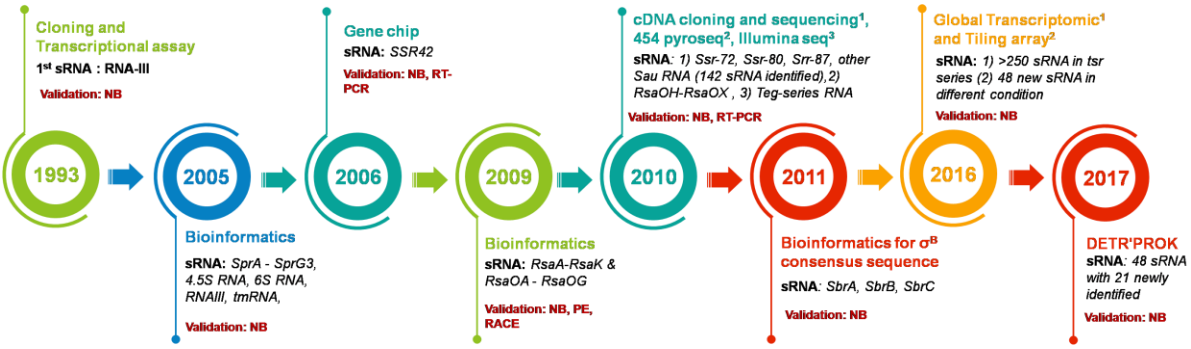

Figure 10: Timeline of sRNA discovery in *Staphylococcus aureus*
The sRNA discovery reported for *S. aureus* dates back to 1993. In the last two decades, multiple sRNAs have been discovered and validated experimentally using northern blot (NB) and real-time PCR (RT-PCR).

1.10 sRNA regulon in *S. aureus*

A large number of small RNAs expressed in *S. aureus* mediate fine-tuning of complex regulatory circuits and genes, involved in biofilm formation, antibiotic resistance, cell wall metabolism, and host-pathogen interaction (reviewed in Fechter *et al*., 2014). The sRNA encoded in *S. aureus* are scattered in the core genome and also in a specific region called Pathogenicity Island (Figure 11 and Table 2).

1.10.1 sRNAs involved in metabolic regulation

Many of the sRNAs encoded by the core genome are conserved in *Staphylococcus* spp. and play a central role as metabolic regulators. RsaE is one of the well-characterized sRNAs, involved in the regulation of several metabolic pathways in *S. aureus*. RsaE is a highly conserved sRNA derived from the common gram-positive ancestor and expressed in *Micrococcus* and *Bacillus*. In *S. aureus*, RsaE regulates the expression of many enzymes involved in lipid, purine, and carbohydrate metabolism, peptide transport, and the TCA cycle. The expression of amino acid and peptide transporter protein *oppB*, succinyl-CoA synthetase *sucD* and arginase *rocF* is controlled by RsaE by interacting with ribosome binding sequence to inhibit mRNA translation (Geissmann *et al*., 2009; Rochat *et al*., 2018). Through its diverse regulatory network, RsaE acts as a metabolic

switch that contributes to cell reprogramming by promoting energy metabolism, amino acid transport in response to nutrient availability and also plays an indirect role in *S. aureus* virulence regulation (Bohn *et al.*, 2010; Somerville and Proctor, 2009). Another example of sRNA involved in metabolic regulation is the RsaI. RsaI is also conserved and expressed in *S. aureus* and controls the signaling pathway responding to glucose uptake. Targetome of RsaI analyzed using the MAPS approach revealed multiple target mRNAs that are involved in sugar metabolism, arginine catabolism, biofilm formation as well as other sRNAs as interacting partners. RsaI interacts with RsaG, RsaD and RsaE sRNA which links its regulatory role in NO detoxification and adaptation to diverse host environment encounters during infection (Bronesky *et al.*, 2019).

