EXPERIMENTAL RNOMICS: THE IDENTIFICATION AND CHARACTERIZATION OF SMALL NON-PROTEIN-CODING RNA (npcRNA) IN *Vibrio cholerae*

HOE CHEE HOCK

UNIVERSITI SAINS MALAYSIA

2013

EXPERIMENTAL RNOMICS: THE IDENTIFICATION AND CHARACTERIZATION OF SMALL NON-PROTEIN-CODING RNA (npcRNA) IN *Vibrio cholerae*

by

HOE CHEE HOCK

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

July 2013

ACKNOWLEDGEMENTS

It would not have been achievable to compile this doctoral thesis without the assistance and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I would like to express my sincere and profound gratitude to my main supervisor, Assoc. Prof. Tang Thean Hock for his constructive criticism, patience, support, and guidance during all times of my work. I am grateful for the chance given to me to work with his enthusiastic group. His capacity to combine critique with an immediate empathy and commitment towards his students and friends engaged in struggle will always be inspiring.

I would like to thank my co-supervisor, Prof. Manickam Ravichandran for his advice and support, as well as providing the bacteria strain to work with.

I am extremely grateful to Prof. Juergen Brosius, Dr. Timofey S. Rozhdestvensky (my field supervisor), Carsten A. Raabe, Gerrit Randau, and Thomas Robeck from Uni. Muenster, Germany. The dynamic collaboration with this group of researchers as well as their hospitality offered during my short attachment to the ZMBE, Uni. Muenster, Germany in 2009 is greatly appreciated.

I am grateful to Prof. Matthew K. Waldor from Howard Hughes Medical Institute (HHMI) and Prof. Andrew Camilli from HHMI and Tufts University School of Medicine, for donating the VC NHfq strain and pJML01 plasmid, respectively

I am indebted to Universiti Sains Malaysia (USM) and Advanced Medical & Dental Institute (AMDI) for providing USM Fellowship Scheme and AMDI Postgraduate Reseach Scheme, respectively during my doctoral studies. I am grateful as well to IPS, USM for providing IPS Student Research Grant for Research

Attachment, which made my research attachment at Institut für Experimentelle Pathologie (ZMBE), Uni. Muenster, Germany during Feb-April 2009 a dream came true.

I would like to thank all the management staffs in AMDI, especially Bursary Department, Academic Office and Reseach Secretariat. The presence of helpful staffs such as Mdm. Norhayati Mohd. Salleh and Mdm. Jamnah Ahmad from Bursary Department; Ms. Jusrina Muhammad from Academic Office; Mdm. Salmiah Saad from Research Secretariat has rendered me with invaluable experiences as well as interesting life outside of the laboratory.

I would like to thank all the family members of Infectious Disease Cluster of AMDI, in particular Mdm. Siti Aminah Ahmed, Citartan Marimuthu, Lee Li Pin, Thiviyaa Othaya Kumar, Nithiya Ravichantar, Cheah Hong Leong, Priya Kaniappan, Toh Saw Yi, and Tan Yue Rong, for their assistance and support.

My sincere appreciations and acknowlegement are extended to all my friends, specially Dr. Heselynn Hussein, Dr. Lee Guan Serm, Dr. Hong Sok Lai, Dr. Yvonne Tee Get Bee, Dr. Tan Zi Ning, Yeo Lai Fook, Yau Chee Theng, and Zhang Yilan, for their encouragement and support through this study.

I owe my loving thanks to my parents, Hoe Ting Fong and Tan Kiaw Hiok @ Tan Kim Hiok, my sister, brother-in-law, brother, and sister-in-law for their love, support, encouragement, and understanding during the long years of my education.

Last but not least, I would like to extend my thanks to everyone who had involved directly or indirectly, in this work.

I dedicate this thesis to my loving parents, family members,

and friends

for their support and unconditional love.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF SYMBOLS AND ABBREVIATIONS

°C	degree Celsius
APS	ammonium persulfate
Arc	aerobic respiratory control
asRNAs	antisense RNAs
ATP	adenosine triphosphate
AWK	programming language for processing text-based data
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like Alignment Tool
bp	base pair
BSA	bovine serum albumin
c-di-GMP	3',5'-cyclic diguanylic acid
cAMP	cyclic adeosine monophosphate
cDNA	complementary DNA
CDS	coding sequence
CRP	cAMP receptor protein
CTP	cytidine triphosphate
CTR	C-terminal region
Da	Dalton
dCTP	deoxycytidine triphosphate
ddH_2O	de-ionized distilled water
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP	dideoxynucleotide triphosphate
DTT	dithiothreitol
EAL	protein domain with homologous function as phosphodiesterase
EDTA	ethylenediaminetetraacetic acid
ESTs	expressed sequence tags
Fe-S	iron-sulfur
FNR	fumarate and nitrate reduction
Fur	ferric uptake regulator
g	gram
g	relative centrifugal force (RCF)
G6P	glucose-6-phosphate
GGDEF	protein domain with homologous function as c-di-GMP cyclase
GTP	guanosine triphosphate
h	hour
H-NS	histone-like nucleoid-associated

H_2O	water
H_2O_2	hydrogen peroxide
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hfq	host factor for Q replication
IGR	intergenic region
IPTG	isopropyl β-D-1-thiogalactopyranoside
IS	insertion sequence
J/cm^2	Joules per square centimeter
KC	potassium chloride
KH_2PO_4	potassium dihydrogen phosphate
KOH	potassium hydroxide
kV	kilovolts
l	lit
	Luria-Bertani Lysogeny media
M	Molar
mA	milliampere
mer	length of oligonucleotide, similar to nt
mg	milligram
mg/ml	milligram per milliliter
$MgCl_2$	magnesium chloride
MH	Mueller-Hinton
min	minute
miRNAs	micro RNAs
MITEs	miniature inverted-repeat transposable elements
ml	milliliter
mM	millimolar
mm	millimeter
$MnCl_2$	manganese chloride
molar	gram molecule
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	messenger RNA
Na_2HPO_4	disodium hydrogen phosphate
NaCl	sodium chloride
NAD	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NaH_2PO_4	sodium dihydrogen phosphate
NaOAc	sodium acetate
NB	Northern Blot
NCBI	National Center for Biotechnology Information
ncRNAs	non-coding RNAs
nM	nanomolar
NO	nitric oxide

npcRNAs	non-protein-coding RNAs
nt	nucleotide(s)
NTD	N-terminal domain
OD_{600}	optical density at 600 nm wavelength
OM	outer membrane
OMP	outer membrane protein
OMVs	outer membrane vesicles
ORF	open reading frame
PAP	poly(A) polymerae
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pmol	picomole
PNPase	polynucleotide phosphorylase
ppGpp	guanosine tetraphosphate
PQS	*Pseudomonas* quinolone signal
PRE	pH-responsive RNA element
PTS	phosphotransferase system
PYO	pyocyanin
RBS	ribosome binding site
RE	restriction enzyme
Rfam	database collection of RNA families
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
RNase	ribonuclease
RNP	ribonucleoprotein
ROSE	repression of the heat shock gene expression element
rpm	revolutions per minute
rRNA	ribosomal RNA
s	second
SD	Shine-Dalgarno
SDS	sodium dodecyl sulfate
sRNAs	small RNAs
SRP	signal recognition particle
TAE	tris/acetate/EDTA
TAP	tobacco acid pyrophosphatase
TBE	tris/borate/EDTA
TCA	tricarboxylic acid
TE	Tris/EDTA
TEMED	N,N,N´,N´-tetramethylethylenediamine
TFs	transcription factors
THI box	TPP riboswitch

TI	transcriptional interference
TIR	translation initiation region
tmRNAs	transfer-messenger RNAs
TPP	thiamin pyrophosphate
tRNA	transfer RNA
U	unit
U/µl	Unit per microliter
UCSC	University of California, Santa Cruz
uof	upstream open reading frame
UTP	uridine triphosphate
UTR	untranslated region
utRNAs	untranslated RNAs
V	Volt
v/v	volume per volume
VSP-1	*Vibrio* Seventh Pandemic Island I
W	Watt
w/v	weight per volume
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside
αMG6P	α-methyl-glucoside 6-phosphate
µF	microfarad
µg	microgram
µg/ml	microgram per milliliter
µl	microliter
µM	micromolar
Ω	ohm

RNOMIK EKSPERIMENTAL:

PENGENALPASTIAN DAN PENCIRIAN "NON-PROTEIN-CODING RNA"

