View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repository@USM

EXPERIMENTAL RNOMICS: TOWARDS THE IDENTIFICATION AND CHARACTERIZATION OF NON-PROTEIN-CODING RIBONUCLEIC ACIDS IN PATHOGENIC AGENT, *SALMONELLA* TYPHI

ROBAIZA BIN ZAKARIA

UNIVERSITI SAINS MALAYSIA

2009

EXPERIMENTAL RNOMICS: TOWARDS THE IDENTIFICATION AND CHARACTERIZATION OF NON-PROTEIN-CODING RIBONUCLEIC ACIDS IN PATHOGENIC AGENT, *SALMONELLA* TYPHI

by

ROBAIZA BIN ZAKARIA

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

June, 2008

DEDICATION

To,

My late father and mother.
My family.
My friends.

ACKNOWLEDGEMENTS

In the name of Allah the most merciful and the most compasionate. All praise and gratitude is due to Allah and may peace and blessing be upon the prophet Muhammad (saw) and his family and his companion.

First of all I would like to thank the Almighty God. I truly believed that without Him paving the road, I won't be able to meet all the "great hearts" and "helping hands" when ever I am in difficulties.

I am most grateful to my supervisor Prof. M. Ravichandran and Associate Prof. Tang Thean Hock, who have supervised the whole of this thesis work, for their guidance, discussions and most of all for their support, understanding and patience. I will never forget how great their hearts are in helping me and giving me hope throughout the challenging period of hard time during this work.

I would like to thank Prof. Asma Ismail as the Director of the Institute for Research in Molecular Medicine (INFORMM), who has gave the opportunity for me to pursue my PhD and to carry-out this study in INFORMM and for accepting me as the family member of INFORMM.

I would also like to express my gratitude to Associate Prof. Mohd Radzi (Head Department of Microbiology and Parasitology) and Prof. Abdul Aziz Baba (Dean of PPSP) for their great support. My gratitude also goes to the IPS and Human Resource-RLKA of USM for giving me the oppurtunity to pursue my PhD.

I would like to express my deepest gratitude to Dr Timofey S. Rozhdestvensky and Venkata Suresh Babu Chinni at the University of Münster, Germany who have encouraged, supported, argued, criticized and worked shoulder to shoulder as a team.

I would like to thank all my friends in INFORMM and Department of Microbiology and Parasitology, and colleagues. Their presence had given me invaluable experiences and made life interesting inside and outside the lab.

Last but not least, my deepest gratitude to my wife, Zanaidah and my childrens - Abang Alif, Abang Afiq and Adik Akif for their unremitting support and patience.

TABLE OF CONTENTS

Page

DEDICATION....................ii

ACKNOWLEDGEMENTS....................iii

TABLE OF CONTENTS....................iv

LIST OF TABLES....................ix

LIST OF FIGURES....................x

LIST OF ABBREVIATIONS....................xii

ABSTRAK TESIS....................xiv

THESIS ABSTRACT....................xvi

CHAPTER 1 - INTRODUCTION

(1.1) General Introduction....................1

(1.2) Non-Protein-Coding RNAs (npcRNAs)....................5

(1.3) Characteristic and Biogenesis of npcRNAs....................7

(1.4) Known Function of npcRNAs....................11

(1.4.1) npcRNA Function in Post-transcriptional Gene Silencing....................13

(1.4.2) npcRNA Functions in RNA Editing....................16

(1.4.3) npcRNA Functions in Pre-mRNA Splicing....................16

(1.4.4) npcRNA Functions in Protein Stability....................19

(1.4.5) npcRNA Functions in Nucleotide Modification of RNA....................19

(1.4.6) npcRNA Functions in mRNA Stability and Translation....................21

(1.4.7) npcRNA Functions in Growth Phases or Developmental Regulation........... 26

