# 1 INTRODUCTION AND REVIEW OF LITERATURE

The ability to constantly sense and adapt to environmental changes is essential for all organisms to maintain its cellular functions. Organisms show a remarkable regulatory flexibility that allows them to thrive under different external conditions and to survive harsh environments. Regulation of gene expression is pivotal for optimal energy management and adapted metabolic responses to fluctuating environmental conditions and stresses, and is achieved by complex mechanisms regulated by intricate networks of multiple regulators.

## 1.1 Regulation of gene expression in bacteria

Pathogenic bacteria continuously interact with the host during an infection and require rapid regulatory circuits to survive and adapt to a myriad of environmental stress such as temperature and pH shifts, nutrient availability, antimicrobials or dangerous chemical reactive species presented by the host. This exquisite level of adaptation requires modulation of their physiological and metabolic features (Antoine et al., 2021; González Plaza, 2020). Host invasion, colonization and survival of a pathogen despite the host's immune responses and antimicrobial agents during infection require a tight coordination of gene expression. A Group of transcription factors, regulatory proteins and regulatory RNAs that function alone or in combination, to up- or down- regulate the gene expression have been recognized as essential factors of virulence control in many pathogens. Bacterial regulatory RNAs play prolific roles in mediating gene expression both at the transcriptional and post-transcriptional levels in many pathogens giving new insights into gene regulation (Chakravarty and Massé, 2019).

### 1.1.1 Regulation at transcription level

Bacterial cells tune in their transcriptional regulation to changing environments, through various mechanisms by regulating the activity of RNA polymerase or promoters to which the RNA polymerase can recognize to bind and initiate transcription (Browning and Busby, 2016; Mejía-Almonte et al., 2020). The global regulator 6S RNA is the first example of regulatory sRNA that regulates sigma factor by mimicking a σ70 promoter complex. 6S RNA forms a complex with RNA polymerase that results in regulation of transcription (Wassarman, 2018).

Transcription factors (TF) regulate the activity of promoter through various mechanisms that activate or repress the transcription. Some TFs bind to operator sequences, the conserved –10 or –35 promoter element, thereby preventing the binding of RNA polymerase (Bervoets and

Charlier, 2019). The best-studied example of transcription repression is by the regulatory protein GalR, a repressor of the galactose operon in *E. coli*. Binding of GalR to two distally located operators of the *galETK* operon induces DNA looping between these sites, which prevents binding of RNA Polymerase (Semsey et al., 2004).

TFs can also act as activators and stimulate the rate of transcription. The activator TF binds upstream of the promoter at -35 element or at the space between the –10 and –35 promoter elements and recruits RNAP and enhance the transcription (Bervoets and Charlier, 2019). Examples of transcription activation is, CarD and RbpA proteins which bind to both Eσ and the promoter upstream of the –10 element to stabilize the unstable open complex (RPo) in *Mycobacterium sp*. (Kaur et al., 2018).

### 1.1.2 Regulation at translation level

Gene regulation is not reliant only at transcriptional level; regulation at the level of protein synthesis also results in rapid and fine-tuned responses to challenges. Translation initiation places a significant demand on cellular energy pools to maintain protein stoichiometry. Control of initiation can be achieved through mechanisms such as regulation of key initiation determinants, that includes sequestration of the ribosome-binding site and control of the number of available ribosomes (Tollerson and Ibba, 2020).

The translational regulatory network of bacteria is governed by diverse interaction between target mRNAs, regulatory RNAs, and regulatory proteins. Regulation of translation mediated by sRNAs plays an important role in bacterial gene regulation in response to diverse environmental conditions. Sequestration of ribosomal binding site by mRNA folding or binding of sRNA, reduces the ribosome accessibility and represses the translation. Yet base pairing of sRNA at the upstream region of RBS unfold the secondary loop structure, making the RBS accessible for translation. The formation of a double-stranded RNA duplex between sRNA and the target mRNA often activates translation, through protecting the target mRNA from the selective degradation mediated by RNases (Jeong et al., 2017; Watkins and Arya, 2019).

## 1.2 Small Regulatory RNAs

Regulatory RNAs are small in size, vary in size between 50 to 500 nucleotides and often called as small RNAs (sRNAs) or non-coding RNAs (ncRNAs) (Bloch et al., 2017; Oliva et al., 2015). Non-coding RNAs represent a class of RNAs that are not translated into proteins, but exert various regulatory functions encoded by the RNAs itself or in complexes with proteins. However, RNAs like RNAIII (Morfeldt et al., 1995) and SgrS (Wadler and

Vanderpool, 2009) revealed that small RNAs might code for small peptides. Thus, this class of RNA molecules is now called as small regulatory RNAs. These sRNAs are usually expressed from Intergenic Region (IGR) or genomic islands under specific stress responses or during virulence, iron uptake, quorum sensing, biofilm formation and pathogenicity (Romilly et al., 2012).

Bacterial regulatory RNAs are established as pivotal players in regulating gene expression in the last two decades. Bacterial sRNAs fine-tune the expression of genes involved in diverse physiological and biological processes like virulence, stress response and quorum sensing. The sRNAs exert their regulatory role by base pairing with cis- or trans- encoded target mRNAs, to influence their stability and translation (Chakravarty and Massé, 2019).

Regulatory RNAs can modulate gene expression at various steps like transcription, translation and mRNA stability, and DNA maintenance or silencing. sRNAs exert their diverse regulatory functions through a variety of mechanisms that include changes in RNA conformation, interacting with proteins, base pairing with their target mRNA, and interactions with DNA which all leads to the regulation of metabolism, growth processes and stress adaptation (Bervoets and Charlier, 2019). Besides these roles, small RNAs have emerged as regulators of bacterial pathogenesis and gained focus in last two decades, since the discovery of the role of RNAIII, a small regulatory RNA of *S. aureus*, in targeting several mRNAs that encode virulence factors (Michaux et al., 2014).

After the discovery of sRNA MicF of *Escherichia coli* in 1984, the first sRNA regulator reported that inhibits the translation of the mRNA encoding the major outer membrane porin OmpF, a plethora of small regulatory RNAs have been revealed in diverse bacterial genera where they help the cell to face environmental pressures by regulating gene expressions.

### 1.2.1 Classification of sRNAs in Bacteria

The bacterial small regulatory RNAs exert their regulatory activity either by base pairing with target mRNAs and thereby affecting its translation and/or stability, or by binding and modifying the activity of proteins. Most sRNAs influence gene expression in response to specific environmental changes. sRNAs can be classified in to: i) sRNAs that binds to mRNA and ii) sRNAs that binds protein to modulate their activity and that has been summarized in Figure 1.

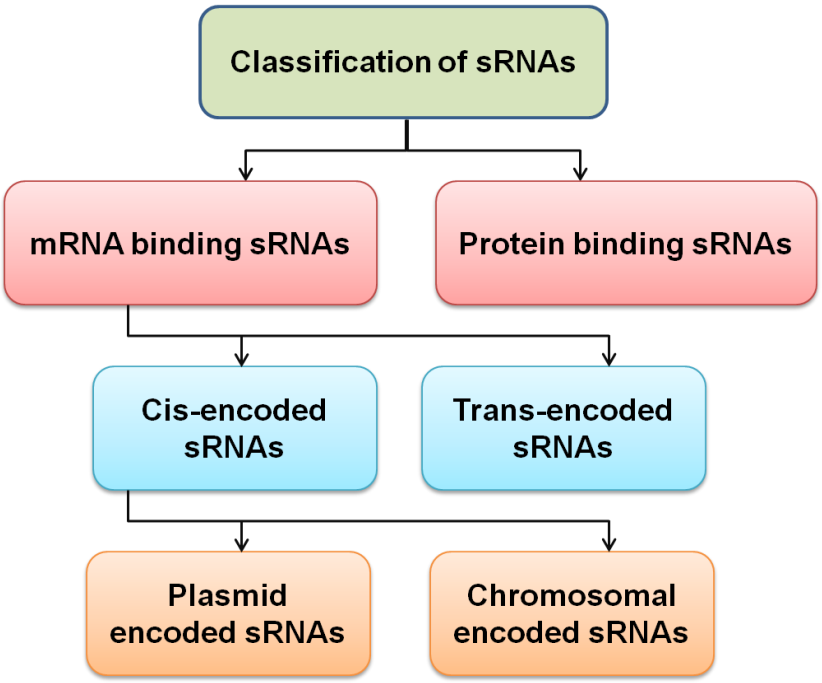


**Figure 1 : Classification of Small Regulatory RNAs**
sRNAs that act by base pairing are classified into Cis and Trans encoded sRNAs based on the genomic locus of their target. Cis-encoded sRNAs can be encoded on plasmid or chromosome of the bacteria. A class of sRNAs act by binding with protein are categorized as protein binding sRNAs.

#### 1.2.1.1 mRNA binding sRNAs

sRNAs that act by base pairing with mRNA are classified into Cis- and Trans-encoded sRNAs based on the genomic locus of their target.

##### 1.2.1.1.1 Cis-encoded sRNAs

Cis-encoded sRNAs interact with target mRNAs which are encoded from the same genetic locus, and are complementary with their target mRNAs with perfect base pairing (antisense regulatory RNAs) (Svensson and Sharma, 2016). Cis-encoded sRNAs are expressed constitutively throughout the cell growth (Waters and Storz, 2009). These antisense RNAs and play a major role in gene regulation by duplex formation with target mRNAs which leads to transcription termination or mRNA cleavage (Watkins and Arya, 2019). An example of a cis-encoded sRNA is the 5'ureB of *Helicobacter pylori*, located at 5'- antisense to the *ureB* gene of the *ureAB* operon involved in the regulation of urease production, a major colonization factor of *H. pylori*. UreB negatively regulates expression of *ureAB* operon by blocking the translation in the 5'UTR of *ureB* (Tejada-arranz et al., 2021). A pictorial representation of regulatory role of *cis*-encoded sRNA is shown in Figure 2. Cis-encoded sRNAs can be further classified in to plasmid encoded and chromosomal encoded sRNAs.

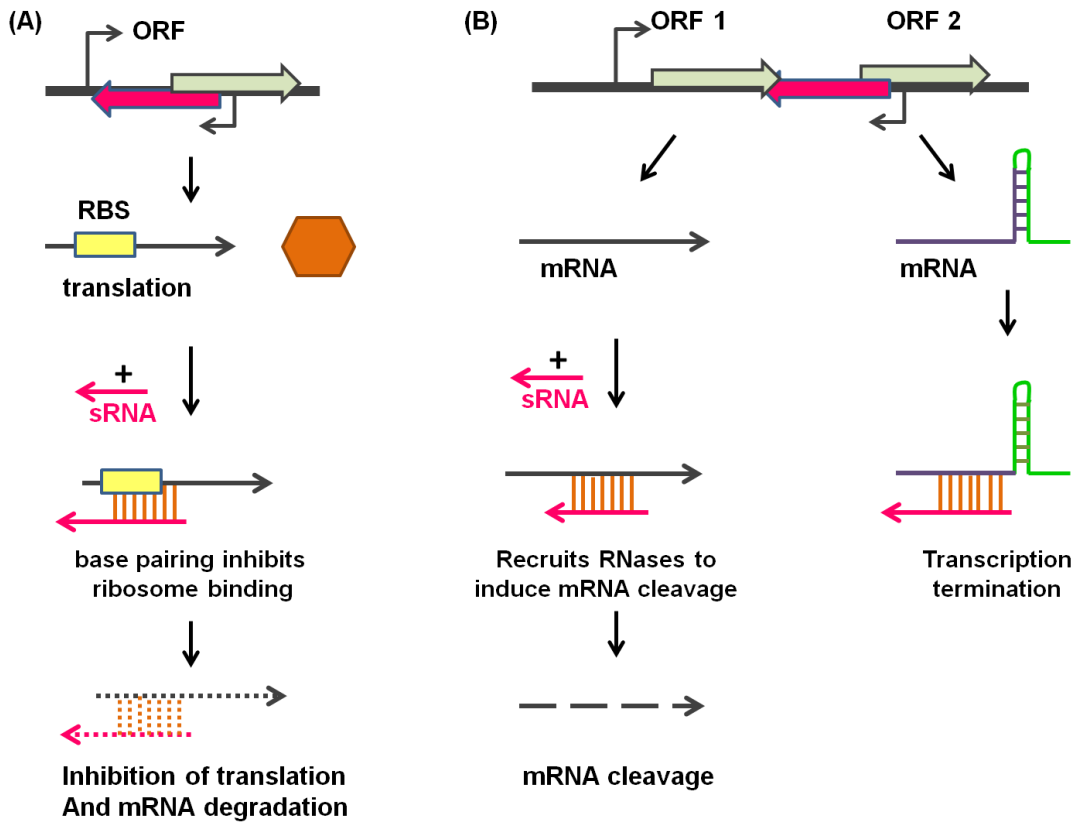


**Figure 2 : Diagrammatic representation of regulatory role of *cis*-encoded sRNA**
(A) *cis*-encoded sRNA, expressed from the antisense strand at the 5' UTR of its target mRNA, shares extensive complementarity. Base pair interaction between sRNA and mRNA inhibits ribosome binding and often leads to mRNA degradation. (B) sRNA expressed from the antisense strand 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 lead to transcription termination. The cis-encoded sRNAs are highlighted in red and their target in black.

#### 1.2.1.1.1.1 Plasmid encoded sRNAs

Most of the identified sRNAs are chromosomally encoded, and only a few are plasmid-encoded sRNAs. The plasmid-encoded sRNAs function through modulating plasmid replication, conjugation frequency, stability and maintenance. They express constitutively but are metabolically unstable, and changes in plasmid copy number affect the levels of these RNAs. The first plasmid encoded sRNA, RNAI, was identified in 1981 from the *E. coli* plasmids, ColE1 and R1. This cis-encoded antisense RNA regulates plasmid replication and controls the plasmid copy number (Kong et al., 2018).