(npcRNA) PADA *Vibrio cholerae*

ABSTRAK

RNA bukan-pengkod-protein (npcRNA) merupakan RNA yang tidak diterjemahkan kepada penghasilan protein. Sebaliknya, ia memainkan peranan dalam mengawal-atur pelbagai proses dalam sel yang merangkumi tindak balas di bawah tekanan serta kevirulenan bakteria. Tesis ini melaporkan pengenalpastian dan pencirian npcRNA daripada agen etiologi kolera, *V. cholerae* El Tor. Sebanyak 224 calon npcRNA dikenal pasti hasil daripada RNomik eksperimental. Sebanyak 92 calon npcRNA digolongkan dalam kategori "*cis*-antisense" manakala sebanyak 132 calon npcRNA digolongkan dalam kategori "intergenic". Sebanyak 56 calon npcRNA telah mencatatkan expresi positif dengan analisis pemblotan Northern. Dalam masa yang sama, satu kumpulan penyelidik lain telah mengenal pasti 627 calon npcRNA daripada *V. cholerae* El Tor N16961 dengan menggunakan teknologi penjujukan yang dikenali sebagai "RNA-Seq". Akan tetapi hanya sekadar 39 calon npcRNA yang ditemui bersama dalam kedua-dua laporan. Susulan daripada itu, kerja penyelidikan ini mendapati bahawa sumber RNA, status pemfosforilan, penambahan nukleotida "C" pada rantaian RNA, penambahan "adapter" mahupun perbezaan jujukan pada "adapter", adalah merupakan antara faktor-faktor yang menyumbang kepada perbezaan data di antara kedua-dua kumpulan. Penemuan ini menunjukkan kesahihan data dalam kajian ini dan kerja-kerja pencirian calon-calon npcRNA diteruskan. Gen bagi dua calon npcRNA daripada kategori "intergenic" telah

menunjukkan potensi sebagai penanda diagnostik, terutamanya dalam bidang epidemiologi molekul. Selain itu, pengekspresan yang berbeza bagi 9 calon npcRNA dalam pelbagai keadaan tekanan yang dihadapi oleh VC N16961 dan VC NHfq serta kajian interaksi dengan protein Hfq juga dilaporkan dalam kajian ini. Tuntasnya, kajian ini telah berjaya dalam mengenal pasti calon-calon npcRNA yang mampu membawa kepada pandangan serta idea yang bernas dalam memahami patofisiologi bagi *V. cholerae*.

EXPERIMENTAL RNOMICS:

THE IDENTIFICATION AND CHARACTERIZATION OF

SMALL NON-PROTEIN-CODING RNA (npcRNA) IN

Vibrio cholerae

ABSTRACT

Non-protein-coding RNAs (npcRNAs) are RNA that are not translated into protein but are involved in a myriad of cell regulatory processes including orchestrating bacterial general stress responses and bacterial virulence. The identification and characterization of npcRNAs from the etiologic agent of cholera, *Vibrio cholerae* (*V. cholerae*) El Tor, was reported in this study. Using experimental RNomics approaches, 224 npcRNA candidates have been identified; 92 belong to the class of putative *cis*-antisense npcRNAs, whereas 132 are the intergenic npcRNAs. Among these npcRNA candidates, differential expressions for 56 of them could be verified via Northern blot analysis. Parallelly, 627 npcRNA candidates were reported by another group in a transcriptomic profile survey in *V. cholerae* El Tor N16961 using RNA-seq. Intriguingly, only 39 npcRNA candidates were common to both datasets. A series of defined tests prompted that RNA substrate, phosphorylation status, C-tailing, adapter ligation, and different sequence of adapter are significant sources of the bias between conventional sequencing and deep sequencing. This suggested that the data in our study is valid and further characterization of the npcRNA candidates was carried out. Two intergenic npcRNA genes were developed as potential diagnostic markers, which could in a multiplex PCR assay serve to be an important tool in molecular epidemiological studies of *V. cholerae*. Differential

expression of 9 selected npcRNA candidates in various stress conditions of VC N16961 and VC NHfq background as well as interaction with Hfq protein were reported. Collectively, this study has successfully identified a plethora of novel npcRNA candidates, which set to provide valuable insights in understanding the pathophysiology of *V. cholerae*.

CHAPTER 1

INTRODUCTION

1.0 Introduction

The total RNome of a cell consists of two classes of RNAs: the RNAs that are translated into protein (messenger RNAs, or mRNAs) and various types of RNA that are not translated into protein but are involved in cell regulatory functions (non-protein coding RNAs, or npcRNAs). These regulatory RNAs are also often referred to as small non-messenger RNAs (snmRNAs), small non-coding RNAs (ncRNAs), untranslated RNAs (utRNAs), small RNAs (sRNAs), or non-protein-coding RNAs (npcRNAs), as they do not contain sizeable open reading frames (ORFs) (Brosius & Tiedge, 2004; Chinni et al., 2010; Davis et al., 2005; Heidrich et al., 2006; Hüttenhofer et al., 2002; Tang et al., 2002; Tjaden et al., 2006; Vogel et al., 2003). The term npcRNA will be used throughout this thesis.

npcRNAs are detected in all three domains of life. Collectively, they have gained momentum in changing our apprehension on the labyrinth of biological regulatory network (Modi et al., 2011). Ranging from ~20 to 400 nucleotides (Waters & Storz, 2009), npcRNAs can act as RNA itself or in association with accessory proteins in ribonucleoprotein (RNP) complexes. They are often involved in a myriad of cell regulatory processes, *i.e.* transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability and translation, growth phase developmental regulation, as well as protein degradation and translocation (Gottesman & Storz, 2011; Hershberg, 2003; Massé et al., 2003; Storz, 2002).

In bacteria, a plethora of npcRNAs have been found to be key players in general stress responses and bacterial virulence (Papenfort & Vogel, 2010; Storz et al., 2005; Storz et al., 2011; Waters & Storz, 2009). These npcRNAs facilitate the fast adaptation to changing environment in the host and instant switch from saprophytic to virulent lifestyle (Shimoni et al., 2007). The first plasmid-encoded antisense RNA, RNA I (~108 nucleotides), was discovered about three decades ago and was found to inhibit ColE1 plasmid replication by blocking the primer formation (Stougaard et al., 1981; Tomizawa et al., 1981). In 1983, identification of another antisense RNA, transcribed from the pOUT promoter of the Tn10 transposon that represses transposition was reported (Simons & Kleckner, 1983). The first chromosomally encoded antisense RNA regulator, MicF RNA, was reported in 1984; it inhibits translation of the mRNA encoding the major outer membrane porin OmpF (Mizuno et al., 1984).

Since then, only a handful of npcRNAs were identified fortuitously, with majority of them being housekeeping RNAs. Among them included RNase P RNA, SRP RNA, and tmRNA which are involved in tRNA maturation, protein translocation and ribosome rescue, respectively (Wassarman et al., 1999). In 2001-2003, experimental RNomics approach has been employed to identify npcRNA genes in several eukaryotic organisms (Hüttenhofer et al., 2001), in the archaeon *Archaeoglobus fulgidus* (Tang et al., 2002a; Tang et al., 2002b), and in *Escherichia coli* (Vogel et al., 2003). During that period, many new npcRNAs have also been reported in *E. coli* based on bioinformatics studies (Argaman et al., 2001; Chen et al., 2002; Rivas et al., 2001; Wassarman et al., 2001). To date, a burgeoning list of npcRNAs have been identified and predicted in wide range of bacteria due to

technical advancement in various screening methods (Altuvia, 2007; Pichon & Felden, 2008; Sharma & Vogel, 2009).

In bacteria, a large number of npcRNAs exist as relatively short transcripts (~50–400 nucleotides) (Waters & Storz, 2009). They mostly function in sensing the environmental cues, *i.e*. differences in pH, temperature, and the availability of nutrient and metabolite, thus reacting rapidly in order to survive in an often hostile environment (Gripenland et al., 2010). The best characterized class of npcRNAs acts via antisense base pairing with target mRNAs. Based on their genomic location, they can be classified into: 1) *cis*-encoded npcRNAs that are located on the strand of DNA opposite to their mRNA targets and hence have extensive complementarity; 2) *trans*-encoded npcRNAs that are identified at genomic locations remote from their mRNA target and often involve less complementarity (Figure 1.1) (Brantl, 2009; Richards & Vanderpool, 2011).