(1.4.8) npcRNA Functions in DNA Transcriptions 26

(1.4.9) npcRNA Functions in Chromosome Replication 27

(1.4.10) npcRNA Functions in RNA Processing 27

(1.4.11) npcRNA Functions in Protein Synthesis 28

(1.4.12) npcRNA Functions in Protein Translocation 28

(1.5) RNomics: Identification and Characterization of npcRNAs 28

(1.5.1) Computational RNomics 30

(1.5.2) Experimental RNomics.. 31

(1.6) The Pathogenic Agent: *Salmonella enterica* serovar Typhi 33

(1.6.1) *Salmonella* Typhi Genome 34

(1.7) Research Objectives 36

CHAPTER 2 - MATERIALS AND METHODS

(2.1) Materials 38

(2.1.1) Chemicals and Enzymes.. 38

(2.1.2) Membranes 38

(2.1.3) Kits and Other Materials.. 38

(2.1.4) Electrophoresis 39

(2.1.5) Radio Isotopes 40

(2.1.6) Bacteria Strains... 40

(2.1.6.1) *Salmonella* Typhi USM05 Isolate 40

(2.1.6.2) *Escherichia coli* K12 Strain.. 40

(2.1.6.3) *Escherichia coli* TOP10 Competence Cells 40

(2.1.7) Oligonucleotides 41

(2.1.7.1) Oligonucleotides Used for cDNA Synthesis 41

(2.1.7.2) Oligonucleotides Used for Amplification of the cDNA-Plasmid-DNA. 41

(2.1.7.3) Oligonucleotides Used for Northern Blot Hybridization 41

(2.1.7.3.1) Oligonucleotides for *S.* Typhi 41

(2.1.7.3.1.1) Oligonucleotides for npcRNAs in the Intergenic Regions 41

(2.1.7.3.1.2) Oligonucleotides for npcRNAs Antisense to ORFs 42

(2.1.7.3.1.3) Oligonucleotides for npcRNAs Overlapping the ORFs 44

(2.1.7.3.1.4) Oligonucleotides for npcRNAs Present at Multiple Locations 44

(2.1.7.3.2) Oligonucleotides for *E. coli* K12 45

(2.1.7.3.2.1) Oligonucleotides for npcRNAs in the Intergenic Regions 45

(2.1.7.3.2.2) Oligonucleotides for npcRNAs Antisense to ORFs 45

(2.1.7.3.2.3) Oligonucleotides for npcRNAs Overlapping the ORFs 46

(2.1.7.3.2.4) Oligonucleotides for npcRNAs Present at Multiple Locations 46

(2.2) Methods 47

(2.2.1) Preparation of Solutions and Chemicals 47

(2.2.1.1) Preparation of DEPC-Treated Deionized Distilled Water 47

(2.2.1.2) Preparation of Glassware and Plastic Container 48

(2.2.2) Preparation of *S.* Typhi USM05 Cell 48

(2.2.2.1) Determination of *S.* Typhi USM05 Standard Growth Curve 48

(2.2.2.2) Culturing and Harvesting of *S.* Typhi Cell 49

(2.2.3) Nucleic Acids Extraction 49

(2.2.3.1) Extraction of Total Genomic RNA 49
(2.2.3.2) Extraction of Genomic DNA 50
(2.2.3.3) Phenol/Chloroform Extraction 51
(2.2.3.4) Elution of Nucleic Acids from Gel Slices 51
(2.2.4) Construction of the cDNALibrary for npcRNAs 52
(2.2.4.1) Size Selection of Total RNA 52
(2.2.4.2) Treatment of Size-selected RNA with Tobacco Acid Pyrophosphatase (TAP) 52
(2.2.4.3) C-tailing of Size-selected RNAs 52
(2.2.4.4) Ligation of *Sal* I Adapter to the 5'-end of C-tailed RNA 53
(2.2.4.5) cDNA Synthesis 53
(2.2.5) Sequence Analysis and Characterization of cDNA Clones 55
(2.2.6) Polymerase Chain Reaction (PCR) 56
(2.2.7) Purification of Plasmid DNA/PCR Products 56
(2.2.8) Northern Blot Analysis 56
(2.2.9) End-labeling of Oligonucleotides 57
(2.2.10) DNA Sequencing 57
(2.2.11) Sequence Analysis 58
(2.2.12) Databases 58

CHAPTER 3 - RESULTS AND DISCUSSION

(3.1) Determination of Growth Phases for *S.* Typhi USM05 59
(3.2) Research Strategy 60

(3.3) Analysis of the cDNA Library Representing Novel npcRNA Candidates from *S.* Typhi USM05 71

(3.4) Sequences Homology of *S.* Typhi USM05 Novel npcRNA Candidates in *S.* Typhi Ty2 and *E. coli* K12 767

(3.5) Expression Analysis of Potential Novel npcRNA Candidates of *S.* Typhi USM05 and *E. coli K*12 by Northern Blot Hybridization 80

(3.5.1) Expression of Novel npcRNA Candidates 82

(3.5.2) Expression Features of Novel npcRNA Candidates 83

(3.6) Classification of potential novel npcRNAs candidates in *S.* Typhi USM05 93

(3.6.1) npcRNAs Transcribed from the Intergenic Region 93

(3.6.2) npcRNAs Transcribed in the Antisense Orientation of an ORF 104

(3.6.3) npcRNAs Overlapping ORF 113

(3.6.4) npcRNAs Located at Several Transcription Locations 123

CHAPTER 4 - CONCLUSION AND FUTURE STUDY

Conclusion and Future Study 131

REFERENCES 134

APPENDICES 148

PUBLICATIONS 164

LIST OF TABLES

Page

Table 3.1 OD_{600} reading for *S.* Typhi USM05 culture..**61**

Table 3.2 Analysis of potential novel npcRNA candidates in *S*. Typhi USM05 and *E. coli* K12...**84**

Table 3.3 Compilation of 28 novel npcRNA candidates transcribed from intergenic region...**95**

Table 3.4 Compilation of 29 novel npcRNA candidates transcribed in the antisense orientation of an ORF...**105**

Table 3.5 Compilation of 18 novel npcRNAs candidates overlapping ORF............**114**

Table 3.6 Compilation of seven novel npcRNAs candidates with sequence homology located at multiple locations...**124**

LIST OF FIGURES

Page

Figure 1.1 The proportion of non-protein-coding and protein-coding sequences in several model organisms..**3**

Figure 1.2 The organization of npcRNA genes...**10**

Figure 1.3 Overview and model of DsrA/Hfq/*rpoS*-mediated regulation....................**12**

Figure 1.4 Mechanisms of action for siRNA and miRNA..**15**

Figure 1.5 Mechanism of U insertion/deletion RNA editing..**17**

Figure 1.6 Spliceosome assembly...**18**

Figure 1.7 Mechanism and function of tmRNA..**20**

Figure 1.8 Structural features of C/D box snoRNAs...**22**

Figure 1.9 Structural features of H/ACA box snoRNAs..**23**

Figure 1.10 Mechanism of translational inhibition/activation in *E. coli* directed by npcRNAs...**24**