#### 1.2.1.1.1.2 Chromosomal encoded sRNAs

By contrast to plasmid encoded sRNAs, chromosomally encoded antisense RNAs have been found to be expressed only under definite conditions. The first chromosomally encoded regulatory RNA, reported in 1984 was MicF RNA, discovered in *E. coli* that inhibits

translation of an mRNA encoding OmpF, a major trans-membrane protein (Delihas, 2015; Mizuno et al., 1984). The other examples of chromosomally encoded sRNA include the two sRNAs expressed from the *Bacillus subtilis* genome, RatA and SR4, that are part of type I (RNA-regulated) toxin/antitoxin (TA) systems which regulates toxins TxpA and BsrG respectively (Brantl et al., 2012).

#### 1.2.1.1.2 Trans encoded sRNAs

In contrast to cis-encoded sRNAs, trans-encoded sRNAs are transcribed from different genomic locations than the genes they regulate (Azhikina et al., 2015). Trans-encoded sRNAs are often encoded intergenically, can also be derived from 5' or 3' UTRs and even from coding regions. These sRNAs interact with multiple mRNA targets via imperfect base-pairing and regulate their translation or stability (Svensson and Sharma, 2016). A pictorial representation of regulatory roles of trans-encoded sRNAs is shown in Figure 3.

The majority of the *trans*-encoded sRNAs negatively regulates their target mRNAs, usually leads to repression of protein levels through translational inhibition, mRNA degradation, or both. The bacterial sRNAs primarily bind to the 5' UTR of target mRNAs and occlude the ribosome binding site leading to the inhibition of translation and reduction in the protein levels (Watkins and Arya, 2019).

### 1.2.2 Functional division of sRNAs

The regulation mechanisms employed by trans-encoded sRNAs involves, (i) Direct sRNA-mRNA base pair interaction with the target at the RBS leading to inhibition of gene expression. (ii) sRNA-mRNA duplex formation, recognized by RNases leading to mRNA degradation. (iii) binding of sRNA with mRNA to prevent the formation of an inhibitory structure that otherwise sequesters the ribosome binding site (Waters and Storz, 2009). Ultimately, these lead to either translation inhibition or activation. Activation of $\alpha$-toxin encoding gene, *hla*, by RNAIII in *S. aureus* is the first discovered anti-sense type of target activation.

Regulation by trans-encoded sRNAs in gram-negative bacteria is mediated by RNA-binding proteins/chaperons such as Hfq, ProQ, and CsrA, that helps to stabilize and promote sRNA-mRNA base-pairing interaction. However, in gram-positive bacteria, sRNAs exert their regulatory functions without the assistance of RNA chaperone (Jørgensen et al., 2020; Smirnov et al., 2016).

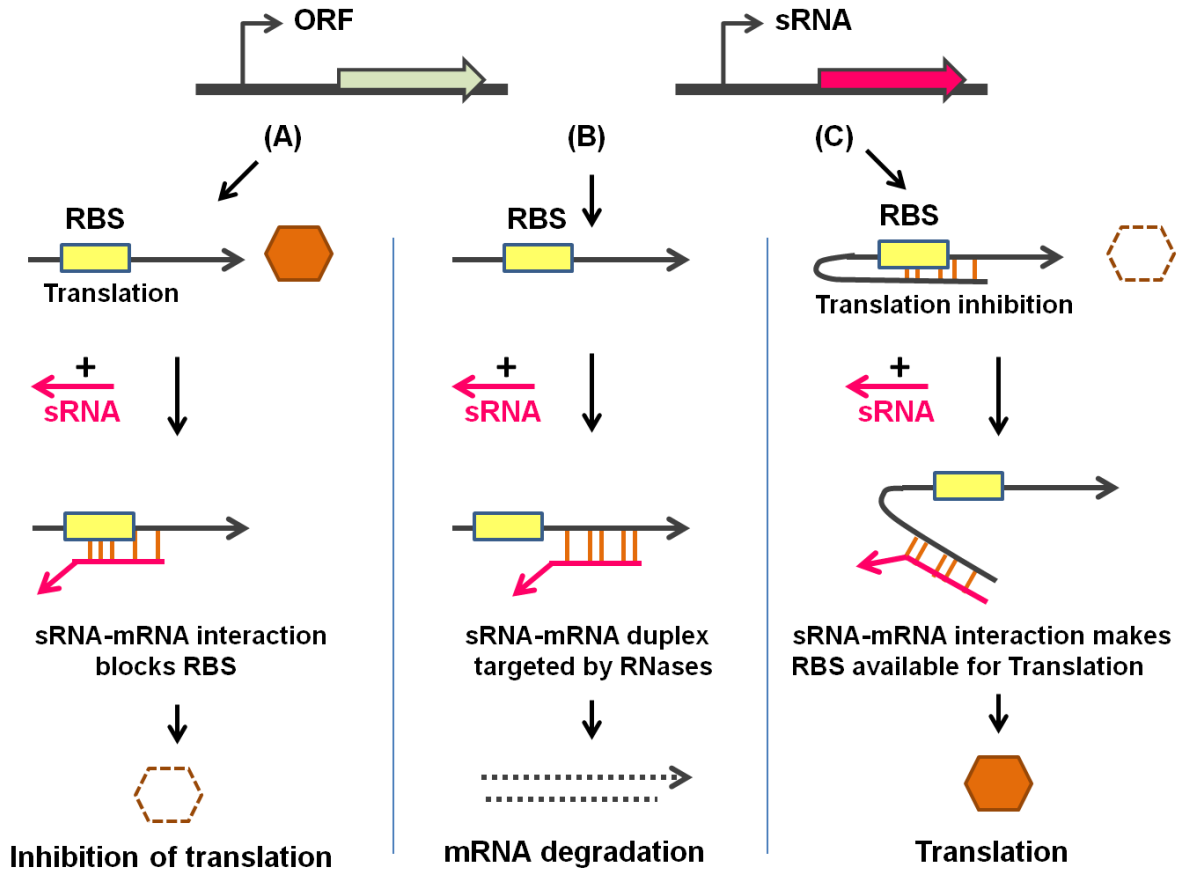


**Figure 3 : Pictorial representation of regulatory role of *trans*-encoded sRNA**
Genes encoding Trans-encoded antisense sRNAs (pink) are located on different genomic locus from the genes encoding their target RNAs (grey). Trans-encoded sRNA share limited complementarity with mRNA. (A) sRNAs base pair at 5' UTR of target mRNA and negatively regulates its expression through translation inhibition. (B) sRNA-mRNA duplex targeted by RNases for degradation which leads to mRNA degradation. (C) Trans-encoded sRNAs act as a positive regulator by preventing the formation of an inhibitory structure which sequesters the ribosome binding site.

Quorum regulatory sRNAs (Qrr sRNAs), LuxO and HapR in *Vibrio cholerae* control the expression of type VI secretion system (T6SS). The Qrr sRNAs down regulates T6SS by direct base pairing at the 5'UTR of the mRNA encoding T6SS cluster or by suppressing the T6SS -activator HapR (Shao and Bassler, 2014).

Although many of the sRNAs have shown negative regulation, an increasing number of sRNAs have been identified that activates translation, in many cases by preventing or overcoming the formation of an inhibitory secondary structure. These sRNAs activate target mRNAs via interactions with the 5' UTR, coding sequence or 3' UTR of transcripts. The PhrS sRNA activates translation of the *pqsR* mRNA, a critical activator of pyocyanin production, a redox-active pigment and a toxin of *P. aeruginosa* by interacting with a sequence in the 5'UTR (Sonnleitner et al., 2011).

### 1.2.2.1 Protein binding sRNAs

In addition to base pairing with mRNA, Protein targeting sRNAs can directly interact with proteins and modulate their activity by mimicking their natural substrates. Global translational regulatory proteins bind to the 5'UTR of the mRNA, preventing ribosome binding which leads to translational inhibition. Protein targeting sRNAs bind to this repressor protein allowing ribosomes to bind to the RBS leading to translation initiation (Figure 4) (Marzi and Romby, 2012).

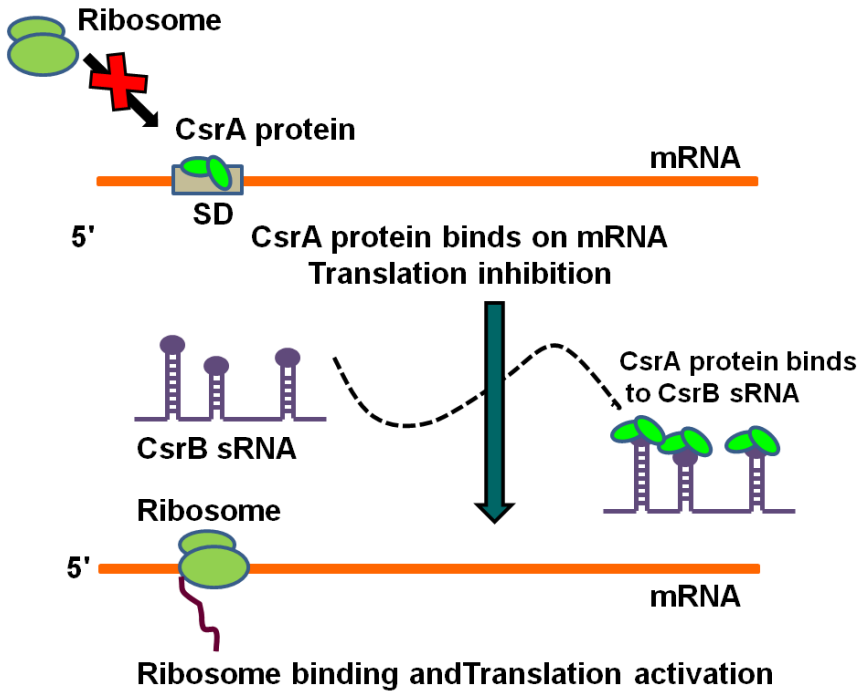


**Figure 4 : Protein binding sRNAs**

Translational repressor protein binds to the 5'UTR of the mRNA and prevents ribosome binding leads to translational inhibition. sRNA binds to this repressor protein allowing ribosomes binding to the RBS to initiate translation.

CsrA is a conserved, abundant sRNA binding protein found in *E. coli* and other gram negative bacteria where it plays a role in the regulation of carbon metabolism, biofilm formation and virulence. The regulatory protein CsrA binds to single-stranded GGA motifs, around the Shine-Dalgarno sequence of the target mRNA where it inhibits or activates translation or influence RNA processing. sRNAs like CsrB/CsrC has repeating sequence of GGA that are regarded as binding sites for the CsrA protein thereby nullifying the effect of CsrA on its target mRNAs (Müller et al., 2019; Romeo and Babitzke, 2018).

#### 1.2.2.2 Dual function sRNAs

Small regulatory RNAs are also called as non-coding RNAs as they generally do not code for proteins and function by base pairing with their mRNA targets. Some bi-functional sRNAs not only regulate their mRNA targets by base pairing mechanism but also code for small peptides that have a role to play in the same or in a different pathway as that of the sRNA that codes for them (Gimpel and Brantl, 2017). The first dual function sRNA discovered was RNAIII in *S. aureus*. RNAIII is a 514 nt long sRNA that modulates the expression of many virulence factors and also encodes the small protein delta hemolysin (Boisset et al., 2007). Other known examples of dual function sRNAs include Psm-mec sRNA in *S. aureus* that encodes α-helical peptide (PSM) and also interacts with the *agrA* mRNA, inhibiting its translation(Kaito et al., 2013). Pel sRNA in *Streptococcus pyogenes* encodes a haemolytic peptide and serves a s a regulatory RNA that activates the transcription of mRNAs encoding virulence factors like M protein, sic mRNA (complement inhibitor) and modulates maturation of the cysteine protease (Mangold et al., 2004). SgrS in *E.coli* protect cells against increased levels of glucose phosphate by regulating the stability and translation of mRNAs encoding proteins involved in glucose transport and catabolism (Görke and Stülke, 2008). The 5' UTR of SgrS encodes SgrT, a 43 aa peptide which counters glucose uptake by directly binding to the PtsG protein (Wadler and Vanderpool, 2007). AzuCR in *E. coli* encodes a 28 amino acid small protein, AzuC which interacts with GplD, glycerol-3-phosphate dehydrogenase protein, required in the catabolism of glycerol and enhances the activity of GlpD (Raina et al., 2021). SR1 in *Bacillus subtilis* encodes SRP1 which binds to GapA involved in glycolysis and it also inhibits translation initiation of *ahrC* mNRA encoding the transcription activator of the arginine catabolic operons (Haq et al., 2021).

#### 1.2.2.3 Riboswitches and thermosensors

Riboswitches are part of the coding mRNA molecule and genetic control elements found within 5' UTRs, that are considered to be cis-acting and have receptors to sense specific environmental signals or cellular metabolites in order to modulate transcription, translation or processing of mRNAs (Cg, 2019; Chakravarty and Massé, 2019). The ligand binding typically induces structural rearrangements of riboswitches that sequester the SD sequence, thereby inhibiting translation initiation (Chatterjee et al., 2021).

The SreA and SreB riboswitches in *Listeria monocytogenes*, can act as trans ncRNAs to inhibit translation of prfA mRNA. PrfA is a transcription regulator that coordinates the expression of many virulence genes in *L. monocytogene* (Krawczyk-Balska et al., 2021).