1.1 Mechanistic Aspects of the *cis*-encoded npcRNAs

Bacterial *cis*-encoded npcRNAs are rather diverse although they share a common feature: they are transcribed from the opposite strand of a known transcriptional unit and thus share extensive complimentarity with the corresponding transcripts. Functional characteristics for several phage- and plasmid-encoded npcRNAs have been well established (Brantl, 2007; Wagner & Simons, 1994). Despite the advancing number of the *cis*-encoded npcRNAs reported (Georg et al., 2009; Sharma et al., 2010; Toledo-Arana et al., 2009), information on the molecular mechanism of individual *cis*-encoded npcRNAs is relatively lagging behind. Nevertheless, unique mechanisms employed by these npcRNAs will be discussed as follows.

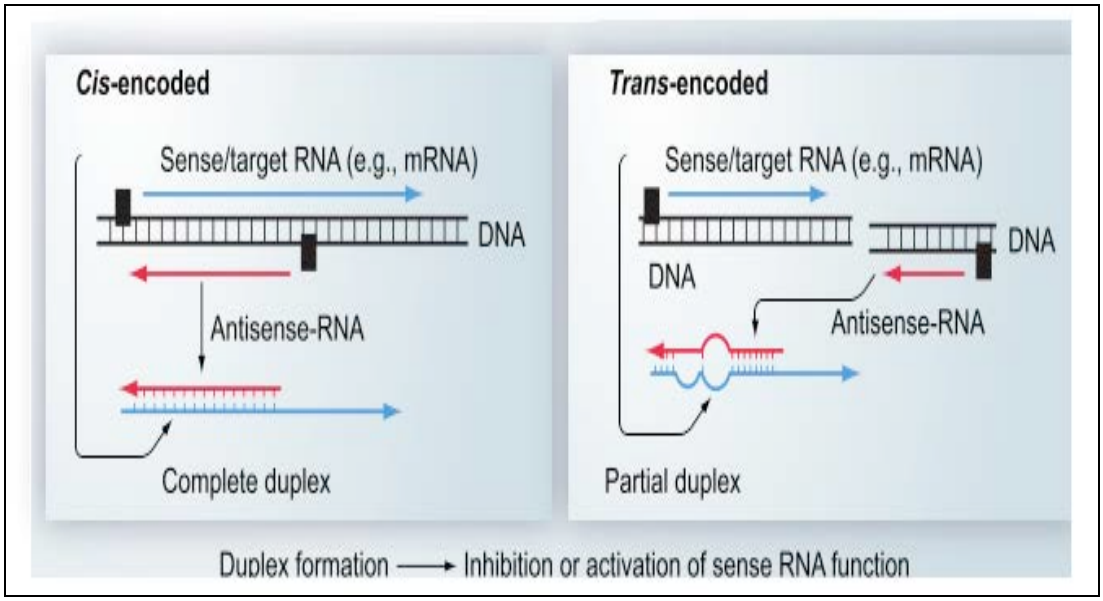

Figure 1.1: Overview of *cis*-encoded npcRNA and *trans*-encoded npcRNA. Antisense RNAs are drawn in red, sense RNAs in blue. Black rectangles denote promoters. *Cis*-encoded npcRNAs have extensive complementarity while *trans*-encoded npcRNAs usually have partial complementarity (adapted from Brantl, 2009).

1.1.1 Transcriptional Termination

The *Vibrio anguillarum* iron transport-biosynthesis operon provides an insight for this mechanism. The operon, which is located on the virulence plasmid, pJM1, consists of four ferric siderophore transport genes (*fatDCBA*) and two siderophore biosynthesis genes (*angR* and *angT*), gives rise to two antisense RNAs (asRNAs), RNAα and RNAβ (Chen & Crosa, 1996; Salinas et al., 1993). Under iron-limiting conditions, the expression level of *fatDCBA* mRNA is ~17 times higher than that of the full-length mRNA (*fatDCBA-angRT*) despite being part of the same polycistronic mRNA. The differential expression observed is due to the RNAβ encoded on the opposite strand of *fatDCBA-angRT*. The stem loop interaction between the growing polycistronic *fatDCBA* and RNAβ causes the formation of potential hairpin close to *fatA* stop codon, which leads to transcription termination (Figure 1.2B) (Stork et al., 2007).

Similarly, RnaG, an asRNA is reported to promote premature termination of transcription of *icsA* mRNA in *Shigella flexneri* (Giangrossi et al., 2010). Formation of two long hairpin structure in the 5' region of the *icsA* mRNA apparently resembles an antiterminator structure. Structural probing assays have suggested that the binding of RnaG to the actively transcribed mRNA inhibits the formation of the antiterminator but favors the formation of a terminator hairpin. This observation is further supported as the RnaG-mediated regulation of *icsA* transcription is disrupted when the hairpin structure of the proposed terminator is mutated (Giangrossi et al., 2010).

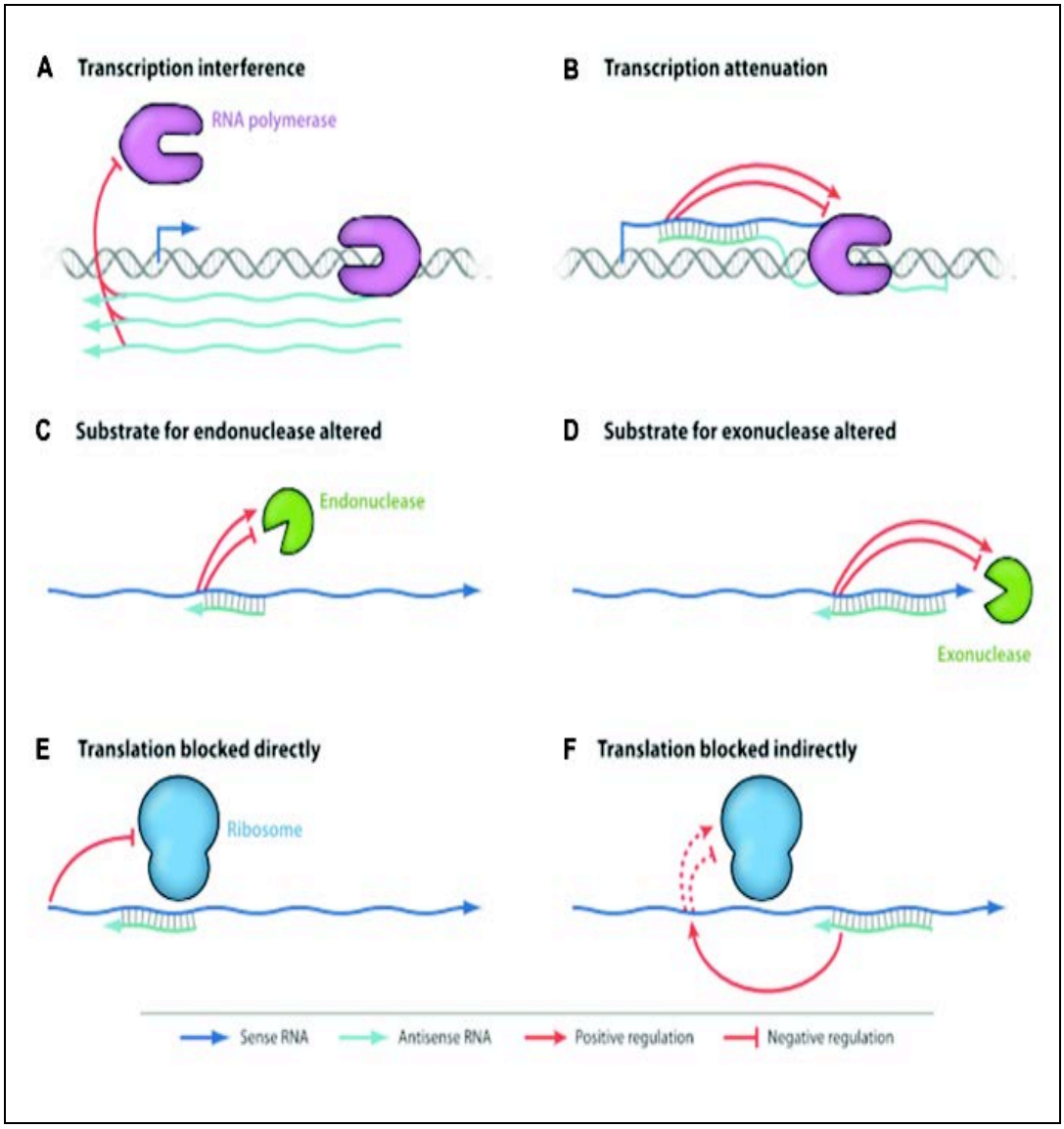

Figure 1.2: Overview of mechanisms deployed by *cis*-encoded npcRNAs. A) Antisense RNAs can induce transcription interference, where transcription from one promoter blocks transcription from a second promoter by preventing RNA polymerase from either binding or extending a transcript encoded on the opposite strand. **B)** Base pairing of the antisense RNA to the target mRNA causes changes in the target mRNA structure, ultimately causing transcription termination. **C)** and **D)** Duplex formation between *cis*-encoded npcRNA and target mRNA can affect target mRNA degradation by endonuclease or exonuclease; *cis*-encoded npcRNA can also indirectly affect the binding of the ribonuclease at a distance from the site of base pairing. **E)** Duplex formation between *cis*-encoded npcRNA and target mRNA can directly block the RBS or **F)** indirectly positively or negatively impact ribosome binding by affecting the target mRNA structure (adapted from Thomason & Storz, 2010).