Figure 1.11 Schematic presentation of the systemic identification of novel npcRNAs candidates...**32**

Figure 3.1 Growth curve of *S.* Typhi USM05..**62**

Figure 3.2 Total RNA extracted from the lag, exponential and stationary phase of *S.* Typhi USM05...**63**

Figure 3.3 The treatment of size-selected total RNA with TAP.......................... ....**64**

Figure 3.4 Schematic presentation of a modified systemic identification of novel npcRNAs candidates.. ......**66**

Figure 3.5 The *Not*I/*Sal*I digested amplicons separated on a 2% agarose gel.......... .....**68**

Figure 3.6 An initial screening of 228 *S.* Typhi USM05 cDNA clones.........................**70**

Figure 3.7 Sequences analysis of *S.* Typhi cDNA clones...**72**

Figure 3.8 Final grouping and sequence homology of the npcRNA candidates...........**75**

Figure 3.9 Polyacrylamide gel electrophoresis of *S.* Typhi USM05 and *E. coli* K12 total RNA from three different growth phases...**81**

Figure 3.10 The expression patterns of npcRNA candidates..**85**

Figure 3.11 Schematic overview of the expression features observed in 38 novel npcRNA candidates with NB result..**87**

Figure 3.12 Genomic context and NB result for StyR-99, StyR-111 and StyR-280......**97**

Figure 3.13 Genomic context and NB result for StyR-56, StyR-59, StyR-161, StyR-381, StyR-3, StyR-47, StyR-362 and StyR-333..............................**100**

Figure 3.14 Genomic context and NB result for StyR-122, StyR-137, StyR-196, StyR-252 and StyR-369...**107**

Figure 3.15 Genomic context and NB result for StyR-254, StyR-264 and StyR-263..**109**

Figure 3.16 Genomic context and NB result for StyR-151, StyR-199, StyR-234, StyR-248 and StyR-358...**112**

Figure 3.17 Genomic context and NB result for StyR-296, StyR-51, StyR-74, StyR-341, StyR-219 and StyR-250..**116**

Figure 3.18 Genomic context and NB result for StyR-29, StyR-328, StyR-50 and StyR-281..**120**

Figure 3.19 Genomic context and NB result for StyR-103, StyR-186, StyR-207, StyR-216, StyR-327 and StyR-329..**126**

LIST OF ABBREVIATIONS

ATP - adenosine triphosphate
BSA - bovine serum albumin
cDNA - complementary DNA
CTP - cytosine triphosphate
ddH_2O - de-ionized distilled water
DEPC - diethyl pyrocarbonate
DNA - deoxyribonucleic acid
dNTP - dideoxynucleotide triphosphate
DTT - dithiothreitol
EDTA - ethylene diamine tetra acetic acid
ESTs - expressed sequence tags
fRNAs - functional RNAs
gRNA - guide-RNA
Hfq - host factor 1 protein
IGR - intergenic region
IS - insertion sequence
LB - Lauria Bertani media
M - molar
MDR - multidrug resistance
miRNAs - micro RNAs
$MnCl_2$ - Manganese chloride
mRNA - messenger-RNA
NaCl - Sodium chloride
NaOAc - Sodium acetate
NB - Northern Blot
NCBI - National Center for Biotechnology Information
ncRNAs - non-coding RNAs
nmRNAs - non-messenger RNAs
npcRNAs - non-protein-coding RNAs
OD_{260} - optical density at 260 nm wavelength
OD_{600} - optical density at 600 nm wavelength
$OD_{260/280}$ - the ratio of optical density at 260 and 280 nm wavelength
ORF - open reading frame
PCR - polymerase chain reaction
pre-mRNAs - precursors of messenger RNAs
pre –rRNAs - precursors of ribosomal RNAs
pre-tRNAs - precursors of transfer RNAs
RE - restriction enzyme
RISC - RNA-induced silencing complex
RNA - ribonucleic acid
RNAi - RNA interference
RNP - ribonucleoprotein

rRNA	- ribosomal-RNA
SDS	- sodium duodecyl sulfate
SELEX	- Systematic Evolution of Ligands by Exponential Enrichment
siRNAs	- small interfering RNAs
snRNAs	- small nuclear RNAs
snRNP	- small nuclear RNA-ribonucleoprotein complex
snmRNAs	- small non-messenger RNAs
sncRNAs	- small non-coding RNAs
snoRNA	- small nucleolar RNA
snoRNPs	- small nucleolar RNA-ribonucleoprotein particles
SR	- signal recognition particle - membrane-associated receptor complex
sRNAs	- small RNAs
SRP	- signal recognition particle
stRNAs	- small temporal RNAs
TAP	- tobacco acid pyrophosphatase
TAE	- tris-acetate-EDTA
TBE	- tris borate EDTA
TEFB	- transcription elongation factor-B
tRNA	- transfer-RNA
tmRNAs	- transfer-messenger RNAs
UTR	- un-translated region
utRNAs	- un-translated RNAs
v/v	- volume per volume
w/v	- weight per volume

RNOMIKS EKSPERIMENTAL: KEARAH PENGENALPASTIAN DAN PENCIRIAN ASID RIBONUKLEIK BUKAN-PENGKOD-PROTEIN DI DALAM AGEN PATOGENIK, *SALMONELLA* TYPHI