Temperature is an environmental cue that affects essentially every cellular process. Various mechanisms involving regulatory proteins or utilizing sensory and regulatory RNAs regulate gene expression to cope with sudden temperature changes. RNA thermosensors or thermometers (RNATs) are elements usually located in the 5′untranslated region (UTR) of mRNAs. They operate by altering their secondary structures that respond to differences in temperature, and exert a major effect on translation efficiency of the downstream gene because of the close proximity to the protein-coding region. RNAT forms a stable structure at low temperature (<30°C), masking the RBS and blocking translation (Loh et al., 2018).
SA1313 from *Staphylococcus aureus* and BA5598 from *Bacillus anthracis*, which likely control the expression of putative ATP binding cassette (ABC) transporters, are regulated by a change in environmental temperature (Fernández et al., 2020). A functional thermosensor was identified within the 5' UTR of *cidA*, a gene of *S. aureus* implicated in biofilm formation and survival of the pathogen (Hussein et al., 2019).

### 1.3 Mechanism of gene regulation mediated by sRNAs

Base pairing of sRNA with target mRNA involves a region of minimum 6-8 contiguous base pairs. Depending upon whether the 5'UTR, 3'UTR or the coding region of mRNA base pairs with the sRNA, various regulatory outcomes are manifested. Bacterial small regulatory RNAs have become the most abundant class of post-transcriptional regulators (Pitman and Cho, 2015). Post transcriptional regulation involves modulation of mRNA stabilization or degradation which subsequently affects the efficiency of mRNA translational. Post-transcriptional studies were mainly focused on 5' UTRs (Untranslated Regions) in the past, but recent studies showed the involvement of 3'UTRs in regulating gene expression (Menendez-Gil and Toledo-Arana, 2021; Ren et al., 2017). RNA-binding proteins bind to 5' UTRs to regulate gene expression by modulating the accessibility of RBSs (Ribosomal Binding Sites) on mRNAs (Ren et al., 2017). 3' UTRs not only affect mRNA stability and translation but also control mRNA localization (Menendez-Gil and Toledo-Arana, 2021).

#### 1.3.1 Translation repression

sRNAs that base pair with their mRNA, targets at the 5'UTR occluding the Shine-Dalgarno sequence, thereby preventing the ribosomal subunit from binding to the mRNA, resulting in translational repression (Figure 5A). sRNA mediated mRNA degradation or destabilization is not only dependent on binding of sRNA to 5'UTR of mRNA, but also involves an enzyme complex containing RNase E or RNase II (Le Scornet and Redder, 2019). For example, in *Listeria monocytogenes*, LhrA sRNA regulates chitinase (*chiA*) mRNA by sequestering its

ribosomal binding site and inhibits translation initiation. The *chiA* catalyzes chitin hydrolysis and is involved in *L. monocytogenes* pathogenesis (Nielsen et al., 2011). VrrA, sRNA in *V. cholerae* down-regulate translation of outer membrane protein OmpT by base-pairing with the 5′ region of the *ompT* mRNA in a Hfq (RNA chaperone protein) dependent manner (Song et al., 2010).

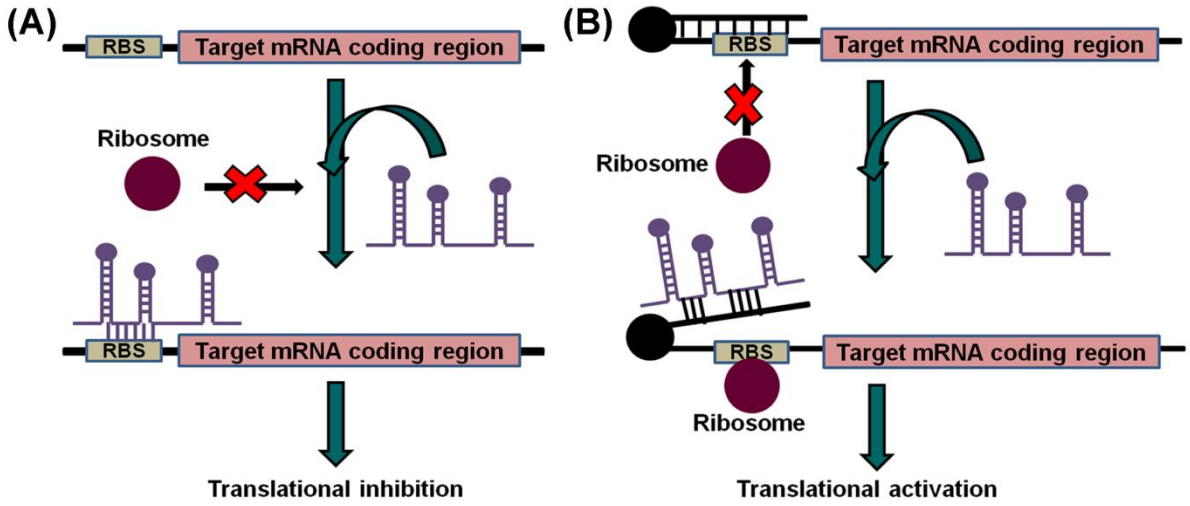


**Figure 5 : Regulatory mechanism of gene expression by sRNA**
(A) Translational repression: Base pairing of the sRNA at the 5'UTR of mRNA occludes the ribosome binding site (RBS) and prevents ribosome binding to initiate translation. (B) Translational activation: An intramolecular secondary structure formed at the 5'UTR of the mRNA sequesters the RBS. The sRNA base pairing at the 5'UTR prevents the formation of inhibitory structures, thus allowing ribosomes to access the RBS and initiate translation.

### 1.3.2 Translational activation

Some examples of mRNA stabilization and translational activation have come up, changing the common perception of sRNAs as negative regulators of gene expression (Papenfort and Vanderpool, 2015). The mechanism of translational activation of the target gene involves unfolding of an intrinsic translation inhibitory structure making mRNA accessible to ribosomes thereby influencing mRNA stability (Papenfort and Vanderpool, 2015) (Figure 5B). RNAIII, most widely studied sRNA of *S. aureus* positively regulates the expression of MgrA which is a global transcriptional regulator by stabilizing *mgrA* mRNA (Gupta et al., 2015).

The mechanism of interference with ribonucleolytic decay involves stabilization of the sRNA-mRNA duplex and thus prevents it from degradation by RNases (Figure 6). RydC sRNA, conserved in a few enteric species including *Escherichia, Salmonella, Citrobacter, Enterobacter, Klebsiella* and *Shigella* activates expression of *cfa* mRNA and subsequently increases protein production. RydC positively regulates *cfa* by base pairing at 5' UTR and

inhibiting RNase E-dependent decay (Fröhlich et al., 2013). FasX sRNA of *Streptococcus* base pairs at the 5'UTR of *ska* mRNA to prevent RNase E-mediated degradation which in turn increases stability of *ska* mRNA promoting enhanced translation and production of the encoded streptokinase (Figure 6) (Ramirez-Peña et al., 2010).

### 1.3.3 Interference with ribonucleolytic decay

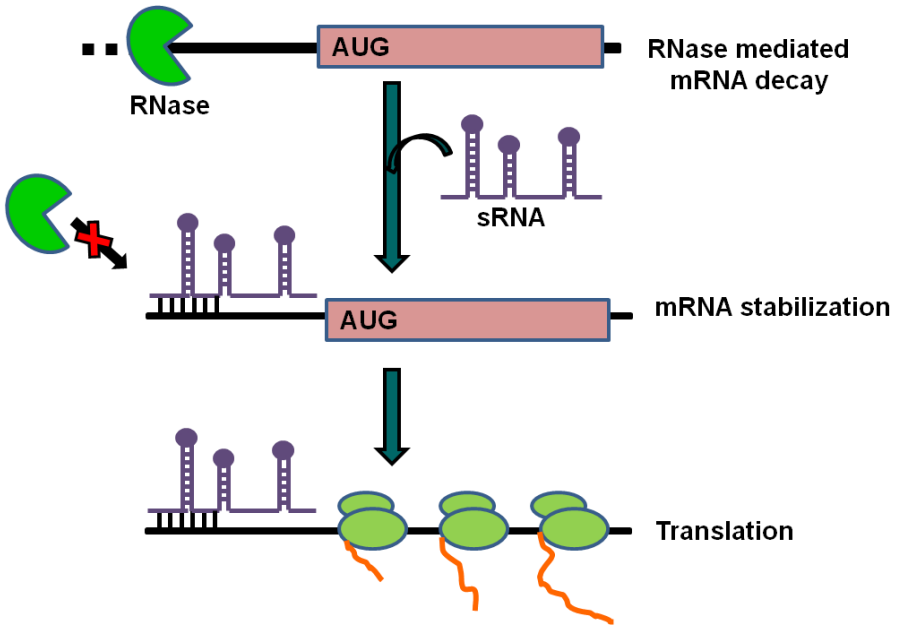


**Figure 6: Translational activation by interference in the ribonucleolytic activity**
sRNA base pairs with target mRNA and interferes with RNase mediated mRNA degradation. The sRNA-mRNA interaction stabilizes the target mRNA by hindering accessibility of RNase leading to enhanced translation.

## 1.4 Identification of small regulatory RNAs in Bacteria

Earlier, potential sRNA candidates were identified by the genome wide searches in intergenic regions using comparative genomics, shotgun cloning or RNomics, microarrays and/or computational searches by examining characteristics like; small in size, expression from the intergenic region, presence of Rho-independent terminator and absence of ORFs (Svensson and Sharma, 2016). Recent studies suggest that small RNAs can be transcribed from 5' or 3' UTR of mRNAs, and few sRNAs also encode small ORFs. An extensive in silico and experimental strategies to identify regulatory RNAs and advancement in genome-wide RNA sequencing approaches such as dRNA-seq (Babski et al., 2016; Sharma and Vogel, 2014), CLIP-seq and Grad-seq (Smirnov et al., 2016) provides a platform for the discovery of a large number of diverse small noncoding RNAs in several bacteria.

### 1.4.1 Bio-computational approach

The computational approach for sRNA identification utilizes some features of sRNAs like sequence homology, conserved secondary structure, structure similarity, presence of promoters, terminators and comparative genomics, employed either alone or in combinations to locate the appropriate sRNA regions (Sridhar and Gunasekaran, 2013).

#### 1.4.1.1 Comparative genomics

sRNA identification based on comparative genomics approach involves main four steps, the first step is to find closely related genomes to a given bacterial genome. The second step is to identify intergenic regions using BLAST program. The pair-wise BLAST hits are gathered into clusters and aligned using ClustalW. The resulting alignments are scored using programs like RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and RNAZ (http://rna.tbi.univie.ac.at/cgi-bin/RNAz/RNAz.cgi). The third step is to carry out structural conservation analysis for the intergenic regions. The fourth step is to predict whether the conserved intergenic regions contain the signal of promoter, transcript factor binding sites or Rho-independent terminator (Backofen and Hess, 2010; Li et al., 2012).

Several computational methods based on a comparative genomics approach have been developed for finding ncRNA sequences, such as QRNA (Rivas and Eddy, 2001), RNAz (Washietl et al., 2005), MSARI (Coventry et al., 2004), and Evofold (Pedersen et al., 2006).

Comparative genomics has been considerably used for the sRNA identification in several bacterial genomes such as *Streptomyces, Cyanobacteria, Sinorhizobium meliloti, Francisella tularensis,* and *Clostridium* spp (Sridhar and Gunasekaran, 2013).

#### 1.4.1.2 sRNAPredict

sRNAPredict uses the relative genomic locations of conserved sequences, transcription signals and ORFs, to identify putative sRNAs encoded in intergenic regions. The sRNAPredict search result gives the details of coordinates, strand orientations and length of the predicted sRNAs, their distance from ORFs, orientation of ORFs and the locus names. sRNAPredict also utilize the user-generated databases of tRNAs/rRNAs and known/putative sRNAs to separate sRNAs that correspond to previously identified transcripts.

Any overlap between the location of a predicted sRNA and a region encoding a tRNA or rRNA is sufficient to exclude that sRNA from the list of potential novel sRNAs, regardless of its strand orientation. sRNAPredict was used to search for sRNAs in *V. cholerae* and identified 9 known sRNAs and 32 novel sRNAs (Livny et al., 2005).

#### 1.4.1.3 sRNA scanner

sRNAScanner utilizes a transcriptional signal-based method to identify intergenic sRNA transcriptional units (TUs) in completely sequenced bacterial genomes. It consists of algorithms to construct PWMs (Position Weight Matrix) from sRNA-specific transcriptional signals, to search for complete genome sequences to identify intergenic promoter and terminator locations. It performs the coordinate based incorporation of promoter/terminator signals to define putative intergenic transcriptional units (TU) and select predicted TUs based on cumulative sum of scores (CSS) values above a nominated threshold. Analysis of the *S. typhimurium* LT2 genome by sRNAscanner under standard conditions yielded a total of 38 known and 118 novel candidate sRNAs (http://bicmku.in:8081/sRNAscanner) (Sridhar et al., 2010; Tsai et al., 2015).

### 1.4.2 Experimental approach for sRNA identification

sRNA candidates identified by bio-computational analysis need experimental validation for their functional significance. Methods like direct labeling and sequencing of RNAs, microarray analysis, and RNomics followed by high-throughput cDNA sequencing (RNA-Seq) were used to identify sRNAs in bacteria. The experimental approaches have undergone constant changes with the new techniques and modifications for the better results. Advancement in RNA sequencing approaches such as dRNA-seq (Bischler *et al*., 2015; Sharma and Vogel, 2014), CLIP-seq and Grad-seq (Smirnov *et al*., 2016) has accelerated sRNA discovery.