1.1.2 Transcriptional Interference

The interaction between an elongating RNA polymerase (RNAP) with DNA-bound transcription factors (TFs), structural DNA-binding proteins and other RNAPs, will affect neighboring transcriptional processes. These often result in transcriptional interference (TI), the direct and in *cis* suppression of one transcriptional process by another transcriptional process (Figure 1.2A; Figure 1.3) (Shearwin et al, 2005). Arrangement of interfering promoters, which can be convergent, divergent, or tandem is instrumental in mechanisms involved in TI (Georg & Hess, 2011).

Collision between RNAPs transcribing opposite strands result from convergent promoter arrangements lead to premature termination of the transcriptional progress of one or both complexes (Shearwin et al., 2005). Through collision (Figure 1.3A), elongating RNAP forces the opposing RNAP to stall and backtrack (Crampton et al., 2006). Also, RNAPs collision has been reported to exert a suppressive effect on transcription from the weaker promoter in PR-PRE and pR-pL promoter pairs of bacteriophage λ and 186, respectively (Callen et al., 2004; Ward & Murray, 1979). Other examples of convergent transcription systems that are likely to be affected by RNAPs collision include *ubiG-mccBA* operon in *Clostridium acetobutylicum* which is controlling the conversion of methionine to cysteine (Figure 1.3A) (André et al., 2008), *prgX/prgQ* operon in *Enterococcus faecalis*, regulating conjugative transfer of the antibiotic resistance plasmid pCF10 from donor cells to recipient cells (Chatterjee et al., 2011), as well as *scbA/scbR* gene pair in *Streptomyces coelicolor*, mediating the synthesis of a signaling molecule, the c-butyrolactone SCB1 and controls the onset of antibiotic production (Chatterjee et al., 2011).

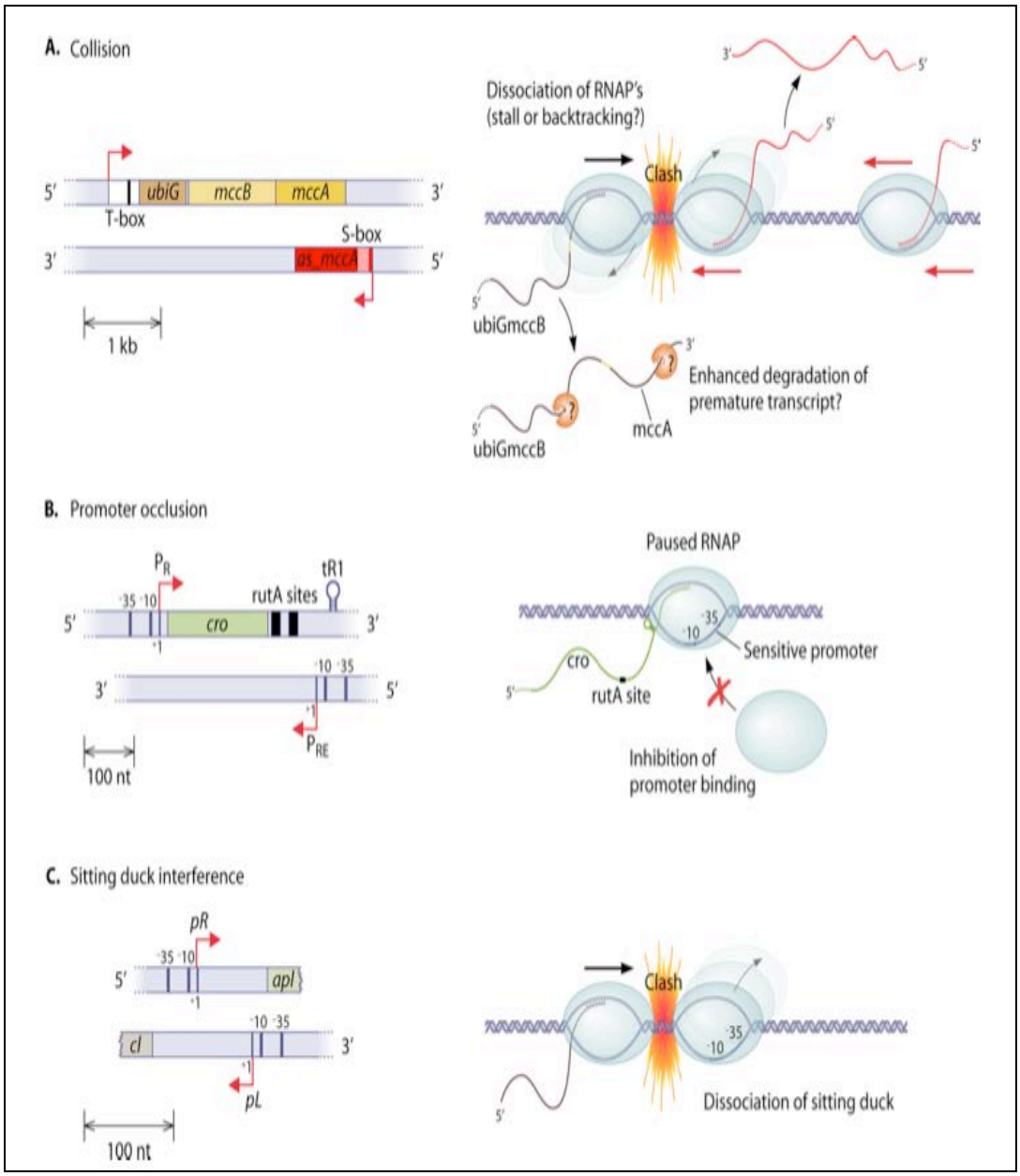

Figure 1.3: Transcriptional interference where antisense RNA is a by-product of interfering promoters. **A)** Proposed collision between the two divergently elongating RNA polymerases, transcribing the asRNA (*as_mccA*) and the *ubiG-mccAB* operon give rise to the rapid degradation of the prematurely terminated transcript and *as_mccA*. **B)** Promoter occlusion. **C)** The sitting-duck mechanism of transcriptional interference (adapted from Georg & Hess, 2011).

Sitting duck interference (in which RNAP complexes waiting to fire at the promoter are removed by passing RNAP) (Figure 1.3C), was proposed by Callen et al. (2004) in their observation that interference of the weak lysogenic promoter by the strong lytic promoter in coliphage 186 disappeared if termination occurred before the transcription from the aggressive promoter reached the sitting duck complex (Callen et al., 2004). This is due to RNA polymerase bound at an open complex of the "sensitive" promoter is rescued from being removed by the collision of another strong elongating RNA polymerase complex. Interference mechanisms are the strongest when promoters are closely spaced (but not overlapping) and of moderate strength, especially when the firing (on-rate) of RNA polymerase at the "sensitive" promoter is tantamount with the initiation rate of elongation (Sneppen et al., 2005).

Promoter occlusion occurs when the formation of an initiation complex at the "sensitive" promoter is arrested by an elongating RNAP coming from an "aggressive" promoter, which passes over a "sensitive" promoter element (Figure 1.3B). Since the elongating RNAP blocks the "sensitive" promoter region for just a transient period, efficient interference by promoter occlusion normally needs a very strong "aggressive" promoter. Indeed, the extent of the interference depends on the strength of the aggressive promoter, the size of the sensitive promoter, and the transcription acceleration across the "sensitive" promoter (Shearwin et al., 2005; Sneppen et al., 2005). All in all, it is worth noting that the work by Palmer et al. (2009) suggested that promoter occlusion together with pausing could lead to strong asymmetric interference between two promoters (Palmer et al., 2009).