ABSTRAK TESIS

RNA bukan-pengkod-protein (npcRNA) merupakan satu kelas pengawalatur-ribo yang bertindak di dalam bentuk kompleks RNA-protein (sebagai RNPs) di dalam pelbagai laluan pengawalaturan. Tesis ini memberikan tumpuan ke atas pengenalpastian npcRNA secara eksperimental daripada bakteria patogenik *Salmonella enterica* serovar Typhi (*S.* Typhi), penyebab penyakit demam kepialu. Melalui pendekatan RNomics Eksperimental, 82 calon novel npcRNA daripada perpustakaan cDNA *S. typhi* telah dikenalpasti dan dicirikan. Daripada jumlah ini, 28 telah ditranskrip daripada IGR, 29 ditranskrip di dalam arah antisense kepada ORF dan 18 dikenalpasti bertindihan dengan ORF. Sementara 7 calon yang lain telah ditranskrip daripada kawasan repititif dan beberapa kedudukan bukan repitatif yang lain. Sebelas npcRNA merupakan npcRNAs yang telahpun dilaporkan. Menariknya, 55 daripada 82 calon yang dikenalpasti mempamerkan homologi kepada *Escherichia coli* serta belum lagi dilaporkan dari kedua-dua organisma. Melalui analisis Pemblotan Northern, pengungkapan 28 novel npcRNA telah berjaya disahkan. Sejumlah 38 calon novel npcRNA adalah dikawalatur semasa fasa pertumbuhan, di mana 14 calon adalah dikawalatur semasa fasa pertumbuhan di dalam kedua-dua *S.* Typhi USM05 dan *E. coli* K12, 11 calon adalah dikawalatur semasa fasa pertumbuhan secara spesies-spesifik hanya di dalam *S.* Typhi USM05 dan 3 calon adalah

dikawalatur semasa fasa pertumbuhan secara spesies-spesifik hanya di dalam *E. coli* K12. Sejumlah 8 calon npcRNA adalah diungkapkan secara ubikuitus, dimana 6 calon diungkapkan secara ubikuitus di dalam kedua-dua *S.* Typhi USM05 dan *E. coli* K12. Dua calon adalah diungkapkan secara ubikuitus-spesies-spesifik hanya di dalam *S.* Typhi USM05. Tidak ada calon yang diungkapkan secara ubikuitus-spesies-spesifik di dalam *E. coli* K12. Dua daripada 38 telah ditunjukkan dikawalatur semasa fasa pertumbuhan dan turut diungkapkan secara ubikuitus. Ciri bagi corak pengungkapan berbeza yang telah diperhatikan boleh dihubungkaitkan dengan kelas npcRNA berbeza yang mana mereka dikumpulkan iaitu npcRNA untuk "housekeeping" dan untuk pengawalaturan. Beberapa calon novel npcRNA telah ditunjukkan terletak berdekatan dengan gen-gen penting yang kekal di dalam *E. coli* K12 dan di antara *Salmonella.* Justeru, adalah dicadangkan bahawa mereka mungkin adalah penting daripada segi fungsi. Secara signifikan, kajian ini telah menetapkan suatu pemahaman ke atas kepentingan keadaan pertumbuhan yang spesifik terhadap pengungkapan npcRNA. Dengan mempertimbangkan bahawa pengawalatur-pengawalatur berkenaan mungkin diungkapkan apabila diaruhkan, molekul-molekul ini mungkin dapat diperkayakan dengan begitu banyak di dalam perpustakaan cDNA yang disediakan daripada keadaan pertumbuhan atau keadaan tekanan yang berkaitan.

EXPERIMENTAL RNOMICS: TOWARDS THE IDENTIFICATION AND CHARACTERIZATION OF NON-PROTEIN-CODING RIBONUCLEIC ACIDS IN PATHOGENIC AGENT, *SALMONELLA* TYPHI

THESIS ABSTRACT

Non-protein-coding RNA (npcRNA) is a large class of riboregulators that act in complex with proteins (as RNPs) in diverse regulatory pathways. This thesis focused on the experimental identification of small npcRNAs from *Salmonella enterica* serovar Typhi (*S.* Typhi), the aetiological agent of typhoid fever. By an Experimental RNomics approach, 82 species of uncharacterized novel npcRNA candidates were identified from library generated from different growth phases of a clinically isolated *S.* Typhi. From this, 28 were transcribed from the IGRs, 29 were transcribed in the antisense orientation of the ORFs and 18 were identified to overlap the ORFs. Another 7 candidates were transcribed from repetitive regions and several non-repetitive locations. Eleven known npcRNAs were also detected. Interestingly, 55 candidates exhibit homology to *Escherichia coli* and were not previously annotated for both organisms. By Northern Blot analysis, the expression of 38 novel npcRNA candidates was confirmed. A total of 28 novel npcRNA candidates were growth phase regulated where 14 candidates were growth phase regulated in both *S.* Typhi USM05 and *E. coli* K12, 11 candidates were species-specific growth phase regulated only in *S.* Typhi USM05 and three candidates were species-specific growth phase regulated only in *E. coli* K12. A total of 8 npcRNA candidates were ubiquitously expressed with 6 candidates were ubiquitously expressed in