#### 1.4.2.1 RNomics approach

This approach is based on shotgun cloning of RNA, in which total RNAs were first subjected to size fractionation, followed by directional cDNA cloning and sequencing. This approach allows the identification of sRNAs in a defined size range that are expressed under specific conditions, irrespective of whether they are encoded independently or generated by processing (Barquist and Vogel, 2015). Steps involved in identification of sRNA by RNomics approach include RNA sequencing (enzymatically or chemically), generation of cDNA libraries by parallel cloning of many sRNAs, microarrays to predict sRNAs that are expressed under a given experimental condition. The transcriptome study has revealed multiple new sRNAs in *V. cholerae* (Papenfort et al., 2015b) and in *M. tuberculosis* (Ostrik et al., 2021) by the RNA sequencing technique.

#### 1.4.2.2 Transposon-sequencing (Tnseq)

Many of the first functionally characterized sRNAs were identified on mobile genetic elements such as plasmids, phages, and transposons. Deep sequencing of transposon (Tn) insertion (TnSeq) libraries has become a powerful tool for evaluating the essentiality of bacterial genomic features. The location of transposon insertions in high-density mutant libraries can be identified by Tn-seq. This provides the resolution necessary to identify short genomic regions that might encode sRNAs or regulatory regions. Additionally, combination of Tn-seq with whole transcriptome RNA-seq data allows for the identification of transcriptionally active regions (Svensson and Sharma, 2016).

Tn-sequencing has been used in identification of virulence associated sRNAs of several pathogens including *N. gonorrhoeae* (Remmele et al., 2014), *S. pneumoniae* (Mann et al., 2012), *M. tuberculosis* (Dejesus et al., 2017), *Salmonella* spp.(Ellis et al., 2017), and *Haemophilus influenzae* (López-López et al., 2021).

#### 1.4.2.3 Differential RNA-sequencing (dRNA-seq)

The dRNA-seq is a deep-sequencing based approach for selective sequencing of primary transcripts. The primary transcripts in bacteria carry a 5' Triphosphate (5' PPP) end while processed transcripts carry 5' Monophosphate (5' P). The differential 5' P-dependent terminator exonuclease (TEX) treatment is employed that distinguishes between primary and processed transcripts. TEX specifically degrades 5' P RNAs, thus enriching primary transcripts with transcriptional start site (TSSs) and provides evidence of specific novel transcripts at the single-nucleotide level in relative terms (Bischler et al., 2015; Sharma and Vogel, 2014). dRNA-seq also uncovers regulatory sRNAs emerging from 3' UTR of bacterial mRNAs (Miyakoshi et al., 2015; Smirnov et al., 2016). dRNA-seq has been used to discover >60 sRNAs in *H. pylori* disproving a prediction that this organism has no ribo-regulators due to its lack of the common RNA chaperone Hfq (Bischler et al., 2015).

#### 1.4.2.4 CLIP-seq

Cross-linking Immuno-precipitation (CLIP) is a modification of the RNA ImmunoPrecipitation (RIP)-seq approach to identify direct RNA ligands to reveal high resolution data of RBP (Ribosomal Binding Protein) binding sites. In CLIP-seq, cells are irradiated with UV light, leading to formation of a covalent bond between protein and RNA followed by cell lysis and immune-precipitation using antibodies against the RBP or with protein-specific antibodies. RNA-protein complexes are separated, purified and protein bound RNA is released (Andresen and Holmqvist, 2018; Barquist and Vogel, 2015). This

strategy was employed for the determination of transcriptome-wide binding sites of Hfq and CsrA in *Salmonella* (Holmqvist et al., 2016).

#### 1.4.2.5 Grad-seq

The gradient profiling by sequencing (Grad-seq) is an advanced approach which uses RNA–protein interaction analyses to draft global RNA landscape by partitioning all cellular transcripts into diverse coding and non-coding groups based on RNA–protein interactions. Grad-seq relies on the sedimentation of cellular RNAs and proteins in a glycerol gradient. Following this biochemical partitioning, the RNA is subjected to Illumina cDNA sequencing and the sedimentation profile of individual transcript is visualized. These transcripts and associated proteins are then identified by transcriptomic and proteomic analyses (Smirnov et al., 2016). This approach was used for characterizing the regulatory role of novel RNA binding protein ProQ and identified its role in sRNA regulation (Gerovac et al., 2021; Smirnov et al., 2016). In *E. coli,* a stable association of RyeG ncRNA with the 30S ribosomal subunit reveals RyeG as an mRNA that encodes a small toxic protein (Giorgio et al., 2020).

## 1.5 Identification of mRNA targets for sRNA

With more and more sRNAs being identified in various genera of bacteria, determining their functions has become an important part of sRNA biology. Most of the sRNAs act by binding to their mRNA targets via a small stretch of 6-8 bp (Gottesman and Storz, 2011). The sRNA-mRNA base pairing is highly imperfect and different regions of sRNA base pair with the targets making the target identification difficult. Following are some of the bioinformatics programs used to predict mRNA targets of sRNAs and experimental methods used for the validation of the predicted targets.

### 1.5.1 Computational programs for target prediction

Computational tools for sRNA target prediction determine the putative targets based on the input parameters like non-canonical base pairing, length of base pairing interaction, region of mRNAs (5'UTR, coding region, 3'UTR) involved in interaction. Following are the few computational programs for the for sRNA target prediction.

#### 1.5.1.1 TargetRNA

TargetRNA program takes sRNA genomic sequence as the input to determine the hybridization potential of the submitted sRNA sequence to each mRNA of the selected genome. TargetRNA uses the modified model of Smith-Waterman dynamic program that is

based on base-pairing potential instead of homology potential. TargetRNA determines base-pairing scores based on the statistical significance of each potential sRNA-mRNA interaction and gives a ranked list of mRNAs as the candidate targets whose binding potential with the sRNA sequence meets a significance threshold (the default P-value $\leq 0.001$) (Tjaden, 2008).

#### 1.5.1.2 TargetRNA2

Target RNA2 is built on already existing programs for prediction of sRNA targets in bacteria like TargetRNA, RNA predator, IntaRNA and Peer and Margalit's different approach which are focused on conserved and accessible target binding regions of sRNAs sequence compared to the other regions of the sRNA sequence. TargetRNA2 employs four important features for the determination of sRNA targets: 1) Conservation of the sRNA, 2) Accessibility of the sRNA, 3) mRNA accessibility and 4) Hybridization energy (Kery et al., 2014).

#### 1.5.1.3 IntaRNA

IntaRNA (Interacting RNAs) efficiently predicts the interaction between two RNA molecules based on the accessibility of interaction sites and the seed size. This algorithm calculates the combined energy score of the interaction as the sum of the free energy required for hybridization and for making the interaction sites accessible. It also provides minimal energy profiles for interacting RNA: RNA pairs that can help in the study of alternative RNA: RNA interaction (Mann et al., 2017). This program can also be used as a guide for wet-lab studies to analyze mutational effects between the interacting RNAs and previously used for validation of targets (Beisel et al., 2012; Robledo et al., 2015).

#### 1.5.1.4 CopraRNA

CopraRNA computes whole genome predictions for conserved sRNA sequence by integrating the phylogenetic information from the diverse organisms and works in alliance with the IntaRNA program. The results are compared with each other based on the p-value of each putative target and the p-value is employed for calculation of q-value which estimates the false discovery rate of the target prediction (Wright et al., 2014).

When TargetRNA, IntaRNA and CopraRNA were compared, CopraRNA was found to have the highest positive predictive value (PPV) and reported to have the lowest rate of false positive results across 18 enterobacterial species (King et al., 2019). CopraRNA has been used for the sRNA target predictions from enterobacterial species, non enterobacterial systems and serve as the comprehensive benchmark tool for target identification (Durand et al., 2015; Lokstein et al., 2014; Pain et al., 2015; Robledo et al., 2015).

#### 1.5.1.5 RNA predator

RNApredator is a web server built upon algorithm RNAplex for the genome-wide prediction of sRNA targets in bacterial genomes. RNApredator considers the accessibility of the target during the target search that improves the specificity of the predictions. RNApredator offers a graphical overview of the accessibility around the RBS of the target upon sRNA binding. After completing the computation of all candidate sRNA–target interactions, it gives a ranked list of mRNA targets sorted by the energy of interaction (Eggenhofer et al., 2011).

#### 1.5.1.6 SPOT (sRNA Target Prediction Organizing Tool)

SPOT uses several existing sRNA prediction algorithms TargetRNA2, sTarPicker, IntaRNA, and CopraRNA in parallel to search for sRNA-mRNA interactions. The software combines predictions from all the algorithms and allows integration of experimental data using customizable filters. The targets predicted by only a single algorithm are listed in a particular order: CopraRNA, TargetRNA or sTarPicker, and IntaRNA (King et al., 2019).

Other online target prediction tools like RNAup (Mückstein et al., 2006), sRNATarget (Cao et al., 2009), sTarPicker (Ying et al., 2011) are also available for mRNA target prediction for sRNA.

### 1.5.2 Experimental approach for target identification

Although various computational algorithms have been developed, their performance in predicting direct regulatory targets of sRNAs is inconsistent. The non-canonical base pairing results in much difficulty in detecting mRNA targets, which makes the experimental validation essential after computational prediction (Li et al., 2012). A number of experimental approaches have been developed to facilitate the identification of regulatory targets for sRNAs. Earlier, the standard screening approaches to identify putative targets used transcriptomic or proteomic measurements of strains either overexpressing or deleted for particular sRNAs. This approach cannot discriminate between direct sRNA targets from the downstream targets specifically under overexpression (Barquist and Vogel, 2015).

#### 1.5.2.1 Ribosome profiling

Ribo-seq is a powerful approach for experimental identification of sRNA targets that can simultaneously determine RNA levels and levels at translation. In this method, translating ribosomes of a bacterial culture are isolated and treated with RNases *in vitro*. Ribosome-protected mRNA fragments called as the “footprints,” are then isolated and sequenced in high throughput, thereby allows taking a snapshot of translation (Glaub et al., 2020). Regulatory

targets of sRNA are identified by comparing total RNA levels or ribosome-footprinted RNA levels for all the genes in sRNA expressing and control cell. Ribo-seq has been used to experimentally identify regulatory targets of the *E. coli* sRNA RyhB (Wang et al., 2015).

#### 1.5.2.2 RIL-seq (RNA interaction by ligation and sequencing)

RIL-seq is an experimental–computational methodology that takes advantage of the binding of the sRNA and its target RNA to the RNA chaperone protein Hfq, Bacteria carrying the Flag-tagged Hfq are UV-radiated to cross-link the protein bound sRNAs. The exposed regions of bound RNAs are trimmed by RNases followed by ligation of neighbouring RNAs, cDNA library preparation and sequencing. The computational pipeline maps the sequenced fragments to the genome and reveals chimeric fragments. A statistical analysis is then applied to select significantly over presented chimeric fragments in the data as interacting RNAs (Melamed et al., 2018, 2016).

This approach was used for the identification of Hfq associated targets in *E. coli* under different stress conditions which uncovered the hundreds of new Hfq associated sRNA-mRNA targets as well as novel small RNAs involved in global cellular processes. The RIL-seq data indicated that 3′ UTR-derived BhsB sRNA of *E. coli* interacts with two targets, *ytfK* and *ompC,* and both were shown to be involved in the cellular response to oxidative stress (Bar et al., 2021).

#### 1.5.2.3 GRIL-seq (Global sRNA target Identification by Ligation and Sequencing)

GRIL-seq method takes advantage of the proximity of the sRNA and mRNA target sites in a complex stabilized by the Hfq protein. This complexity facilitates a preferential ligation of the 3′ and 5′ ends by bacteriophage T4 RNA ligase co-expressed in the same cell allowing identification of transcripts recognized by sRNAs.. GRIL-seq can be used to identify multiple targets of sRNAs based on their ability to form transient complexes with mRNAs as well as with other transcripts (Han et al., 2016). Regulatory targets of PrrF1, an iron-regulated sRNA were identified using this method in *Pseudomonas aeruginosa* (Zhang et al., 2017).

#### 1.5.2.4 MAPS (MS2-affinity purification coupled with RNA sequencing)

In MAPS, the sRNA of interest is tagged at its 5′ end with the MS2 RNA aptamer and expressed in bacteria. Cytoplasmic extracts are purified by affinity chromatography followed by the sequencing of the enriched MS2–sRNA–RNA complexes. The advantages of MAPS include the detection of the poorly expressed targets and discrimination between direct and indirect targets to find targets that base pairs with sRNA. This experimental approach has

been applied successfully to many sRNAs from Gram-negative bacteria such as *E. coli*, *S. typhimurium*, and from Gram-positive bacteria like *S. aureus* (Georg et al., 2020).

### 1.6 Small RNAs in pathogenic bacteria

Over the past two decades small regulatory RNAs have gained massive appreciation for their roles in mediating post-transcriptional gene regulation of various physiological and virulence-related processes in bacteria by synchronizing complex networks of stress adaptation (Piattelli et al., 2020). sRNAs represent a powerful regulatory reservoir to provide the fast adaptive response to rapidly changing environmental conditions. Of large number of sRNAs in bacteria, some of them are known to play role in regulating cell response to various stress factors or pathogenesis (Azhikina et al., 2015).

sRNAs are involved in several mechanisms that help the pathogen in adaptation, counterfeiting, or suppress the host immune system, along with the other molecules participating in the complex process of pathogenicity regulation. The involvement of sRNAs at infection processes target a number of genes or transcription factors, leading to continuous changes of expression and responses to fluctuating environmental stresses (González Plaza, 2020). Identification of such regulatory sRNAs in pathogens, and elucidation of their functions are crucial for understanding the role of these molecules in the progression of infection and in the host-microbe interaction (Ostrik et al., 2021).