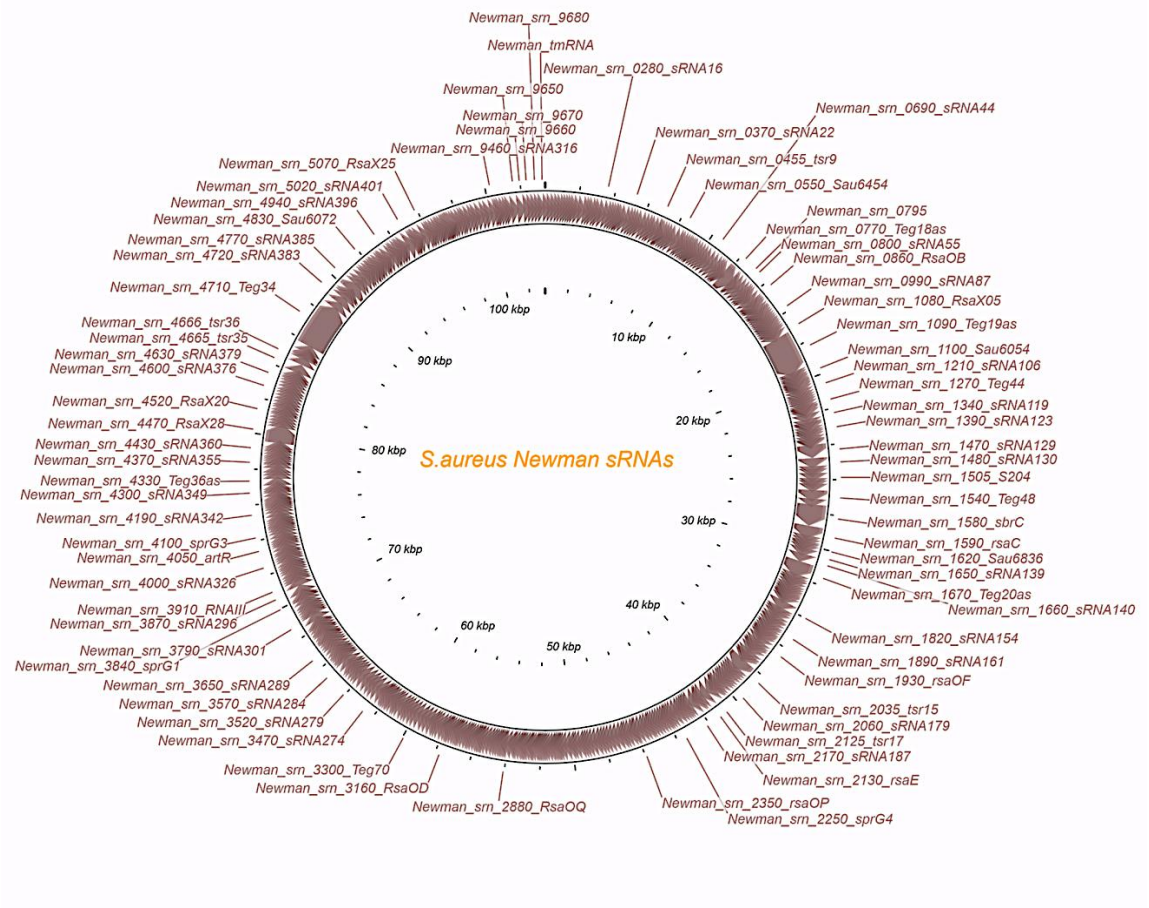

Figure 11: Map of sRNA expressed in *S. aureus* Newman

Small RNAs are widespread in the *S. aureus* genome, and are encoded from discrete genomic loci including intergenic region, *cis* regulatory elements, and pathogenicity islands. In *S. aureus* strain Newman, more than 500 sRNA have been identified and reported in the SRD database. The genomic location of each sRNA was retrieved and used to construct the map of sRNAs in the *S. aureus* Newman genome using CG-viewer.