both *S.* Typhi USM05 and *E. coli* K12 and 2 candidates were species-specific ubiquitously expressed only in *S.* Typhi USM05. There was no species-specific ubiquitously expressed candidate observed in *E. coli* K12. Two out of 38 candidates were shown to be growth phase regulated and ubiquitously expressed. The different feature of expression patterns observed in this study can be associated with the different classes of npcRNA that they might be grouped-in, namely the housekeeping and the regulatory npcRNAs. A number of novel npcRNA candidates were shown to be located close to important genes that were conserved in *E. coli* K12 and among *Salmonella*. Thus, suggesting that they might be functionally important. Significantly, this study has set the understanding on the importance of the specific growth conditions to the expression of npcRNAs. The next phases of challenge are to further characterize and to elucidate the functions of these new classes of small RNA molecules. Considering that such regulators may be highly expressed upon induction, these molecules may be greatly enriched in cDNA libraries prepared from the relevant growth or stress condition.

CHAPTER 1

INTRODUCTION

(1.1) General Introduction

It is believed that modern biochemistry arose with the first appearance of deoxyribonucleic acid (DNA), leading to the birth of "Central Dogma of Molecular Biology". The dogma defined a general pathway for the expression of genetic information stored in the DNA. The genetic information is transferred into messenger-RNA (mRNA) through transcription by RNA polymerase and is then translated on ribosomes, a ribonucleoprotein complex of ribosomal-RNA (rRNA), with the help of specialized adapter, the transfer-RNA (tRNA), that carry specific amino acid, to produce functional proteins. These proteins, then in turn perform all the enzymatic and structural functions in the cell.

Consequently for decades, the belief that the critical functions of the cell and the complexity of an organism depend exclusively on the protein-coding genes according to "one gene, one protein" hypothesis (Morey and Avner, 2004) has progressively gained its ground and eventually created a great impact on genomic studies (*e.g,* genome sequencing project; Venter *et al.*, 2001) in human and several selected model organisms, where the main interest have been put into discovering protein-coding genes (Eddy, 2001; Morey and Avner, 2004).

The over-simplistic view may probably correct for the prokaryotic organisms where most of the earlier genetic studies have been done based on the “simple” and compact genome of the prokaryotes (Pang *et al.*, 2006). However, with the discovery of the non-coding sequences or introns (Hastings and Krainer, 2001) that lies between protein-coding genes or exons, and intronic splicing activities of primary RNAs transcribed from eukaryotic genes, the dogma have been challenged.

Furthermore, recent analysis of genomes strongly suggested that protein-coding genes alone are not enough to account for the complexity of higher organism (Mattick, 2001). This can be seen from the evidences that with the increase of an organism complexity, the protein-coding genes contribution to its genome is tremendously decreased (Eddy, 2001; Szymanski and Barciszewski, 2002; Shabalina and Spiridonov, 2004). The genomes of *Caenorhabditis elegans* and *Drosophila melanogaster* for example, were found to contain only twice as many protein-coding genes as in yeast and bacteria, and in human genome, the numbers is about twice that of the invertebrates (Mattick, 2001; Standish, 2002; Figure 1.1).

The ribonucleic acid (RNA) molecule has been shown to be a versatile bio-organic molecule with a variety of roles and functions in structural, genetic storage, regulatory and catalytic processes in living cell (Caprara and Nilsen, 2000). The identification of catalytic properties of the RNA subunit of ribonuclease P, aptamers, ribozymes, intronic self-splicing and gene silencing activities have further underlined the notion that the function of RNA molecules extended well beyond a transient role in