In many cases, regulatory functions of sRNAs are dependent on "RNA chaperones", which function by binding to secondary and tertiary structures of RNA molecules and inducing structural changes (Djapgne and Oglesby, 2021; Quendera et al., 2020).

sRNAs have been so far found to be involved in regulatory mechanisms of almost all the bacterial species, playing crucial role in metabolism, and virulence of pathogenic bacteria such as *Escherichia coli* (Sy and Tree, 2021), *V. cholerae* (Pérez-Reytor et al., 2017), *Listeria monocytogenes* (Krawczyk-Balska et al., 2021), *Pseudomonas aeruginosa* (Pita et al., 2018), *S. aureus* (Liu et al., 2018), *M. tuberculosis* (Ostrik et al., 2021). Regulatory targets of sRNAs in various pathogens are described in table 1.

#### 1.6.1 sRNAs involved in stress response and metabolism

Bacteria encounter various stresses when they come in contact with the host and to adapt to the host environment, bacteria have developed general stress responses and metabolism remodeling. Under glucose limitation, cellular cAMP levels are increased which activates expression of the sRNA CyaR (cyclic AMP-activated RNA) by the activator protein CRP (cAMP receptor protein) in *E. coli* and *S. typhimurium*. CyaR sRNA represses the expression

of OmpX, a major outer membrane protein that stimulates bacterial adhesion; LuxS, the autoinducer-2 synthase; NadE, an essential NAD synthetase; and YqaE, a predicted membrane protein with an unknown function (Lalaouna et al., 2018; Rau et al., 2015; Wang et al., 2020).

Under nutrient deficient condition, F6 sRNA of *M. tuberculosis* affects the expression of the essential chaperonins, GroEL2 and GroES (Houghton et al., 2020). MrsI sRNA of *M. tuberculosis* involved in the regulation of iron metabolism, targets the *bfrA* mRNA encoding bacterio-ferritin – an iron deficiency protein (Gerrick et al., 2018), and Mcr11 regulates the expression of operons involved in regulation of lipid metabolism (Moores et al., 2017).

RsaI sRNA involved in the metabolic regulation of *S. aureus*, controls the signaling pathway of glucose uptake. Global approaches like MAPS and transcriptomic analysis have shown that RsaI regulates genes involved in sugar uptake and metabolism, cytochrome biosynthesis, anaerobic metabolism, arginine catabolism and biofilm formation. The MAPS approach revealed that RsaI interacts with other sRNAs RsaG, RsaD, and RsaE which links its metabolic role to virulence of *S. aureus* (Bronesky et al., 2019).

### 1.6.2 sRNAs regulating host-pathogen interactions

The important steps in the infection cycle of a pathogen include the capacity to limit and repair damages during the host-related stress conditions, to escape the immune system and to develop antibiotic resistance. The regulation of these steps during host colonization and infection is tightly regulated and sRNAs are key players in these regulatory processes.

In *V. cholerae*, $\sigma^E$ activates the expression of Hfq-dependent sRNAs, VrrA and MicV. MicV, shares the same base-pairing region as VrrA. VrrA represses the translation of ompA mRNA causing increased release of outer membrane vesicles (OMVs). OMVs are essential for niche colonization, transport of virulence factors into host cells, modulation of host defense, and response and communication with the surrounding environment (Ahmadi Badi et al., 2020). VrrA suppresses the major outer membrane protein OmpT, the biofilm matrix protein RbmC and the ribosome-binding protein Vrp, involved in starvation survival (Peschek et al., 2019; Sabharwal et al., 2015).

Two RybB homologous sRNAs, RfrA and RfrB in *Salmonella enterica* play an essential role in the survival of a pathogen within the macrophages and also regulates the expression of type3 secretion system (T3SS) (Leclerc et al., 2013). MTS1338 sRNA is highly expressed during the stationary phase of growth and the dormancy state. This sRNA is highly conserved in genomes of highly pathogenic *Mycobacteria* and its transcription is controlled by the

transcriptional regulator DosR. MTS1338 play a role during the stable phase of infection, when host responses confront mycobacterial multiplication more or less successfully and triggers adaptive biochemical cascades for intracellular persistence (Salina et al., 2019).

### 1.6.3 sRNAs in regulatory networks

Two-component systems (TCSs) along with sRNAs allow pathogenic bacteria to quickly sense and respond to different stimuli during host colonization. TCSs and stress-related sigma factors, regulatory RNAs and proteins comprise the major components of regulatory networks that tightly control gene expression in pathogenic bacteria. Such regulatory networks have been recently described in both gram-negative and gram-positive pathogens and are associated with stress response, colony morphology, motility, competence control, and other virulence-related processes (Brosse and Guillier, 2018; King et al., 2020). The most extensively studied regulatory network of OmrA/B sRNAs is regulated by OmpR-EnvZ TCS. OmrA/B sRNAs negatively auto-regulate their transcription by directly repressing the *ompR-envZ* operon that encodes their transcriptional activators (Guillier and Gottesman, 2008). In addition to the above stated examples, a large number of other sRNAs have been identified and characterized for their involvement in regulating the virulence of pathogens, which are listed in Table 1. Examples of sRNAs targeting TCS genes in different bacteria are presented in Table 2.

**Table 1: sRNAs regulating pathogenicity targets**

| Pathogen | sRNA | Target | Related Phenotype/function | Reference |
|---|---|---|---|---|
| *Mycobacterium tuberculosis* | Mcr7 | *tatC* | Involved signaling by binding to its mRNA | (Ostrik et al., 2021) |
| *Clostridium perfringens* | VR-RNA | *cola* | Toxin expression | (Obana et al., 2010) |
| Group A streptococci | FasX | *ska, cpa* | Streptokinase and pilus expression | (Liu et al., 2012) |
| *Helicobacter pylori* | RepG | *tlpB* | Antisense repressor of chemotaxis receptor level | (Tejada-arranz et al., 2021) |
| *Listeria monocytogenes* | LhrA | *chiA* | Chitinase | (Krawczyk-Balska et al., 2021) |
| | LhrC | *lapB* | Represses a virulence-associated adhesion | |
| | Rli27 | Lmo0514 | Activation of expression of a cell wall protein inside cells | |

| | | | | |
|---|---|---|---|---|
| *Pseudomonas aeruginosa* | PhrS | *pqsR* | couples quorum sensing with oxygen availability | (Malgaonkar and Nair, 2019; Sonnleitner et al., 2011) |
| | PhrD | *rhlR* | Quorum sensing regulator | |
| *Salmonella typhimurium* | InvR | *ompD* | Represses a porin of the core genome | (Pfeiffer et al., 2007) |
| | IsrM | *sopA, hilE* | Invasion, intracellular replication, virulence, and colonization of mice | (Gong et al., 2011) |
| | DapZ | *oppA, dppA* | 3' UTR encoded sRNA Regulates amino acid / oligo peptide metabolism | (Chao et al., 2012) |
| | SgrS | *sopD* | Core genome sRNA that regulates a SPI-1 effector | (Papenfort et al., 2012) |
| *Shigella dysenteriae* | RyhB | *virB* | Affects T3SS/effectors and virulence | (Murphy and Payne, 2007) |
| *Salmonella enteric* | RprA | *rpoS, ricI* | Controls conjugation of pSLT virulence plasmid in response to membrane conditions | (Papenfort et al., 2015a) |
| | AmgR | *mgtC* | Survival in macrophages | (Lee and Groisman, 2010) |
| Streptococcus pyogenes | FasX | *ska, cpa, prt*F1/F2 | Activation of the secreted virulence factor streptokinase and repression of pilus formation | (Klenk et al., 2005) |
| | RivX | *Mga* | Regulates expression of virulence transcription factors | (Roberts and Scott, 2007) |
| *Vibrio cholera* | VqmR | *vpsT, rtx* | Biofilm formation and toxin expression | (Papenfort et al., 2015b) |
| | TarB | *tcpF* | Regulates a virulence factor and affects colonization | (Bradley et al., 2011) |
| | Qrr | *vca0939* | QS control and cyclic-di-GMP synthesis | (Zhao et al., 2013) |
| | VrrA | *ompA* | Outer membrane vesicle synthesis & colonization of mouse intestine | (Sabharwal et al., 2015) |
| | CsrB/C/D | CsrA | QS and virulence factor production | (Lenz et al., 2005) |
| *Xanthomonas campestris* | sX13 | *hrpX* | Affects levels of T3SS regulator | (Schmidtke et al., 2013) |

**Table 2**: **Examples of sRNAs regulating TCS in bacteria**

| Pathogen | sRNA Regulating TCS | Regulated TCS | References |
|---|---|---|---|
| **Gram-negative Bacteria** | | | |
| *Escherichia coli* | RybC | DpiA-DpiB | (Mandin and Gottesman, 2009) |
| *E. coli* | SdsN | NarQ-NarP | (Hao et al., 2016) |
| *E. coli, Salmonella* | GcvB, MicA | PhoP-PhoQ | (Coornaert et al., 2013, 2010) |
| **Gram-positive Bacteria** | | | |
| *Staphylococcus aureus* | psm-mec | AgrA-AgrC | (Kaito et al., 2013) |
| *Clostridioides difficile* | Cd2-2 | CmrR-CmrS-CmrT | (Garrett et al., 2019) |
| *Streptococcus pneumoniae* | srn206 | ComD-ComE | (Acebo et al., 2012) |
| *Entercoccus faecalis, Listeria monocytogenes* | EutX, Rli55 | EutVW | (DebRoy et al., 2014; Mellin et al., 2014) |

### 1.7 *Staphylococcus aureus* – A human pathogen

*Staphylococcus aureus* is a major human pathogen that is capable of producing an expansive repertoire of cell surface-associated and extracellular virulence factors. The ability of *S. aureus* to adhere to extracellular matrix (ECM) and plasma proteins is a key factor in the colonization and dissemination throughout the host. It is the most common cause of hospital and community-acquired infections. As a commensal bacterium *S. aureus* asymptomatically colonizes about 30% of the human population (Tong et al., 2015). Clinically, *S. aureus* is responsible for a large spectrum of diseases, from localized to systematic infections. Amongst the more common severe pathologies caused by *S. aureus* are endocarditis, osteomyelitis, and septicemia and is also a leading cause of pneumonia and other respiratory tract infections, cardiovascular infections, nosocomial bacteremia and atopic dermatitis (Cheung et al., 2021; Nakamura et al., 2014; Sutton et al., 2021).

*S. aureus* infections are challenging due to frequently occurring antibiotic resistance, among which methicillin-resistant *S. aureus* (MRSA) are the most important clinically. Infections caused by MRSA leads to increased mortality and morbidity, as compared to those caused by methicillin-sensitive *S. aureus* (MSSA) (Cheung et al., 2021).

#### 1.7.1 *Staphylococcus aureus* Newman

*S. aureus* strain Newman was isolated first in 1952 from a human infection and has been extensively used to examine the therapeutic efficacy of agents designed to treat *S. aureus* infection due to its robust virulence phenotypes in animal models of staphylococcal disease

(Sause et al., 2017). *S. aureus* strain Newman was first sequenced in the year 2008 and carries a similar combination of major pathogenicity islands, νSaα and νSaβ, as other *S. aureus* strains COL, NCTC8325, and USA300. The genome size of *S. aureus* Newman is 2.8 Mbp encoding 2,614 open reading frames and G+C content of 32.9 % (Baba et al., 2008).

### 1.8 Small regulatory RNAs in *S. aureus*

RNAIII was the first regulatory RNA of *S. aureus* discovered in 1993 by Novick and collaborators (Novick et al., 1993). In 2005, the existence of seven small RNAs encoded from pathogenicity islands (PIs) was discovered and named them Spr for "small pathogenicity island RNA" (Pichon and Felden, 2005). sRNAs were detected by performing bioinformatics analysis and their existence was experimentally confirmed by northern blot. In 2009, Geissmann and colleagues showed the expression of eleven chromosomal sRNAs, named Rsa for RNA from *S. aureus* (Geissmann et al., 2009). Table 3 gives the summary of functionally characterized sRNAs of *S. aureus.*

#### 1.8.1 sRNAs involved in metabolic regulation

Many of the sRNAs encoded by the core genome are highly conserved and play a central role as metabolic regulators. RsaE is one of the well-characterized sRNAs, involved in the regulation of several metabolic pathways and conserved amongst bacteria belonging to genera *Staphylococcus*, *Macrococcus*, and *Bacillus.* In *S. aureus*, RsaE lowers the expression of many enzymes associated with the TCA cycle, folate metabolism, peptide transport and lipid, purine, and carbohydrate metabolism which promotes adaptation of bacterial cells to low concentration of nutrients. The expression of *oppB*, amino acid and peptide transporter protein; *sucD*, succinyl-CoA synthetase and *rocF*, arginase is controlled by RsaE via direct interactions with ribosome binding sequence to inhibit mRNA translation (Rochat et al., 2018).

#### 1.8.2 sRNAs that couples quorum sensing to virulence

Quorum sensing (QS) is a bacterial cell signaling system which regulates bacterial physiology. In *S. aureus* QS system is mainly regulated by Agr (accessory gene regulator) system, which also encodes the major regulatory sRNA RNAIII. RNAIII is a bifunctional sRNA, as it encodes δ-hemolysin (*hld*) which contributes towards *S. aureus* virulence. RNAIII activates *hla* translation by base-pairing at the 5'UTR of *hla* mRNA that leads to the unmasking of the ribosome binding site to activate its translation (Svenningsen, 2018). RNAIII as an effector regulates the expression of important virulence genes including

proteins associated with cell wall metabolism and exotoxins. These include protein A, α-toxin, adherence proteins Sbi, and Map, a cell wall hydrolytic enzyme LytM, and master transcription factors Rot and MgrA. RNAIII also regulates the expression of two-component systems and other mechanisms in the formation of biofilms, synthesis of amino acid and peptidoglycan (Bronesky et al., 2016; Howden et al., 2013; Lioliou et al., 2016).