Table 2: **Functionally characterized sRNAs in *S. aureus***

sRNA	Role	Targets	References
sRNA encoded from core genome			
RNAIII	Regulates the expression of surface and excreted toxins, biofilm formation	*hla, coa, rot, lytM, sbi*	(Boisset *et al*., 2007; Chabelskaya *et al*., 2014; Chevalier *et al*., 2010; Chunhua *et al*., 2012; Morfeldt *et al*., 1995)
RsaE	Carbon metabolism, biofilm formation	*opp-3B/opp-3A, sucD, rocF*	(Geissmann *et al*., 2009; Rochat *et al*., 2018)
RsaA	Immune evasion, biofilm formation, capsule formation, cell wall metabolism	*mgrA, flr, ssaA*	(Tomasini *et al*., 2017)
RsaI	Regulation of gene expression under glucose limited condition, biofilm formation	*glc*U_2, *fn3K*, *tre*B, *HG001_01242, HG001_02520,* RsaE, RsaG	(Bronesky *et al*., 2019)
RsaD	Carbon overflow metabolism, contributing to stationary phase cell death during exposure to weak acid stress	*alsS*	(Augagneur *et al*., 2020)
ArtR	Involved in virulence regulation by activating alpha toxin expression	*sarT*	(Xue *et al*., 2014)
Teg49	Controls virulence, oxidative stress and biofilm formation	*sarA*	(Kim *et al*., 2014)
Teg41	Toxins and virulence regulation	*α*PSM	(Zapf *et al*., 2019)
sRNA encoded from pathogenicity island			
SprX	Virulence regulation, biofilm formation, antibiotic resistance	*hld, clfB, spoVG, ecb*	(Eyraud *et al*., 2014; Ivain *et al*., 2017; Kathirvel *et al*., 2016)
SprC	Attenuates bacterial virulence and host cell phagocytosis	*atl*	(Le Pabic *et al*., 2015)
SprD	Negatively regulates the expression of the immune-evasion molecule	*sbi*	(Chabelskaya *et al*., 2010)
SprF1	RNA antitoxin, promotes persister cell formation and attenuates the translation	*ribosomes*	(Pinel-Marie *et al*., 2021)

1.10.2 sRNA linking quorum sensing and virulence regulation

Quorum sensing is mainly regulated by *agr* system in *S. aureus*, which also encodes the major regulatory sRNA RNAIII. RNAIII functions as an intracellular effector molecule and coordinates the expression of virulence factors (Novick, 2003; Novick *et al*., 1993). RNAIII is a bifunctional RNA, encodes a delta hemolysin, which also contributes towards the virulence of *S. aureus*. The 3' UTR of RNAIII contains a conserved C-rich sequence that interacts with many target mRNAs at the ribosome binding site, and prevents mRNA translation (Boisset *et al*., 2007; Chevalier *et al*., 2010; Huntzinger *et al*., 2005). RNAIII enhances the expression of secreted alpha-hemolysin by base pairing at the 5' UTR of *hla* mRNA that leads to the unmasking of the ribosome binding site to activate its translation. RNAIII not only regulates the virulence but also involved in cell wall metabolism by repressing the many cell wall hydrolases and amidases involved in the peptidoglycan metabolism and contribute to the cell wall integrity at high cell density (Boisset *et al*., 2007; Chunhua *et al*., 2012; Dunman *et al*., 2001; Lioliou *et al*., 2016).

1.10.3 sRNAs expressed from the pathogenicity island

Pathogenicity Islands are horizontally acquired mobile genetic elements that play a role in bacterial evolution, present in the several pathogenic strains, and often absent in nonpathogenic strain. Pathogenicity Island discovered in *S. aureus* encodes several sRNAs including SprA-SprG and Teg series of sRNAs, virulence factors and play an important role in pathogenicity regulation (Beaume *et al*., 2010; Bohn *et al*., 2010; Pichon and Felden, 2005). Despite the long list of sRNA reported, only a few sRNAs have been characterized functionally. These include SprD (Chabelskaya *et al*., 2010), SprC (Le Pabic *et al*., 2015), SprX (Eyraud *et al*., 2014; Kathirvel *et al*., 2016), and SprA (Sayed *et al*., 2012). SprD, the second sRNA that emerged as a virulence regulator after RNAIII, is one of the players in the multifactorial regulation of *sbi* mRNA, which encodes an immune evasion molecule. SprD interacts with *sbi* mRNA, at the ribosome binding site and inhibits translation initiation. Sbi mRNA is encoded from the core genome, suggesting that SprD mediates the crosstalk between pathogenicity island and core genome. Both RNAIII and SprD repress the *sbi* mRNA translation via the same mechanism during the stationary phase (Chabelskaya *et al*., 2010). The involvement of two sRNAs in the regulation of one common target mRNA is not limited to *S. aureus*, as it has been well documented in other bacteria (Gogol *et al*., 2011). SprC, encoded from the pathogenicity island attenuates the *S. aureus* pathogenicity by suppressing the

virulence, spread, and host cell phagocytosis. This inhibitory control is partially mediated by its direct interaction with major autolysin Atl (Le Pabic *et al.*, 2015).