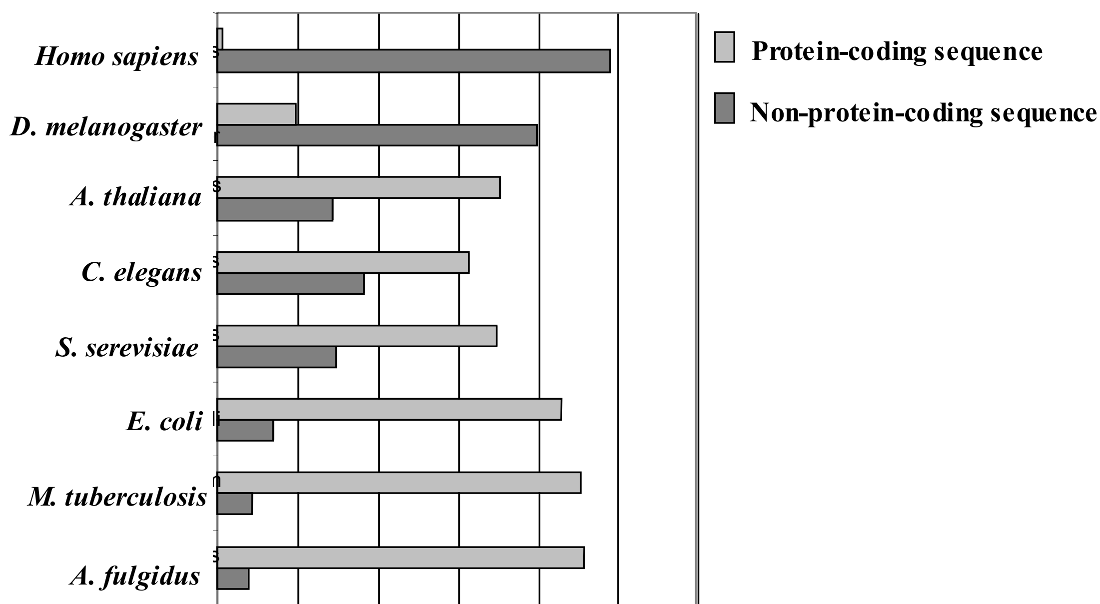

Figure 1.1: The proportion of non-protein-coding and protein-coding sequences in several model organisms. The proportion of protein-coding sequences (light shaded bar) compared to protein-coding sequences (dark shaded bar) is increased in the genome of less complex organisms e.g., *A. fulgidus*, *M. tuberculosis* and *E. coli*. In contrast, the contribution of protein-coding sequences compared to the non-protein-coding sequences is decreased with the increase of the organism complexity as shown by fruit fly (*D. melanogaster*) and human (*Homo sapien*) where non-protein-coding sequences in human contributed approximately 98.5% to its genome (adapted from Szymanski and Barciszewski, 2002).

ensuring the expression of protein-coding genes (Levy and Ellington, 2001). It is estimated that 95% to 98.5% of the transcriptional output of eukaryotic genomes are un-translated RNAs without protein-coding properties that are transcribed from non-coding sequences (Eddy, 2001; Mattick, 2001; Figure 1.1).

In the literatures, the un-translated RNAs are generally referred as small RNAs (sRNAs) or non-coding RNAs (ncRNAs). However, the term sRNAs is predominantly used for the un-translated RNAs in prokaryotes (Wassarman *et al.*, 1999; Suzuma *et al.*, 2002; Pichon and Felden, 2005). While the later, ncRNA is used for the un-translated RNAs found in eukaryotes and is originally used exclusively for polyadenylated eukaryotic RNAs transcribed by RNA polymerase II, carrying a 7-methylguanosine cap structure but lacking open reading frame (ORF). Nowadays, the term ncRNA has been extended to designate any piece of RNA transcripts without protein-coding capacity in both prokaryotic and eukaryotic cells (Eddy, 2001; Mattick and Gagen, 2001; Morey and Avner, 2004; Liu *et al.*, 2006; Pang *et al.*, 2006).

Less frequently used terms are un-translated RNAs (utRNAs), functional RNAs (fRNAs) and small non-coding RNAs (sncRNAs) (Ambros, 2001; Argaman *et al.*, 2001; Erdmann *et al.*, 2001b; Wagner and Flardh, 2002; Brosius, 2005; Vogel and Sharma, 2005). Some authors have also referred them as non-messenger RNAs (nmRNAs) to discriminate them from mRNAs (Hüttenhofer *et al.*, 2001). Since the majority of the known ncRNAs are small in size, that ranged between 20-500 nucleotide (nt), Hüttenhofer *et al.* (2001) coined and popularized the term small non-messenger RNAs

(snmRNAs) to refer to this group of ncRNAs. Recently, non-protein-coding RNAs (npcRNAs) term has also been used (Brosius and Tiedge, 2004) and the term npcRNAs is used in this thesis to refer to all the un-translated RNAs species.

(1.2) Non-Protein-Coding RNA (npcRNA)

The npcRNA molecules have been shown to involve in variety of functions in all living cells. The npcRNAs in-complexed with proteins to form ribonucleoprotein (RNP) complexes that functions in DNA and RNA stability. They also function in diverse regulatory pathways including chromosome modification, transcriptional and translational control, splicing, development timing control, cell differentiation, proliferation, apoptosis, organ development, human diseases and immune system (Ambros, 2001; Szymanski and Barciszewski, 2002; Cheng *et al.*, 2003; Ranum and Day, 2004;; Mocellin *et al.*, 2006; O'Gorman *et al.*, 2006; Royo *et al.*, 2006; Barbarotto *et al*., 2008; Kawaji and Hayashizaki, 2008). Their versatility has been known to be contributed by the chemical and physical properties of the RNA molecules (Johansson, 2005).

The RNA molecules have been known to play important roles and functions well beyond the role of ensuring the expression of protein-coding genes (Caprara and Nilsen, 2000; Levy and Ellington, 2001; Kawaji *et al*., 2008; Royo and Cavaille, 2008). Single-stranded RNA molecules, for example, have been shown to posses the ability to form

local secondary structures through the interactions of complementary segments within the same strand. These secondary structure elements can influence many cellular processes as mentioned earlier (Katz and Burge, 2003).

The RNA molecule acquires complex folded conformations that can be applied in any sophisticated recognition process by providing recognition elements for protein binding, formation of large macromolecule complexes and catalytic reactions. The tertiary structure of the RNA molecules can also form a virtually unlimited number of highly specific ligand-binding sites that permit RNA interaction with chemically and structurally diverse sets of small compounds like metal ions and nucleotide (Erdmann *et al.*, 2001b; Szymanski *et al.*, 2003; Brantl, 2004; Johansson, 2005).

RNA and protein bind together through recognition that occurs by induced fit mechanism rather than by the rigid protein-like “lock and key” docking (Moore and Steitz, 2002). Base pairing allows a small RNA to target other nucleic acids with a great specificity. They are often found to have roles that involve sequence-specific recognition or base complimentarity of another nucleic acid. RNA molecules by its very nature are ideal materials for this role. Base complimentarity allows a very small RNA to be exquisitely sequence specific.