### 1.8.3 sRNAs expressed from pathogenicity island

Pathogenicity islands (PAIs) are horizontally acquired distinct genetic elements that encode various virulence factors in pathogenic strains and often absent in non-pathogenic strains of the same or closely related species. sRNAs located on PAIs suggests that these play an important role during *S. aureus* infections (Felden et al., 2011). sRNAs that are expressed from the pathogenicity islands includes SprA (Sayed et al., 2012), SprC (Le Pabic et al., 2015), SprD (Chabelskaya et al., 2010) and SprX (Buchad and Nair, 2021; Eyraud et al., 2014; Kathirvel et al., 2016). SprC regulates *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). SprD interacts at the RBS of *sbi* mRNA, encoding an immune-evasion molecule and inhibits translation initiation (Chabelskaya et al., 2010). SprX regulates the expression of delta hemolysin and clumping factor B, by direct interaction with the hld and clfB transcripts, and influence the expression of Immunodominant Staphylococcus antigen A (IsaA) and biofilm formation (Kathirvel et al., 2016). SprX positively regulates the autolysin expression by directly interacting with the mRNA of autolysin regulator WalR (Buchad and Nair, 2021).

**Table 3: List of functionally characterized small RNAs in Staphylococcus aureus.**

| sRNAs | Targets | Mode of action | References |
|---|---|---|---|
| **sRNAs encoded from core genome** | | | |
| **RNAIII (514 nt)** | *hla*↑, *hlb*↑, *mgrA*↑ *rot*↓, *protein A*↓, *Spa*↓, *coa*↓, *SA1000*↓, *SA2353*↓ | Activator of the translation of the *hla* mRNA, encoding α-hemolysin. Up-regulate the expression of toxic shock syndrome toxin, enterotoxins and enzymes like lipases, proteases and nucleases. Also activates global regulator MgrA by stabilizing its mRNA. Represses the expression of surface proteins such as protein A and coagulase and inhibits the translation of rot mRNA | (Boisset et al., 2007; Chabelskaya et al., 2010; Gupta et al., 2015; Morfeldt et al., 1995) |

| | | | |
|---|---|---|---|
| **RsaA (139 nt)** | *mgrA*↓ | Represses the synthesis of the global transcriptional regulator MgrA mRNA. Indirectly activates the synthesis of surface proteins that stimulate biofilm formation. It decreases the synthesis of capsule formation. Several mRNAs were validated as direct RsaA targets which encode a family of SsaA-like enzymes that are involved in peptidoglycan metabolism and the secreted anti-inflammatory FLIPr protein. | (Geissmann et al., 2009; Romilly et al., 2014; Tomasini et al., 2017) |
| **RsaC** | *sodA*↓ | Inhibits the synthesis of the Mn-dependent superoxide dismutase SodA synthesis and favors the oxidative stress response mediated by SodM. Other putative targets of RsaC are involved in oxidative stress (ROS and NOS) and metal homeostasis (Fe and Zn). | (Lalaouna et al., 2019) |
| **RsaD** | *alsS* ↓ | RsaD directly regulates *alsS* (acetolactate synthase) mRNA, a positively regulated CodY target that play a role in carbon utilization RsaD also contributes to stationary phase cell death during exposure to weak acid stress. | (Augagneur et al., 2020) |
| **RsaE (100 nt)** | *25 genes* ↓ *39 genes* ↑ | RsaE lead to the downregulation of 25 genes which include numerous metabolic enzymes involved in the citrate (TCA) cycle like *sucC* mRNA, encoding the succinyl-CoA synthetase and *opp-3B/opp-3A* mRNAs, coding amino acid and peptide transporter. Upregulation of 39 genes which include membrane proteins involved in peptide transport, operon for valine, leucine and isoleucine biosynthesis. | (Bohn et al., 2010; Felden et al., 2011; Geissmann et al., 2009) |
| **ArtR (345 nt)** | *sarT*↑ | Positively regulates the expression of α-toxin by directly binding to 5' UTR of transcriptional regulator *sarT* mRNA | (Xue et al., 2014) |

| | | | |
|---|---|---|---|
| **Teg41 (194 nt)** | α-PSM↑ | Increases levels of Phenol soluble modulins (PSM) and corresponding increasing in hemolytic activity | (Zapf et al., 2019) |
| **SSR42 (891 nt)** | *2* mRNA↑ *~80 mRNA*↓ | Modulates the expression of approximately 80 mRNAs in two genetically divergent *S. aureus* strains, UAMS-1 (MSSA; USA200) and LAC (MRSA; USA300).<br>Represses the expression of various known virulence factors such as protein A (*spa*), Ig G binding protein (*sbi*), α-hemolysin and Panton Valentine leukocidin (PVL).<br>Contributes to pathogenesis by mediating erythrocyte lysis, resistance to human polymorpho nuclear leukocyte killing and in a murine model of skin and soft tissue infection. | (Morrison et al., 2012b) |
| **sRNAs encoded from pathogenicity island** | | | |
| **SprA1/SprA1AS (208/60 nt)** | SprA1AS | SprA1 and SprA1AS, both forms a complex *in vivo* through base pairing interactions and prevent internal translation of the SprA1 encoded toxic delta hemolysin Peptide | (Felden et al., 2011; Pichon and Felden, 2005) |
| **SprC (152 nt)** | *atl*↓ | SprC negatively regulates ATL, a major staphylococcal autolysin and thereby reduces virulence and phagocytosis in the host | (Le Pabic et al., 2015) |
| **SprD (142 nt)** | *Sbi*↓ | SprD prevents translation initiation of sbi mRNA by an antisense mechanism and negatively regulates the expression of Sbi immune-evasion molecule impairing both the adaptive and innate host immune responses | (Chabelskaya et al., 2010) |
| **SprX (154 nt)** | *spoVG*, clfB, δ-hemolysin, IsaA, WalR | SprX negatively regulates the *spoVG* expression and thereby influencing vancomycin and teicoplanin glycopeptide antibiotic resistance.<br>SprX1positively regulates the expression of clumping factor B (ClfB), delta hemolysin and the autolysin regulator WalR and negatively regulates the | (Buchad and Nair, 2021; Eyraud et al., 2014; Kathirvel et al., 2016) |

| | | | |
|---|---|---|---|
| | | immunodominant protein, IsaA. | |
| **SprF1/ SprG1 (141/312 nt)** | SprG1↓ | *cis*-antisense encoded SprF1, negatively regulates the expression of SprG1 RNA which encodes the two toxic antimicrobial and hemolytic peptides and thereby prevents the mortality of *S. aureus* | (Pinal-Marie *et al.*, 2014) |

↑ indicates up-regulation, ↓ indicates down-regulation.

### 1.9 Virulence factors of Staphylococcus aureus

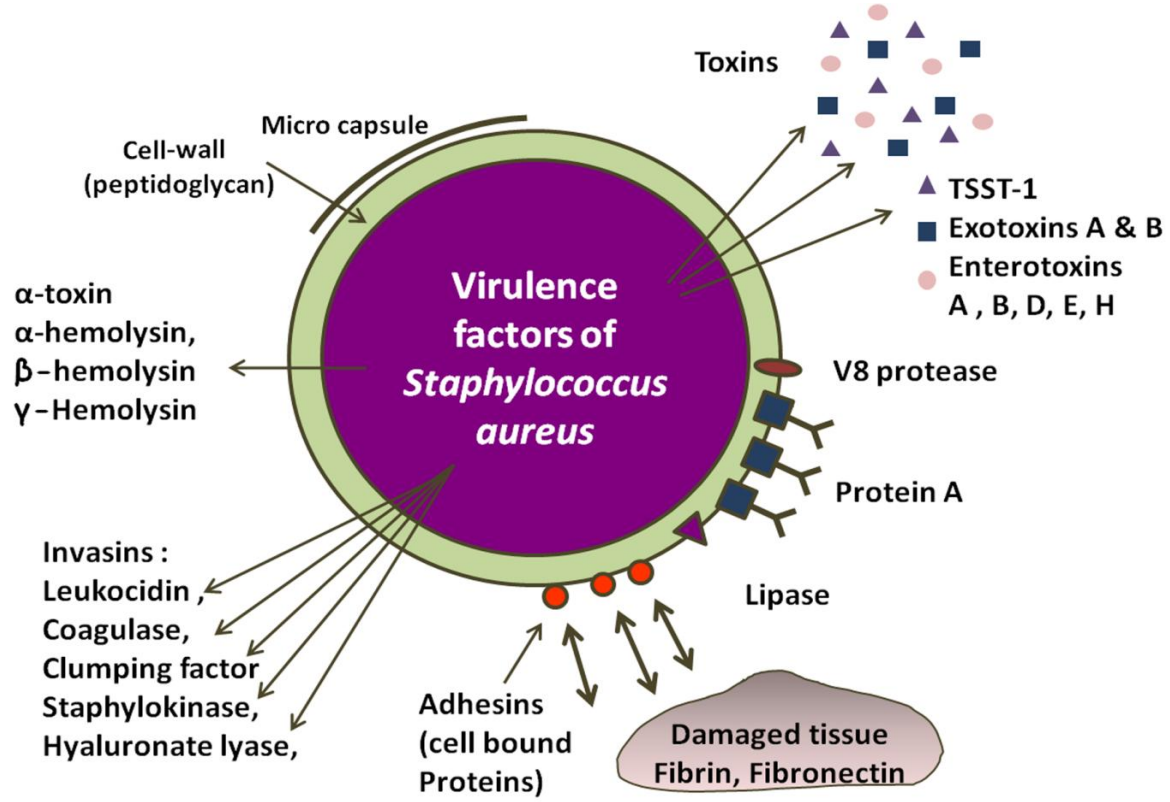


**Figure 7 : Virulence factors of *Staphylococcus aureus***

These virulence factors are categorized according to their functions, (i) surface proteins that promote adhesion, internalization, and colonization; (ii) toxins and enzymes that promote tissue damage, inflammation, invasion and dissemination; (iii) surface factors that affect phagocytosis (iv) factors that enhance survival in phagocytes; (v) super antigens and other molecules that modulate the immune system by altering the function of lymphocytes and antigen presenting cells (Malachowa et al., 2011).

*S. aureus* has an extraordinary repertoire of virulence factors that allow it to survive extreme conditions within the human host. Bacterial adherence and interaction with the host cell is a crucial first step in the initiation of colonization and infection by *S. aureus*. The ability of *S. aureus* to adhere to host matrix is attributed to the regulated expression of surface-associated

adhesions and secreted enzymes and toxins that damage the host cell and helps in establishment of a successful infection (Figure 7).

*S. aureus* virulence factors and their functions are listed in Table 4 as reviewed in (Clarke et al., 2006; Lin and Peterson, 2010; Malak et al., 2020; Nwokediuko and Adeleye, 2019; Tam and Torres, 2019).

**Table 4: Virulence determinants of Staphylococcus aureus**

| | **Virulence factors** | **Functions** |
|---|---|---|
| **Structural components** | Capsule | Inhibits chemotaxis and phagocytosis; inhibits proliferation of mononuclear cells |
| | Slime layer | Facilitates adherence to foreign bodies; inhibits Phagocytosis |
| | Teichoic acid | Binds to fibronectin |
| **Surface adhesions** | Protein A (Spa) | Inhibits antibody-mediated clearance by binding to IgG |
| | Clumping factor (ClfA/B) | Colonization factor, mediates adhesion of fibrinogen and fibronectin |
| | Fibronectin binding protein (FnBPA/B) | Responsible for adhesion and internalization; important for *in vitro* and *in vivo* infections resulting in septic death |
| | *S. aureus* surface protein G (SasG) | Role in adherence and biofilm formation |
| **Secreted factors** | Coagulase | Converts fibrinogen to fibrin |
| | Hyaluronate lyase | Hydrolyses hyaluronic acids in connective tissues, promoting the spread of staphylococci in tissues |
| | Alpha-hemolysin | Lysis of red blood cells |
| | Lipases | Inactivate fatty acids |
| | Serine proteases; cysteine proteases (including staphopains); aureolysin | Inactivate neutrophil proteolytic activity; inactivate antimicrobial peptides |
| | Staphylokinase (Sak) | Plasminogen activation; inactivate antimicrobial peptides |
| | Toxic shock syndrome toxin 1 (TSS); staphylococcal enterotoxins (SeaA-F) | Activate T cells and macrophages |
| | Cytolysins (α-,β-,γ-,δ-toxins); leukocidins | Induce apoptosis and toxic for various cell types, including erythrocytes, lymphocytes, |

| | | |
|---|---|---|
| | (LukD/E) or Panton valentine leukocidin (PVL) | monocytes, epithelial cells |
| | Chemotaxis inhibitory proteins | Inhibit the chemotaxis of neutrophil and monocytes |
| | Exfoliative toxin | activate T cells |

### 1.9.1 Surface associated virulence factors

*S. aureus* surface-associated virulence factors aid in bacterial adhesion, colonization, tissue invasion, and host defense evasion. Adherence to the host matrix is mediated by the surface adhesins, known as MSCRAMMs (for Microbial Surface Components Recognizing Adhesive Matrix Molecules) (Foster et al., 2014).