1.1.3 Modulation of Translation

In *E. coli*, SymR (77 nucleotide RNA) is encoded in *cis* to the AUG start codon and the ribosome binding site (RBS) of a SOS-induced gene whose product (SymE) shows homology to the antitoxin MazE (Kawano et al., 2007). *symE*-SymR sense-antisense pairing hinders the access of 30S ribosomal subunit to RBS and prevents the initiation of translation (Figure 1.2E) (Kawano et al., 2007). Disruption of SymR RNA expression resulted in a 3-fold increase in *symE* mRNA levels, and 7-fold increase in SymE protein. Intriguingly, the high ratio of SymR to *symE* (10:1) even after *symE* induction suggests that other cellular molecules could affect SymR-*symE* base pairing or SymR could regulate other target in *trans*. In addition, the SymR-*symE* system is different from previously characterized toxin-antitoxin modules as the SymR antitoxin RNA is relatively stable while antitoxins in other cases are extremely unstable, i.e. Sok antisense RNA (Kawano et al., 2007). It remains a puzzle how this could benefit bacteria during SOS but it clearly demonstrates an example where both RNA degradation (Figure 1.2C; Figure 1.2D) and the inhibition of translation (Figure 1.2E) can co-operate in repressing a target gene.

1.1.4 Alteration of Target RNA Stability

Apart from blocking the translation of target RNA or repressing a target gene, *cis*-encoded asRNA could act to enhance the stability of a transcript. In *E. coli*, GadY RNA (105 nucleotides) is positioned at an intergenic region between *gadX* and *gadW* and encoded in *cis* to the 3'-untranslated region (UTR) of the *gadX* gene. The AraC/XylS-like transcriptional regulators GadX and GadW are involved in glutamate-dependent acid resistance (Opdyke et al., 2004). In response to acid stress, GadY will induce the cleavage of the bicistronic *gadXW* transcript, thus enhancing

the stability of the *gadX* transcript. Binding of GadY to the 3'-UTR of *gadX* could form a double stranded RNA duplex and impede digestion by RNase E (Opdyke et al., 2004). In line with this, a decrease in survival rate and decreased amounts of GadY and *gadX* transcripts were obtained in an RNase E knockout strain under acidic condition (Takada et al., 2007). In contrast, Opdyke et al. (2011) reported that RNase III and not other RNases are involved in GadY dependent cleavage, suggesting that RNase involvement in the cleavage machinery could be growth condition-dependent (Opdyke et al., 2011).

Another example showing codegradation of *cis*-encoded asRNA and target mRNA is the IsrR-*isiA* system in cyanobacterium *Synechocystis* sp. PCC 6803 (Dühring et al., 2006). IsrR asRNA (177 nucleotides) is constitutively transcribed while expression of *isiA* mRNA is induced upon iron, redox, or light stress. Since massive reorganization of the photosynthesis apparatus is the result of *isiA* expression, *isiA* expression needs to be tightly moderated. Upon iron stress, *isiA* transcripts cannot accumulate until its number surpasses the number of IsrR as RNA duplex formed is immediately degraded by an unknown mechanism (Dühring et al., 2006). As a result, a delay of IsiA protein expression during early stress and a faster recovery from stress is achieved (Legewie et al., 2008; Levine & Hwa, 2008). This regulation, also known as "threshold-linear response" (Levine & Hwa, 2008) certainly provides an energetically brilliant way in preventing unnecessary expression of stress related proteins, which could be potentially catastrophic.

1.2 Mechanistic Aspects of the *trans*-encoded npcRNAs

On the contrary to *cis*-encoded asRNA, most of the bacterial chromosomal npcRNA genes are *trans*-encoded, in which they exhibit limited base pairing to their target RNA to inhibit translation or affect the stability of target RNA. Due to the lack of full complementarity, *trans*-encoded npcRNAs in some cases depend on helper protein such as Hfq (host factor required for phage Qβ RNA replication, also known as host factor 1) to execute its regulatory function. Inhibition of translation has been the most significant mechanistic aspect for *trans*-encoded npcRNAs. Nevertheless, there have been other reported mechanisms as well which will be discussed below.

1.2.1 Translation Inhibition

The most prevalent mechanism in translation inhibition is via direct blocking of the RBS. One common feature among the reported *trans*-encoded npcRNAs in this category is the base pairing between npcRNAs and target mRNA which overlaps the RBS, or/ and the 5' or 3' regions of RBS, masking the RBS and inhibiting translation (Figure 1.4B). For example, imperfect binding of MicA and MicC to the *ompA* and *ompC* RBS respectively, has been shown to inhibit the translation of *ompA* and *ompC* (Chen et al., 2004; Udekwu et al., 2005). Apart from that, the repression of *fhlA* translation by OxyS npcRNA (Altuvia et al., 1998) is achieved by formation of two short kissing complexes overlapping the Shine-Dalgarno (SD) sequence and within the coding sequence of *fhlA* mRNA (Argaman & Altuvia, 2000).

Another strategy to inhibit translation is by blocking a ribosome standby site, which is demonstrated by IstR-1/*tisAB* of *E. coli* (Darfeuille et al., 2007). A ribosome standby site has been identified ~100 nucleotides upstream of the *tisB* RBS,

facilitating ribosome sliding to a transiently open *tisB* translation initiation region (TIR). Base pairing of IstR-1 to the ribosome standby site prevents ribosome loading. Indeed, the pairing results in 5' truncation of *tisB* due to cleavage by RNase III, blocking the standby-dependent translation of the *tisB* reading frame (Darfeuille et al., 2007).

Atypically, SR1/*ahrC* of *B. subtilis* provides yet another insight of translation inhibition by inducing structural changes downstream of the RBS. SR1 acts by base pairing with *ahrC* mRNA, which encodes the transcriptional activator of the *rocABC* and *rocDEF* arginine catabolic operons (Heidrich et al., 2006). SR1 and *ahrC* mRNA share seven complementary regions, A–G. Region G is located ~100 nucleotides downstream from the RBS of *ahrC* mRNA. Toeprinting and translational fusion analysis show that initial contact of SR1 and *ahrC* at region G induces structural alterations in 6 out of 7 complementary regions, resulting in inhibition of translation initiation (Figure 1.4B) (Heidrich et al., 2006; Heidrich et al., 2007).

1.2.2 Translation Activation

In certain cases, 5' untranslated region (UTR) of RNA folds and occludes the SD sequence and/ or AUG start codon, intrinsically inhibiting the translation initiation. The melting of this occluded region is usually achieved by the pairing of npcRNAs to the complementary region, which leads to the translation initiation (Figure 1.4A). For example, 5'-end of ~514 nucleotides RNAIII (encoded by the *agr* locus) of *Staphylococcus aureus* complements the 5' UTR of *hla* (encoding hemolysin α), and hence initiates the translation of *hla* (Morfeldt et al., 1995).