Many functional roles do not require the more sophisticated catalytic prowess of protein and could be carried out by simple RNA molecules. Post-transcriptional regulation, in particular, can be achieved simply by steric occlusion of sites on a target

pre-mRNA or mature RNA (Morey and Avner, 2004). Evolution of a small, specific complementary RNA can be achieved in a single step, just by a partial duplication of a fragment of the target gene into an appropriate context for expression of the new npcRNAs.

Traditionally, many npcRNAs are grouped into specific classes based on their physical properties such as size, structure or sequence motives, protein partner or sub-cellular location (Hüttenhofer *et al*., 2002). They might also be classified according to their functions such as catalyst, guides, catalytic cofactors, antisense RNAs, protein binding-site antagonists/agonist or templates (Hüttenhofer and Schattner, 2006).

(1.3) Characteristic and Biogenesis of npcRNA

Generally, npcRNAs often show only weak primary sequence conservation, lack of open reading frame (ORF) and are not systematically processed (Morey and Avner, 2004). Some npcRNAs such as the 16S and 23S rRNAs in bacteria are relatively large from 2,000 - 4,000 nt in length. However, majority of npcRNAs are usually very small with size between 20 – 500 nt. The npcRNA genes are expressed in only few cells or at specific cell growth stages/phases (Eddy, 2002; Yuan *et al.*, 2003).

Many are expressed at low level, that other may suggest that these RNAs are merely transcriptional noise from illegitimate promoters. Notably, in some large size npcRNAs, evolutionary conservation may not be a reliable signature of functional npcRNAs. Unlike DNA, npcRNAs that lack of conservation does not necessarily

indicates lack of function. A number of well-studied npcRNAs, for example human's *XIST* (17 kb) and mouse's *Air* npcRNAs are poorly conserved (Oudejans *et al.*, 2001; Chureau *et al.*, 2002) but are involved in dosage compensation and X-chromosome inactivation during development.

In eukaryotes, npcRNAs are transcribed by different polymerases. For example, RNA polymerase I transcribed mostly ribosomal RNAs. RNA polymerase III transcribed small structural RNAs like tRNAs and 5S-RNA. Whereas RNA polymerase II transcribed mRNAs and most other npcRNAs. Long regulatory npcRNAs in Eukaryal are believed to be transcribed from within a well-defined gene with polymerase II promoters and sometimes are alternatively spliced and/or polyadenylated (Erdmann *et al.*, 2001a; Morey and Avner, 2004) and therefore might have evolved from genes that have lost their protein-coding capacity. Other long npcRNAs are thought to be of functional importance for cell viability and expressed from both chromosomes (Eddy, 2001).

Some vertebrates' npcRNAs are transcribed as monocistronic units from independent genes (Maxwell and Fournier, 1995). However, many npcRNAs in plant and yeast are processed from polycistronic precursors (Maxwell and Fournier, 1995; Samarsky and Fournier, 1999). There are some of the eukaryotes' npcRNAs that are not independently transcribed at all but they are processed out of the introns of host transcripts (Eddy, 2002) or may be produced from intergenic transcription (Morey and Avner, 2004). Some host genes are non-protein coding and appeared to act exclusively as vehicles for npcRNAs only (Tycowski *et al.*, 1996; Bortolin and Kiss, 1998; Pelczar and

Filipowicz, 1998). There are also npcRNAs that are generated from complex transcription unit of spanning tandem repetitions of the intron and non-protein coding flanking exons (Cavaille *et al.*, 2000). The transcriptions of npcRNAs in eukaryotes are illustrated in Figure 1.2:A.

Bacterial npcRNAs are mostly detected from the short intergenic region of the genomes and probably derived from unidirectional, convergent or divergent orientation of the ORFs (Rogozin *et al.*, 2002). The transcriptions of npcRNAs in prokaryotes are illustrated in Figure 1.2:B. Most antisense npcRNAs appeared to overlap the 3'-ends un-translated region (UTR) of the protein-encoding ORF, whereby the translation termination codon of the ORF falls in the 5'-end or 3'-end half of the npcRNA (Dahary *et al*., 2005). The base-pairing regulatory RNAs or riboregulators (antisense RNAs) usually found in bacteria exert their functions via blocking translation by sequestration of ribosome loading site, promote target RNA decay by creating RNase III substrates, induce premature termination of transcription, interfere with activator pseudoknots or even activate an mRNA by unfolding a ribosome-binding site (Wagner and Flardh, 2002). Majority of antisense npcRNAs in bacteria are trans-encoded in the chromosome.

The antisense RNAs can be separated into *cis*-encoded and *trans*-encoded groups. The *cis*-encoded group consisted of npcRNAs are encoded at the same genetic location but on the opposite strand to the RNA they act upon, thus exhibits a long perfect base-pairing with their potential target. The *trans*-encoded group on the other hand is