#### 1.9.1.1 Staphylococcal protein A (SpA)

SpA is a surface protein encoded by *spa* gene and plays a multifaceted role in survival and virulence of *S. aureus*. By binding the Fc portion of antibodies, protein A can prevent phagocytosis of the bacteria by sequestering antibodies (Falugi et al., 2013). SpA facilitates the adherence and act as a super antigen by binding to the B-cell receptors promoting apoptosis and reduces antibody secreting B-cell (Brignoli et al., 2019).

#### 1.9.1.2 Clumping factor A and B

*S. aureus* expresses two structurally related surface associated fibrinogen binding adhesins, clumping factor ClfA and ClfB. Clumping factor A (ClfA) is a cell-wall–anchored protein that promotes bacterial adhesion to fibrinogen and enhances the colonization ability of *S. aureus* (Herman-Bausier et al., 2018). ClfA significantly protects *S. aureus* against macrophage phagocytosis (Palmqvist et al., 2004), interacts directly with host cells or through fibrinogen bridges and also involved in the pathogenesis of endocarditis (Claes et al., 2017; Josse et al., 2017). Clumping factor B (ClfB) promotes clumping of cells, adherence to fibrinogen coated surface and facilitates skin abscess formation by binding to the host protein loricrin during nasal colonization. ClfB has been identified as promising vaccine antigen that can drive effective immune responses to protect against skin and tissue infections caused by *S. aureus* (Lacey et al., 2019).

#### 1.9.1.3 Autolysins

Autolysins are endogenous cell wall hydrolases playing important roles in processes such as cell division, autolysis, peptidoglycan recycling, and biofilm formation. Out of many autolysins produced by *S. aureus*, Atl, a major autolysin involved in bacterial cell wall

degradation and cell dispersion is the most important. Autolysins are reported to mediate staphylococcal attachment to the host extracellular and plasma proteins and biofilm formation (Binsker et al., 2018; Osipovitch et al., 2015; Porayath et al., 2018).

#### 1.9.1.4 Biofilms

Biofilm matrix is a community of cells that encases all of the cells in the mature protective extracellular polymeric structure, It is composed of host factors, secreted and lysis-derived proteins, polysaccharide, and eDNA (Lister and Horswill, 2014). Biofilm protects the bacteria from phagocytic attacks and is a major defense mechanism of *S. aureus* formed on host tissue and medical implants such as stents, ventilators, artificial heart valves, catheters and joint prosthetics, aspirators, pacemakers, stitch materials, and orthopedic devices which are in direct contact with blood causing chronic infections (Archer et al., 2011; Cheung et al., 2021).

*S. aureus* is the leading species in biofilm associated infection domain representing 80% of the nosocomial infections (Reffuveille et al., 2017). A major constituent of the *S. aureus* biofilm matrix is polysaccharide intercellular adhesin (PIA), also known as polymeric N-acetyl-glucosamine (PNAG) that is produced by enzymes encoded in the *icaADBC* locus.

### 1.9.2 Secreted virulence factors

The pathogenesis of *S. aureus* infections depends on the secretion of a series of extracellular toxins, and enzymes that destruct host cells and tissues facilitating spreading, nutrient uptake and manipulate the host's immune responses. Secreted virulence factors that play an important role in *S. aureus* infection include different cytokines, hemolysins, extracellular enzymes, and proteases (Kong et al., 2016).

#### 1.9.2.1 Staphylokinase

Staphylokinase (Sak) act as a cofactor to promote bacterial dissemination by hijacking host plasmin to activate plasminogen for the breakdown of fibrin clots. It facilitates *S. aureus* invasion of the skin barrier to generate large and open lesions. Sak also reduces biofilm formation and facilitates the detachment of mature biofilm by activating plasminogen (Kwiecinski et al., 2016; Peetermans et al., 2014).

#### 1.9.2.2 Hemolysins

Hemolysins are pore-forming toxins (PFT) secreted by *S. aureus*, which are able to cause the lysis of host cells by producing pores on the cytoplasmic membrane. *S. aureus* encodes four types of hemolysins α, β, γ and δ. α-hemolysin, also known as α-toxin, encoded as *hla*, is

extensively studied cytotoxin which damages a wide range of host cells (Tam and Torres, 2019). It enhances phagocytosis and the intracellular killing of neutrophils, and facilitates the pathogenesis of pneumonia in *S. aureus*/MRSA infections (Wardenburg et al., 2007). β-hemolysin has $Mg^{2+}$ dependent sphingomyelinase activity that destabilizes the lipid bi-layer and membrane fluidity. β-hemolysin plays a role in the recurrence of chronic osteomyelitis, respiratory infections, regulates biofilm formation and enhances *S. aureus* induced keratitis (Divyakolu et al., 2019; Huseby et al., 2010; Katayama et al., 2013). γ-hemolysin is hemolytic to rabbit erythrocytes and leukocytes (Vandenesch et al., 2012) and is associated with bacteremia ,septic arthritis and pathogenesis of toxic shock syndrome (Divyakolu et al., 2019; Kong et al., 2018).

δ-hemolysin acts as a surfactant to disrupt the cell membrane and exhibit wide spectrum of cytolytic activity. δ-hemolysin has a significant role in the development of skin disease, atopic dermatitis, due to mast cell degranulation (Nakamura et al., 2013).

#### 1.9.2.3 Hyaluronate lyase

Hyaluronate lyase is an extracellular enzyme that primarily degrade hyaluronic acid (HA), (Kreil, 1995; Stern and Jedrzejas, 2006). These enzymes are composed of two large, α-domain and β-domain which are joined by a short linker to form a pocket where HA binds and is cleaved. A number of organisms catabolize HA as a sole carbon source during infection (Hirayama et al., 2009; Marion et al., 2012; Starr and Engleberg, 2006). *S. aureus* is the only staphylococcal species that secretes hyaluronate lyase enzyme and the enzyme is closely related to those produced by *Streptococcus pneumoniae* and *Streptococcus agalactiae* (Hart et al., 2009; Hynes and Walton, 2000). Hyaluronidase or hyaluronate lyase, encoded by the gene *hysA*, is conserved across all clonal lineages of *S. aureus*. Hyaluronate lyase cleave the hyaluronic acid at the β-1,4 glycosidic bond to yield glucuronic acid and N-acetyl glucosamine as the final product as shown in Figure 8. In *S. aureus* the expression of *hysA* is positively regulated by global virulence regulator Agr, and negatively regulated by CodY and Sar (Brinsmade et al., 2010; Bronner et al., 2004; Ibberson et al., 2014).

Hyaluronate lyase have shown functions in biofilm formation and the increased activity of hyaluronate lyase results in loss of structural integrity and necrosis of the tissues, allowing pathogen access to deeper tissues. Hyaluronic acid present in the extracellular matrix serving as a structural component of *S. aureus* biofilm is degraded by HysA (Ibberson *et al.*, 2016). As a "spreading factor" of *S. aureus,* hyaluronate lyase is implicated in the dissemination of bacteria through breaking down HA in extracellular matrix and biofilms. Deletion of *hysA*

resulted in reduced skin and lung pathology and lowered bacterial burden in skin and lung infection models (Hart et al., 2013; Ibberson et al., 2014).

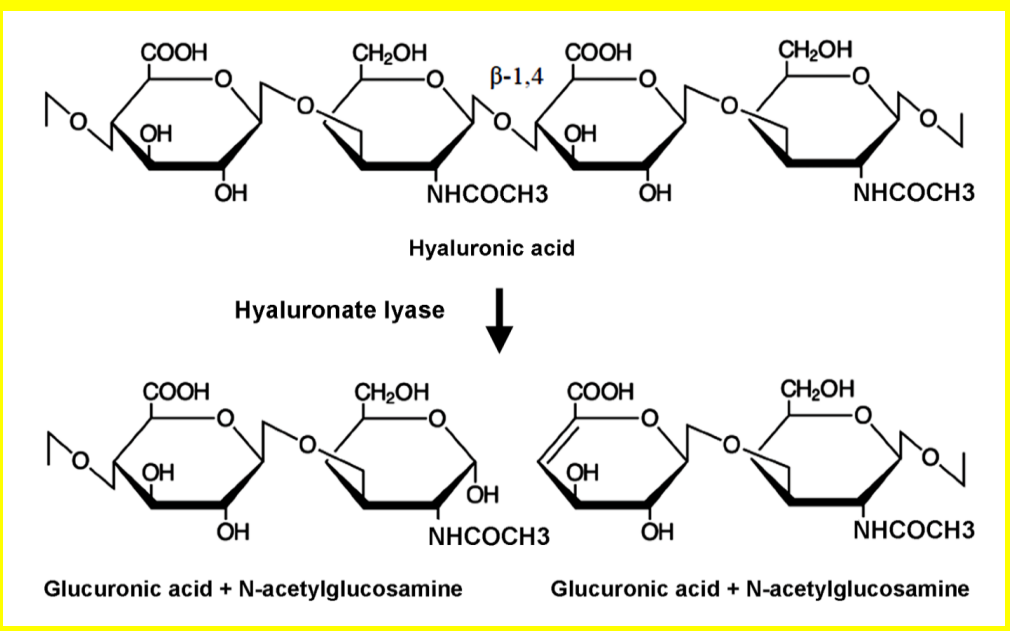


**Figure 8: Breakdown of hyaluronic acid by bacterial hyaluronate lyase.**

### 1.9.2.4 Proteases

The major proteolytic enzymes secreted by Staphylococci consist of a metalloprotease (aureolysin, Aur), two cysteine proteinases referred to as staphopain (ScpA) and the cysteine protease (SspB), and serine proteases (SspA and SplABCDEFF). A primary function of the secreted proteases of *S. aureus* is to control the progression of infection by selectively modulating the stability of virulence factors (Bose et al., 2014; Gimza et al., 2019). Consequently, proteases of *S. aureus* play an important role in the post-translational regulation, in addition to modifying host proteins to the benefit of the bacteria. Each protease have diverging roles in the infection process, including inactivation of the protease inhibitors and antimicrobial peptides in the host, modulation of kinin and chemokine synthesis, degradation of immuno globulins and complement cascade proteins, modification of the bacterial surface, interactions with components of the coagulation and fibrinolysis pathways, etc. (Zdzalik et al., 2013). Aurolysin is a metalloprotease that target diverse substrate including *S. aureus* proteins and host protein, involved in immune defense and phagocytosis (Laarman et al., 2011). SspA and SspB are serine protease and Staphopain A (ScpA) (also known as V8) is cysteine protease that modulates the biofilm formation by reducing bacterial adhesion and enhances bacterial dissemination (Stach et al., 2018).

- **Serine protease like protein (*spl*)**

Spl operon is a cluster of six serine protease-like genes (SplABCDEF), which encode for putative proteases with similarity to staphylococcal glutamyl endopeptidase (V8 protease) (Popowicz et al., 2006). The six *spl* genes in *spl* operon are co-transcribed, during the early stationary growth phase, and their expression is regulated by the global accessory gene regulator (Agr). SplA, SplB, and SplC are biochemically and structurally characterized Spl proteases (Zdzalik et al., 2013). *Spl* operon is located on a pathogenicity island νsaβ, and close to virulence genes enterotoxins and leucocidins (Baba *et al.*, 2008; Paharik *et al.*, 2016). SplD acts on several olfactory receptors which are trans membrane proteins expressed in nares, the primary colonization niche of *S. aureus* (Zdzalik *et al.*, 2013).

## 1.10 Principal genetic regulations involved in *S. aureus* virulence

The regulation of staphylococcal virulence factors is controlled at different stages by an intricate network of two-component signaling systems, transcription regulators including alternative sigma factor $\sigma^B$, RNA binding proteins and small RNAs.

### 1.10.1 Two-component system

*S. aureus* utilize two-component signaling systems to sense and respond to changes in their surroundings by changing gene expression and adapt to hostile stresses by shifting its metabolism. The extracellular stimuli is sensed by a histidine kinase (HK) resulting into auto phosphorylation that transfers phosphoryl group to a conserved aspartate residue on the response regulator (RR). Phosphorylation of RR promotes conformational changes that modify the activity of DNA binding and regulates the expression of multiple targets (Rapun-Araiz et al., 2020). *S. aureus* encodes 16 TCSs which are involved in response to environmental stimuli such as pH, stress, nutrient sensing and metabolism. TCSs listed in table 5 play a significant role in regulation of virulence, antibiotic resistance, autolysis, biofilm development, and cell death by fine-tuning the gene expression (Haag and Bagnoli, 2017).

#### 1.10.1.1 AgrCA

Accessory gene regulator (Agr) encodes a quorum-sensing system and acts as a master virulence regulator in *S. aureus* (Recsei et al., 1986). The Agr system is activated when an auto inducing peptide (AIP) signal accumulates in the extracellular environment, and reaches a certain threshold (Figure 9). The main effector molecule of the Agr system is RNAII and RNAIII, (Srivastava et al., 2014). AgrA binds directly to the promoters of the alpha and beta-

phenol soluble modulins (PSM) encoding operons (Jenul and Horswill, 2019; Tan et al., 2018). The agr system represses the transcription of a number of cell wall-associated proteins (protein A, coagulase, fibronectin binding protein) and up-regulates several exoproteins (alpha-toxin, serine proteases, cysteine proteases, gamma-hemolysin, leucocidins, and lipase) (Cheung et al., 2011; Dunman et al., 2001; Morfeldt et al., 1995). Agr mutants showed smaller lesion sizes and reduced bacterial load in murine skin infection models and infective endocarditis indicating its role in the pathogenesis of staphylococcal disease (Montgomery et al., 2010; Paharik et al., 2017; Schwan et al., 2003).