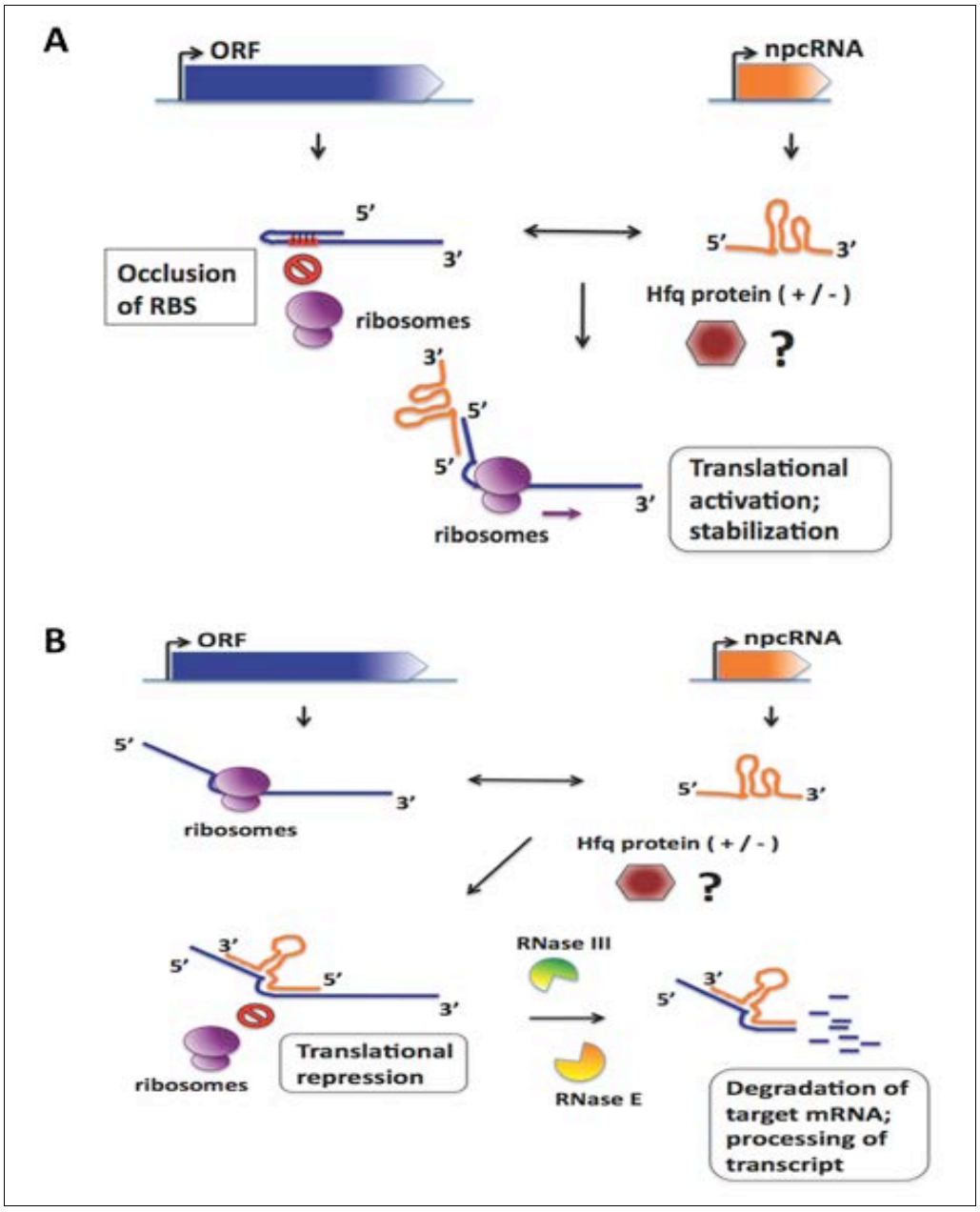

Figure 1.4: General mechanism of *trans*-encoded npcRNAs. **A)** *Trans*-encoded npcRNAs can bind to target mRNAs and relieve the occlusion of the RBS, thus allowing translation initiation. Hfq is usually involved as the binding facilitator between npcRNA and target mRNA. **B)** *Trans*-encoded npcRNA can also act negatively by pairing with the RBS of a target mRNA, thus occluding the RBS and repressing translation. Usually the non-translated mRNA is degraded via an RNase E-dependent pathway. The interaction between npcRNA and target mRNA can also induce or unmask sites for RNase III. This usually leads to degradation of both mRNA and npcRNAs. Involvement of Hfq varies from case to case.

This mechanism, also known as anti-antisense mechanism was further appreciated in translational activation exerted by GlmZ (on *glmS*, encoding glucosamine-6-phosphate) (Urban & Vogel, 2008), RyhB (on *shiA*, encoding shikimate permease) (Prévost et al., 2007), DsrA (Majdalani et al., 1998), and RprA (Majdalani et al., 2002; McCullen et al., 2010) (both on *rpoS* in response to different stress).

1.2.3 Variation of Translational Inhibition and Nuclease Mediated Degradation

In most cases, nuclease mediated cleavage of the mRNA is coupled with the repression of translation (Figure 1.4B). RNase III has been reported to be necessary for the degradation of RNAIII in *S. aureus* (Huntzinger et al., 2005). Besides activating the translation of *hla* (Morfeldt et al., 1995), RNAIII is also involved in the translational inhibition of SA1000 mRNA, *spa*, *rot* (Boisset et al., 2007), and *coa* (Chevalier et al., 2010), which encodes a novel fibrinogen-binding protein, main surface adhesion protein, pleiotropic transcriptional factor Rot, and staphylocoagulase, respectively. In all the examples mentioned, formation of RNAIII-mRNA duplexes leads to inhibition of ribosome binding and favors specific recognition by RNase III.

Interestingly, nuclease mediated degradation has also been demonstrated in certain cases where RBS is not blocked. For instance, base pairing between the *Salmonella* MicC and the *ompD* mRNA within codons 23–26 promotes RNase E-dependent decay of the mRNA without blocking ribosome binding (Pfeiffer et al, 2009). Meanwhile, base pairing between *E. coli* RyhB and the *iscRSUA* mRNA at the *iscR-iscS* intergenic region under iron depletion leads to the cleavage of the

downstream *iscSUA* transcript (Desnoyers et al., 2009). The IscR protein produced could then activate the expression of *suf* operon and overtake the Fe-S cluster biogenesis under iron-deprived conditions (Yeo et al., 2006).

1.3 *Trans*-encoded npcRNAs That Reconcile Protein Activity

Trans-encoded npcRNAs that reconcile protein activity remain global players in cellular regulation although they are the minority group compared to npcRNAs that act via base pairing. Regulation of protein with enzymatic activity was best demonstrated in the *E. coli* 6S RNA, which binds the σ^{70}-containing RNA polymerase (RNAP) (Figure 1.5) (Wassarman, 2007). During the stationary phase, the interaction between 6S RNA and σ^{70}-containing RNAP represses the transcription from σ^{70}-dependent promoters; but favors the transcription from some σ^{S}-dependent promoters. Homologs of *E. coli* 6S RNA have been identified in a number of organisms, including *Bacillus subtilis* and *Legionella pneumophila*, which express two 6S RNAs (Barrick et al, 2005; Faucher et al., 2010). Remarkably, 6S RNA serves as template for transcription of 14-20 nucleotide product RNAs (pRNAs) during the outgrowth from stationary phase (Figure 1.5) (Gildehaus et al., 2007), that is characterized by the rapid release of RNAP from 6S RNA bound complexes and decrease in 6S RNA pools (Cavanagh et al., 2012). Judging from the tight coupling of 6S RNA expression exerted to the complex network of bacterial regulators (Neusser et al., 2008) and the presence of multiple 6S RNAs, it is not surprising to encounter novel role for 6S-like RNAs in near future.