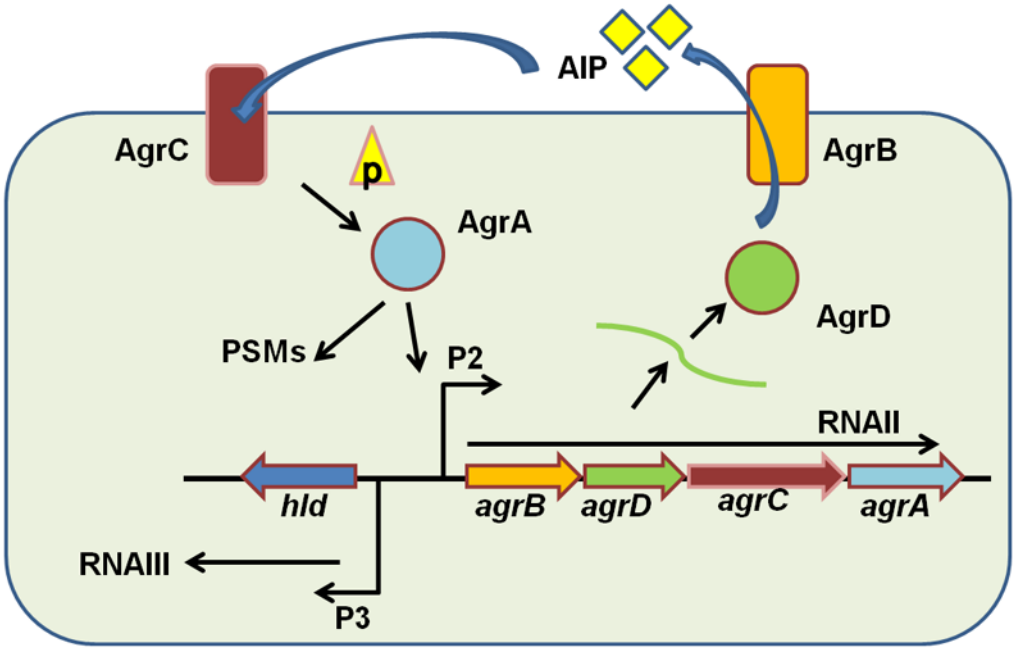


**Figure 9: A schematic representation of the molecular organization, signal biosynthesis and transduction cascade of the agr quorum-sensing system.**
The autoinducing peptide (AIP) is synthesized from the AgrD peptide. Following the modification of AgrD, AIP is transported into the environment by transporter membrane protein AgrB. When the extracellular concentration of AIP reaches a significant level, the signal is sensed by the two-component system AgrC and AgrA. AgrC is activated by the binding of AIP. AgrA undergoes phosphorylation and activates either P2 or P3 promoter. Phosphorylated AgrA also bind to the promoters of the phenol-soluble modulins (PSMs), leading to their expression. Activation of P2 regulates the expression of the *agr* operon, also referred to as RNAII, which controls the quorum sensing process. Activation of P3 regulates the expression of the sRNA RNAIII. RNAIII goes on to then regulate the expression of its target mRNA (Jenul and Horswill, 2019).

### 1.10.1.2 WalKR

The autolytic activity of major staphylococcal autolysins such as Atl and LytM are powered by WalKR (YycGF, VicKR and MicAB), denoting as its main function. WalRK activates the transcription of genes, which are involved in metabolism of cell wall and its degradation. A

decreased level of WalKR, can cause an increased resistance towards the Triton X-100 and cell lysis induced by lysostaphin (Buchad and Nair, 2021).

**Table 5 : Two-component systems involved in regulation of *S. aureus* genes**

| TCS | Positive regulation | Negative regulation | Function |
|---|---|---|---|
| **AgrAC** | *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* | Quorum sensing, cell wall and secreted protein synthesis |
| **SaeRS** | *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* | Secreted factors involved in immune evasion |
| **WalKR** | *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* | Cell wall metabolism |
| **SrrAB** | *tst, spa, and icaR* | *tst, srr* | Aerobic and anaerobic respiration |
| **ArlRS** | *sdrC, sdrD, sdrE, rot, agrA, mgrA, capH* | *spa, coa, geh, splA, splB, splC, splD, splE, splF, sspA, isdA, isdB* | Adhesion, autolysis, multidrug resistance and virulence genes |
| **GraRS** | *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* | - | CAMP sensing |

### 1.10.1.3 SaeRS

SaeRS two-component regulatory system is composed of the sensor histidine kinase SaeS, response regulator SaeR, and two auxiliary proteins SaeP and SaeQ. SaeRS TCS is essential for the production of exoproteins (e.g., α-hemolysin, β-hemolysin, nuclease, and coagulase), for adhering to and invading lung epithelial cells (Liu et al., 2016). The SaeRS TCS induces the production of localized and systemic pro-inflammatory cytokines, including tumor necrosis factor alpha, Interferon gamma and interleukins IL-6 and IL-2 (Cho et al., 2015; Zurek et al., 2014).

### 1.10.2 SarA

Staphylococcal Accessory Regulator (Sar), has been characterized as a pleiotropic transcriptional regulator of virulence factors that upregulates the expression of fibronectin and fibronectin binding proteins, hemolysins, enterotoxins, toxic shock syndrome toxin, and genes involved in oxidative stress and biofilm formation, and represses the expression of proteases, protein A, and collagen-binding proteins (Andrey et al., 2015; Morrison et al., 2012a; Oriol et al., 2021). SarA represses the expression of target proteins of small RNA SprC and srn_9340 involved in *S. aureus* virulence (Mauro et al., 2016).

### 1.10.3 Alternative Sigma factor ($\sigma^B$)

Alternative sigma factors play diverse roles in responding to stress stimuli, surviving adverse conditions, cell differentiation (Rodriguez Ayala et al., 2020). Environmental conditions like salt stress, alkaline shock, and heat shock activates $\sigma^B$ expression and modulates the gene expression of several virulence genes. $\sigma^B$ enhances bacterial adhesions by positively modulating the expression of clfA, fibrinogen-binding clumping factor and fnbA, fibronectin-binding protein A. $\sigma^B$ negatively influences expression of numerous virulence-associated exo-enzymes and toxins (Roberts et al., 2017). *S. aureus* adaptation is mediated by SigB during chronic infections that allows pathogen for long-term intracellular persistence (Tuchscherr et al., 2015).

### 1.10.4 RNases in *S. aureus*

Ribonucleases (RNases) contribute to the RNA degradation and RNA processing. In bacteria, the first step of RNA decay involves removal of pyrophosphate from the 5'end of the RNA followed by an endonucleolytic cleavage allowing exonucleolytic degradation. RNases identified in *S. aureus* are listed in Table 6. Main RNases, i.e. RNase III, RNase Y, RNase J1/J2 and PNPase, are discussed below.

#### 1.10.4.1 RNaseIII

RNase III cleaves at a stem-loop structure in the 5'UTR region of the transcript, releasing the Shine Dalgarno (SD) sequence for recruitment of ribosomes. Through its ds-specific RNase activity, RNase III is a key player in cleaving small regulatory RNA (sRNA)/mRNA duplexes (Boisset et al., 2007; Huntzinger et al., 2005; Liu et al., 2011).

RNAIII regulates the staphylococcal protein A (*spa*) that inhibits phagocytic engulfment. The regulation of *spa* involves the formation of an RNAIII-*spa* mRNA duplex that leads to

translational arrest of *spa* mRNA. This RNA-RNA duplex is then degraded by RNase III (Huntzinger et al., 2005).

**Table 6: Main RNases in *S. aureus***

| **Ribonuclease** | **Gene** | **Function** |
|---|---|---|
| RNase III | *rnc* | ds-RNA endonuclease* |
| Mini-III | *mrnC* | ds-RNA endonuclease |
| RNase Y | *rny*/*cvfA* | ss-RNA endonuclease* |
| RNase J1 | *rnjA* | Strong 5'-3' exonuclease activity*<br>ss-RNA endonuclease |
| RNase J2 | *rnjB* | Weak 5'-3' exonuclease activity<br>ss-RNA endonuclease |
| RNase P | *rnpA* | Endonucleolytic cleavage of RNA,<br>Removing 5'- extranucleotides from tRNA precursor with *rnpB* ribozyme* |
| RNase Z | *Rnz* | Endonucleolytic cleavage of RNA involved in removing extra 3' nucleotides<br>from the tRNA precursor¤ |
| RNase M5 | *rnmV* | ds-RNA endonuclease,<br>maturation of 5S rRNA |
| PNPase | *pnpA* | 3'-5' Exonuclease* |
| RNase R | *Rnr* | 3'-5' Exonuclease |
| YhaM | *yhaM* | 3'-5' Exonuclease |
| RNase HI | *ypqD/rnhA* | RNase HI-family protein of unknown function |
| RNase HII | *rnhB* | Endonuclease,<br>degradation of RNA/DNA duplexes |
| RNase HIII | *rnhC* | Endonuclease,<br>degradation of RNA/DNA duplexes |

*demonstrated experimentally

### 1.10.4.2 RNase Y, RNase J1 / J2

Single-strand endonuclease RNase Y (*rny*) and two paralog RNases named RNase J1 and RNase J2 were discovered in Gram-positive bacteria such as *B. subtilis* or *S. aureus* (Bonnin and Bouloc, 2015). RNase J1 and RNase J2, encoded by *rnjA* and *rnjB*, respectively have the 5' to 3' exonuclease activity. RNase Y has shown to enhance the expression of a global regulator SaeS/SaeR TCS (Marincola et al., 2012; Raj et al., 2020). The phosphodiesterase activity of RNaseY is essential for the expression of several virulence genes of *S. aureus* such as *lukD* (leucotoxin) and *hlgC* (Gamma hemolysin component C) (Cho, 2017).

#### 1.10.4.3 PNPase

Polynucleotide phosphorylase (PNPase) encoded by *pnpA* gene is a bifunctional enzyme having 3' to 5' exonuclease activity that uses inorganic phosphate $P_i$ instead of $H_20$ to degrade RNA releasing nucleotide diphosphate (Deutscher and Li, 2000). PNPase play a significant role in in cold shock adaptation (Anderson and Dunman, 2009). PNPase and RNase Y regulate the turnover of mRNAs involved in virulence but in opposite manner. The other RNases and its main function are listed in the table 6 as reviewed in (Bonnin and Bouloc, 2015).

### 1.11 Potential use of bacterial sRNAs

sRNAs are key regulatory elements that are currently aimed for designing novel therapeutic strategies. The development of RNA-based antimicrobial strategies requires the understanding of the factors involved in the mechanism and activity of each RNA element, their specificity for the targets, and the impact caused by introducing these RNAs to the host (Di Noto et al., 2019).

#### 1.11.1 As biomarkers

sRNAs have potential use as biomarkers in diagnosis and detection of the severity of infection. When *S. aureus* isolates from infected patients were studied, RNAIII expression was found lower in strains isolated from patients with septic shock syndrome than in patients with chronic infections. SprD levels in concurrence with RNAIII expression distinguished colonized patients from patients with blood stream infections (Bordeau et al., 2016). sRNAs like RsaA are responsible for the change from acute to chronic infections of *S. aureus* (Romilly et al., 2014).

#### 1.11.2 Targeting sRNAs for antimicrobial therapy

Regulatory RNAs act at the post-transcriptional level to control bacterial physiology, development, virulence and resistance to antibiotics, make these sRNAs promising targets for antimicrobial chemotherapy (Dersch et al., 2017; Oliva et al., 2015). The current approaches for developing RNA-based antimicrobials aim for specific targets that have little or no effect on the host microbiota (Di Noto et al., 2019).

Riboswitches and thermoregulators are known to control the expression of an adjacent mRNA upon sensing physical or chemical signals. The guanine riboswitch, have shown potential results as targets for antimicrobial compounds against the pathogen *Clostridioides difficile* (Yan et al., 2018).

In an antisense therapy, the known causative genes for a particular disease are inactivated by a small DNA or RNA oligo of 14-25 nt to inhibit transcription or translation. Antisense drugs have been studied for the treatment of diabetes, cancer, amyotrophic lateral sclerosis, muscular dystrophy, asthma and arthritis. The use of bioinformatics, high-throughput sRNA screening, whole-genome transcriptomics and proteomics, coupled with existing robust molecular characterization methods can provide information regarding production, regulation and pathogenic implications of sRNAs. These analyses can be used to design potential therapeutics based on sRNA-complementary peptide nucleic acids (PNAs). The information about sRNA expression can be used to develop a ultrasensitive diagnostic system, which has been conceived for detection of small target samples at extremely low concentrations in a very short time (Mraheil et al., 2010).

For example, the synthesis of a key metabolite GlcN6P by enzyme GlmS involved in cell envelope synthesis is regulated by two homologous sRNAs GlmY and GlmZ in enteric bacteria. The depletion of GlcN6P stimulates the synthesis of sRNAs, which provides protection against antibiotics such as bacilysin that act by inhibition of GlmS. The bactericidal potency of GlmS inhibitors can be enhanced by co-application of a nonmetabolizable GlcN6P analog, which will suppresses the accumulation of GlmY/GlmZ sRNAs (Khan et al., 2016).

Another example is *E. coli* sRNA RybB, which negatively regulates the mRNA encoding the crucial biofilm regulator CsgD. Epigallocatechin gallate (EGCG), a polyphenol present in green tea, activates *rybB* expression, which abolishes biofilm formation and expose the bacteria against antibiotics and host defenses, making EGCG a promising adjuvant that increases antibiotic susceptibility in combined chemotherapy (Serra et al., 2016).