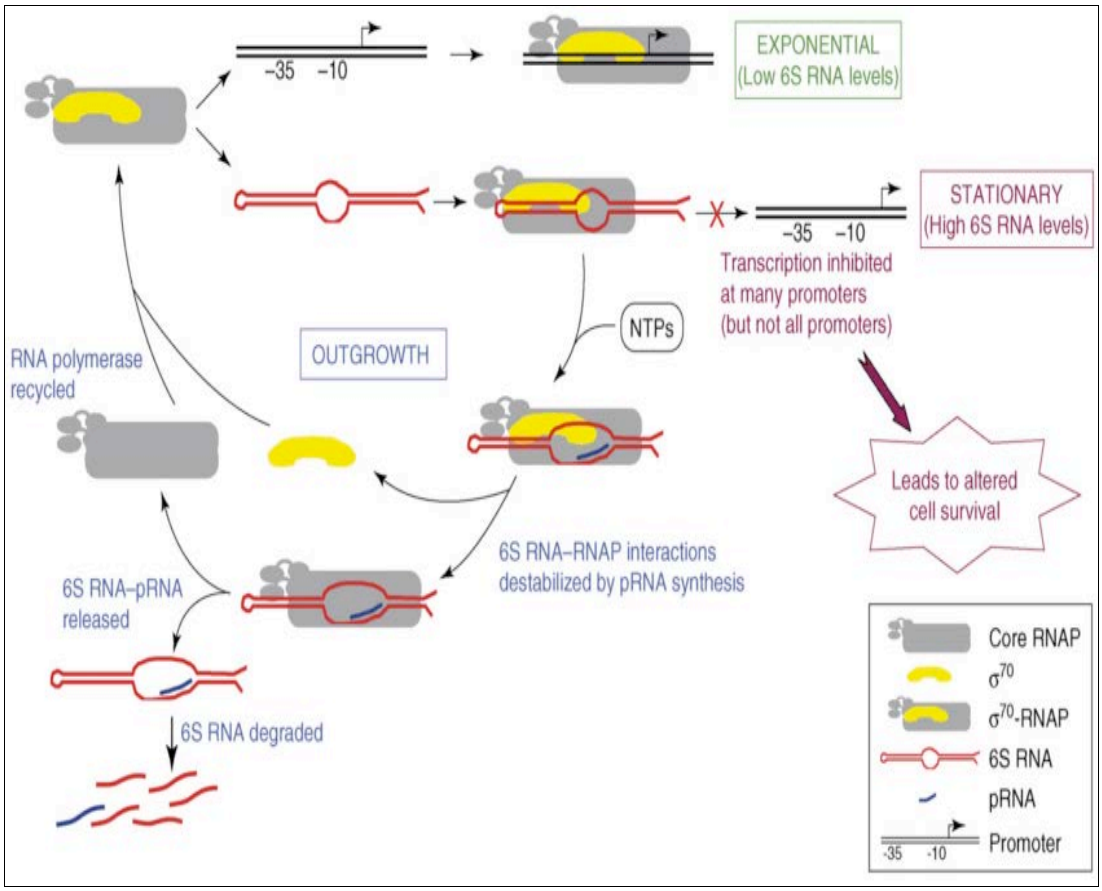

Figure 1.5: General mechanisms of 6S RNA action. During stationary phase, 6S RNA levels are high and 6S RNA–σ^{70}-RNA polymerase predominates; during exponential phase, most of the σ^{70}-RNA polymerase is bound to DNA. 6S RNA inhibition of transcription at specific promoters leads to increased competitive and long-term survival and decreased survival of stresses in stationary phase. When stationary phase cells are moved to a nutrient rich environment, they enter an outgrowth phase in which NTP concentrations increase significantly and 6S RNA is used as a template for pRNA synthesis, resulting in release and degradation of 6S RNA as well as the recycling of RNA polymerase (adapted from Wassarman, 2007).

Another major system in titrating the bacterial carbon metabolism is the Csr (carbon storage regulator) system, with CsrA as the central protein regulator (Romeo & Gong, 1993). Since the discovery of CsrB (Liu et al., 1997) and CsrC (Weilbacher et al., 2003) npcRNAs, it has become apparent that CsrB family members bind to and antagonize the effects of the CsrA, indirectly modulating the global regulatory circuits upon entry into stationary phase or nutrient limited conditions (Babitzke et al., 2009). In *E. coli*, CsrB and CsrC npcRNA carry multiple GGA motifs that mimic sequences bound by CsrA, thus effectively sequestering the CsrA protein away from mRNA leaders. Apart from the induction by two-component system BarA/UvrY (Babitzke & Romeo, 2007), these npcRNAs are also regulated by CsrD protein that directs the degradation via RNase E (Suzuki et al., 2006). Multiple CsrB-like npcRNAs have been reported to regulate CsrA-like protein in other bacterial species. For example, *Pseudomonas aeruginosa* utilizes two CsrB-like RNAs (Rsm/Y/Z) to control the quorum-sensing machinery and the expression of extracellular products via GacS/GacA two-component system (Kay et al., 2006).

Another npcRNA, GlmY has been reported to act on YhbJ, a putative RNA-binding protein encoded in the *rpoN* operon, which in turn governs the processing of another npcRNA, GlmZ (Reichenbach et al., 2008). Although both GlmY and GlmZ share high similarities in sequence and predicted secondary structure, they promote the accumulation of the GlmS (glucosamine-6-phosphate [GlcN-6-P] synthase) via a different mechanism. GlmZ activates *glmS* mRNA translation by an anti-antisense mechanism, with the facilitation by RNA chaperone, Hfq; when GlcN-6-P is depleted, GlmY accumulates and binds to YhbJ, outcompeting the binding of GlmZ, thus resulting in accumulation of full length GlmZ and activation of *glmS* expression.

The regulation of *glmS* is further complicated as polyadenylation of the GlmY by poly(A) polymerase (PAP) could also result in the degradation of GlmY by polynucleotide phosphorylase (PNPase) (Reichenbach et al., 2008; Urban & Vogel, 2008).

1.4 Correlation between Hfq and RNases

Generally, the short and imperfect base pairing of *trans*-encoded npcRNAs in Gram-negative bacteria is mediated by Sm-like hexameric RNA chaperone, Hfq (Valentin-Hansen et al., 2004; Vogel & Luisi, 2011). Hfq is a 102 amino acid protein that was first identified as a host factor required for phage Qβ RNA replication in *E. coli* (Franze de Fernandez et al., 1968). Hfq preferentially binds to an AU-rich single-stranded region upstream of the terminator (Geissmann et al., 2006; Møller et al., 2002; Zhang et al., 1998). It is now generally recognized that the Hfq N-terminal domain (NTD) presents two binding faces: the proximal face that interacts with AU-rich RNA and the distal face with specificity for poly(A) and A–R–N triplets (where R is a purine and N is any nucleotide) (Brennan & Link, 2007; Link et al., 2009). In addition to that, there have been studies reported about the affinity of Hfq towards single stranded regions that are adjacent to stem loops (Brescia et al., 2003; Sun & Wartell, 2006) as well as tRNA (Lee & Feig, 2008), implying the ability of Hfq to bind structured RNA. Growing data also suggests that poly(U) tail (Otaka et al., 2011; Sauer & Weichenrieder, 2011) and polyadenylation (Derout et al., 2003; Mohanty et al., 2004) of the npcRNA could also influence the interactions of the npcRNAs with the proximal and distal face of Hfq, respectively. Interestingly, Vincent et al. (2002) inferred that C-terminal region (CTR) of the Hfq could assist in directing the RNA to the correct face of Hfq (Vincent et al., 2012), contradicting to

the earlier idea that CTR of Hfq has no implication in riboregulation (Olsen et al., 2010). Obviously, the detailed mechanisms of Hfq (Figure 1.6) as RNA chaperone remains ambiguous and the possibility that the mechanisms are npcRNA dependent cannot be excluded.

Hfq is found in many Gram-negative and Gram-positive bacteria, as well as in the archaeon *Methanococcus jannaschii* (Valentin-Hansen et al., 2004). Intriguingly, remote Hfq homologues (from the bacteria *Neisseria meningitidis* and *Aquifex aeolicus*, and the archaeon *Methanocaldococcus jannaschii*) tend to interact with npcRNAs when expressed heterologously in *Salmonella typhimurium* (Sittka et al., 2009). Besides that, multiple Hfq proteins have been identified in *Burkholderia cenocepacia* (Ramos et al., 2011), demonstrating functional diversification of this protein.

The pleiotropy of an *hfq* deletion mutation was first apparent from the multiple stress response-related phenotypes in *E. coli* (Tsui et al., 1994). It was then reported to be partly attributed to the reduced efficiency of translation of *rpoS* mRNA, encoding the major stress sigma factor, σ^S (Brown & Elliott, 1996; Muffler et al., 1996). Nevertheless, accumulating evidence are suggesting that Hfq could widely impact bacterial physiology, including the σ^S-independent control of virulence factors in a plethora of pathogenic bacteria (Chao & Vogel, 2010; Chiang et al., 2011; Ramos et al., 2011).

The involvement of Hfq in npcRNA/mRNA interactions varies among bacteria and seems to be influenced by the overall GC-content of the genome, free

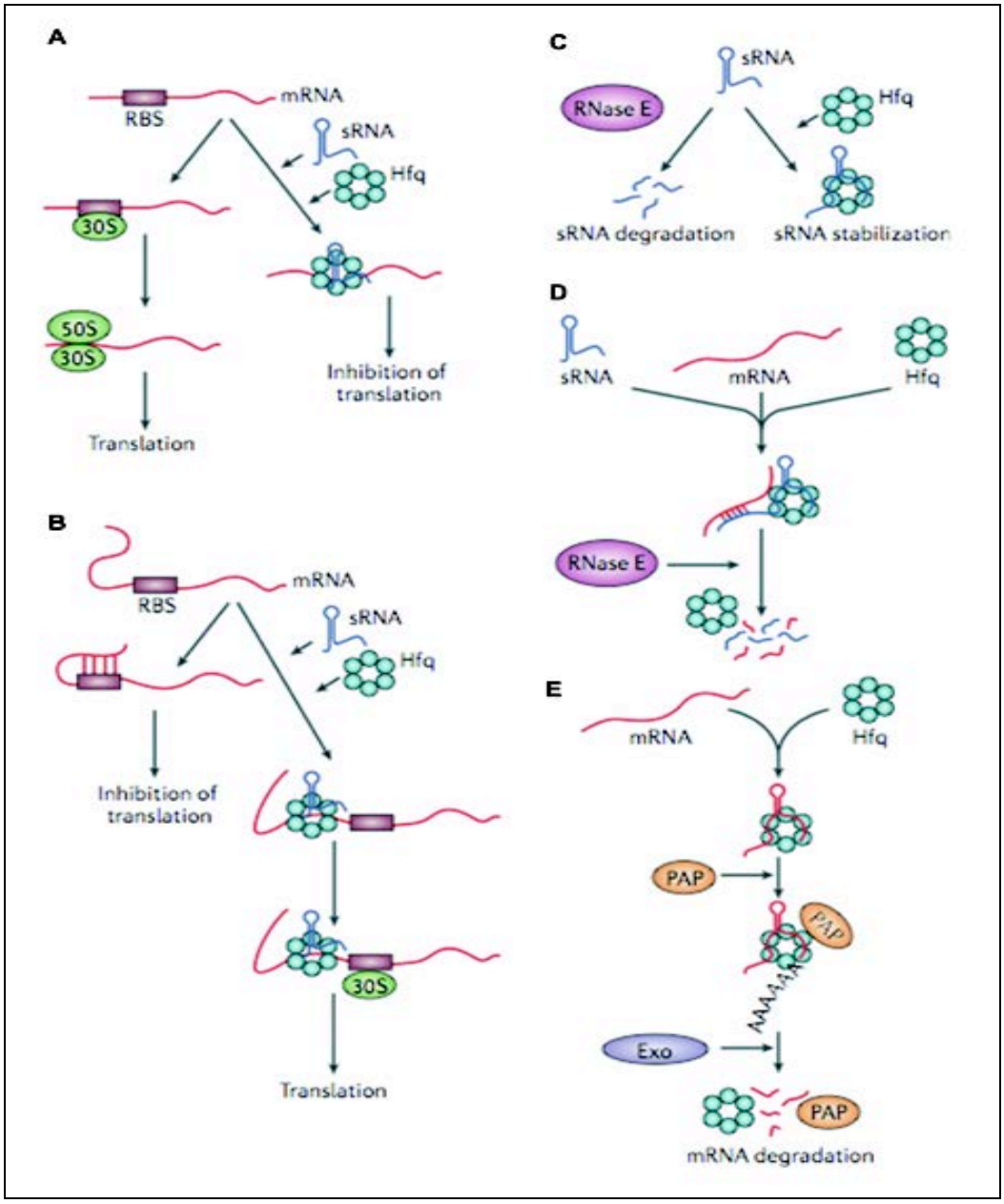

Figure 1.6: Different modes of Hfq activity. A) Hfq associated with npcRNA (sRNA) may occlude the ribosome binding site (RBS) of a target mRNA from binding of 30S and 50S ribosomal subunit, thus repressing translation. **B)** Binding of Hfq to npcRNA (sRNA) unmask the occlusion of RBS by secondary structure formed at 5'-UTR, thus activating translation. **C)** Hfq may protect npcRNA (sRNA) from ribonuclease cleavage. **D)** In some cases, Hfq may induce ribonuclease cleavage of npcRNA (sRNA). **E)** Hfq may stimulate the polyadenylation of an mRNA by poly(A) polymerase (PAP), which in turn triggers the 3'-to-5' degradation by an exoribonuclease (Exo), such as polynucleotide phosphorylase (PNPase), RNase R, or RNase II (adapted from Vogel & Luisi, 2011).

energy of the npcRNA/mRNA pairing interaction, genome size, as well as the structural variations of Hfq protein (Jousselin et al., 2009). For example, Hfq is needed for interaction of RyhB/*sodB*, GcvB/*dppA*, OxyS/*fhlA*, and GcvB/*oppA* (due to their low predictive ΔG value); but is dispensable for interaction of RNAIII/*sa1000* and IstR-1/*tisAB* (due to their high predictive ΔG value) (Jousselin et al., 2009). Apart from that, possession of the *hfq* gene in low-GC Gram-positive *S. aureus* does not endow any significant cellular role (Bohn et al., 2007). Appealingly, role of Hfq in antisense riboregulation of Gram-positive bacteria has now been demonstrated in *Listeria monocytogenes* (Nielsen et al., 2010). Obviously, versatility of this protein will need further studies to shed light on its role in riboregulation.

In most cases, there seems to be a correlation between RNase E and Hfq in RNA duplex formation. RNase E is a major ribonuclease in degradosome (consists of PNPase, enolase, and an RNA helicase, RhlB) as well as a central endonuclease that often recognizes AU-rich regions as cleavage sites (Carpousis, 2007). Degradation of the mRNA seems to be the fate when RNase E and Hfq interact with npcRNA/mRNA complex (Figure 1.6C; Figure 1.6D) (Morita et al., 2005). Nonetheless, Hfq can sometimes confer protection to target mRNA and *trans*-encoded npcRNA due to the competition between RNase E and Hfq towards AU-rich region (Moll et al., 2003). Recently, RyhB was proposed to induce mRNA degradation by promoting a distal downstream RNase E-dependent cleavage site within the target mRNA, *sodB* (Prévost et al., 2011). This remarkable finding provides yet another insight on how RNase E and Hfq could determine the fate of target mRNA and npcRNA.

Undeniably, there are other bacterial proteins that could serve as RNA chaperones in species with no apparent Hfq homolog or in addition to Hfq. The *E. coli* ProQ protein which possesses a CTR, akin to Hfq, has been recently proposed a chaperone role (Chaulk et al., 2011). In *B. subtilis*, an operon expressing three small basic proteins (FbpABC) has also been postulated to function as RNA chaperones, aiding FsrA in regulating *sdhC* expression (Gaballa et al., 2008). Recent work by Rieder et al. (2012) in identifying RNA chaperones in *H. pylori* with no Hfq invigorate the advancement of the characterization of those proteins in facilitating npcRNA base-pairing (Rieder et al., 2012). All in all, there is still room for the yet to be determined mechanism of Hfq or Hfq homologues as binding motifs for these proteins should not be constraints in the wake of versatility of npcRNA or target mRNA.

1.5 Involvement of Bacterial npcRNAs in Regulation of Stress

Bacteria encounter a wide range of stress in their constantly changing environments. Variations in temperature, pH, solute concentrations, nutrients, and oxygen level can exert environmental stress on their growth. In order to adapt and survive in an often hostile atmosphere, bacteria have developed ways to sense changes and orchestrate a cascade of alterations in gene expression and protein activity. Together with two-component signal transduction and regulatory proteins, npcRNAs have been implicated in integrating environmental stress signals and regulating a plethora of stress responses. To date, numerous npcRNAs have been identified and predicted in a wide range of bacteria (Abu-Qatouseh et al., 2010; Altuvia, 2007; Pichon & Felden, 2008; Raabe et al., 2011), including those associated with bacterial ribosomal

protein operons (Khayrullina et al., 2012). Here, the involvement of bacterial npcRNAs in orchaestating selected stress conditions will be discussed.

1.5.1 npcRNAs in Temperature Stress

Temperature is one of the most important parameters that bacterial cells need to closely monitor. Often, signal transduction systems featuring complex feedback loops are responsible for reactions to temperature fluctuations. In contrast to signal transduction system responses that usually have a lag period, RNA-based feedback to temperature fluctuations is more rapid and usually is exerted via changes in structural conformation of regulatory RNA regions termed "RNA thermometers." Frequently, temperature fluctuations affect expression of heat shock, cold shock, and virulence genes. Hence, it is not surprising that these classes of genes employ RNA thermometers to control translation initiation in response to the surrounding temperature.

Typically, RNA thermometers are located within 5'-UTRs of an mRNA and form a secondary structure that occludes the Shine-Dalgarno (SD) sequence from binding to the 30S ribosome subunit (i.e., the switch off state) at low temperature. As the temperature gradually rises, the secondary structure is destabilized via a zipper-like mechanism and formation of translational initiation complexes is permitted (i.e., the switch on state) (Figure 1.7A) (Kortmann & Narberhaus, 2012). The most common RNA thermometer is the Repression Of the heat Shock gene Expression (ROSE) element, which is always associated with genes encoding small heat shock proteins. ROSE elements have been found in rhizobia (Nocker et al., 2001a; Nocker et al., 2001b) and α- and γ-proteobacteria (Waldminghaus et al., 2005). A conserved