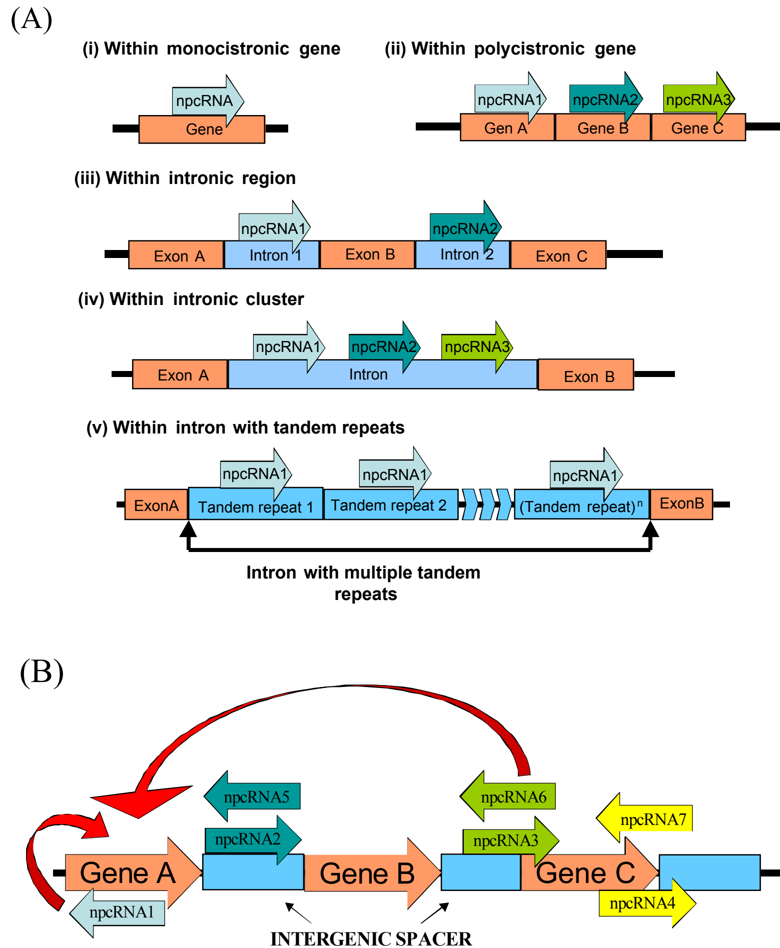

Figure 1.2: The organization of npcRNA genes. (A) npcRNA genes organization in eukaryote. In eukaryotes, npcRNAs can be transcribed either from **(i)** monocistronic units of independent protein-coding genes, **(ii)** polycistronic precursors of protein-coding genes cluster or intergenic region of the introns of host transcripts that are maybe **(iii)** single intronic region or **(iv)** long expanding intronic cluster. There are npcRNAs that are generated from **(v)** complex transcription unit of spanning tandem repetitions of the intron and protein-coding flanking exons. Some host genes are non-protein-coding genes and appeared to act exclusively as vehicles for npcRNAs only; they are similarly organized as in **(i)** and **(ii)**, however the genes are non-protein-coding genes. **(B) npcRNA genes organization in prokaryote.** In prokaryote, the npcRNAs can be derived from unidirectional, convergent or divergent to the ORF's orientation (i.e, npcRNA1, npcRNA2, npcRNA3 and npcRNA4 are unidirectionally transcribed to GeneA, GeneB and GeneC, respectively; npcRNA5 and npcRNA6 are convergently transcribed to GeneA and GeneB, respectively; npcRNA5 and npcRNA6 are divergently transcribed to GeneB and GeneC, respectively). Bacterial npcNAs are mostly detected from the short intergenic spacer region of the genomes i.e, npcRNA2 and npcRNA5. The cis-antisense npcRNAs (i.e, npcRNA1) are transcribed from the opposite strand of the ORFs (i.e, GeneA). The trans-antisense npcRNAs (i.e, npcRNA1) may be transcribed from the same or opposite strand (i.e, npcRNA6) at a distance away from the ORFs (the red arrows indicated the target gene). Some npcRNAs can also be found to overlap the 3' end or 5' end of the ORFs where large sequences of the npcRNAs were in the intergenic region (i.e, npcRNA3, npcRNA4, npcRNA6 and npcRNA7). NpcRNA: non-protein-coding RNA (adapted from Hüttenhofer *et al.*, 2004).

consisted of npcRNAs that are encoded at a chromosomal location distinct from the RNA target they act upon and generally do not exhibit perfect base-pairing with their potential target (Figure 1.2:B) (Storz *et al.*, 2005).

Interestingly, there are significantly large numbers of bacterial *trans*-encoded antisense npcRNAs such as MicC, MicF and DsrA, that show strong requirement for the RNA chaperone protein complex; an Sm and Sm-like homolog regulator in the eukaryotes; known as host factor 1 (Hfq) (Moll *et al.*, 2003; Storz *et al.*, 2005; Vecerek *et al*, 2008). Most of these npcRNAs share a similar structure composed of three stem-loops (Sauter *et al.*, 2003; Zhang *et al.*, 2003). The 11.2 kD Hfq protein forms a homo-hexameric ring-shaped structure that bind to AU-rich sequences in npcRNAs (Wagner and Flardh, 2002; Moll *et al.*, 2003; Storz *et al.*, 2005). Hfq binding can either promotes npcRNA-mRNA base-pairing, npcRNA/mRNA accessibility to RNases or protecting npcRNA/mRNA against Rnase-E via structural change in the npcRNA or mRNA (reviewed by Storz *et al.*, 2005). The overview and model for the Hfq-mediated DsrA/RpoS regulation in *E. coli* is illustrated in Figure 1.3.

(1.4) Known Function of npcRNAs

Most npcRNAs are regulatory molecules that seemed to fine-tune cellular responses to environmental changes by integrating environmental signals into global regulation. They are involved in many important cellular processes such as post-transcriptional gene silencing, RNA editing, pre-mRNA splicing, protein stability, nucleotide modification of RNA, mRNA stability and translation and growth phases or

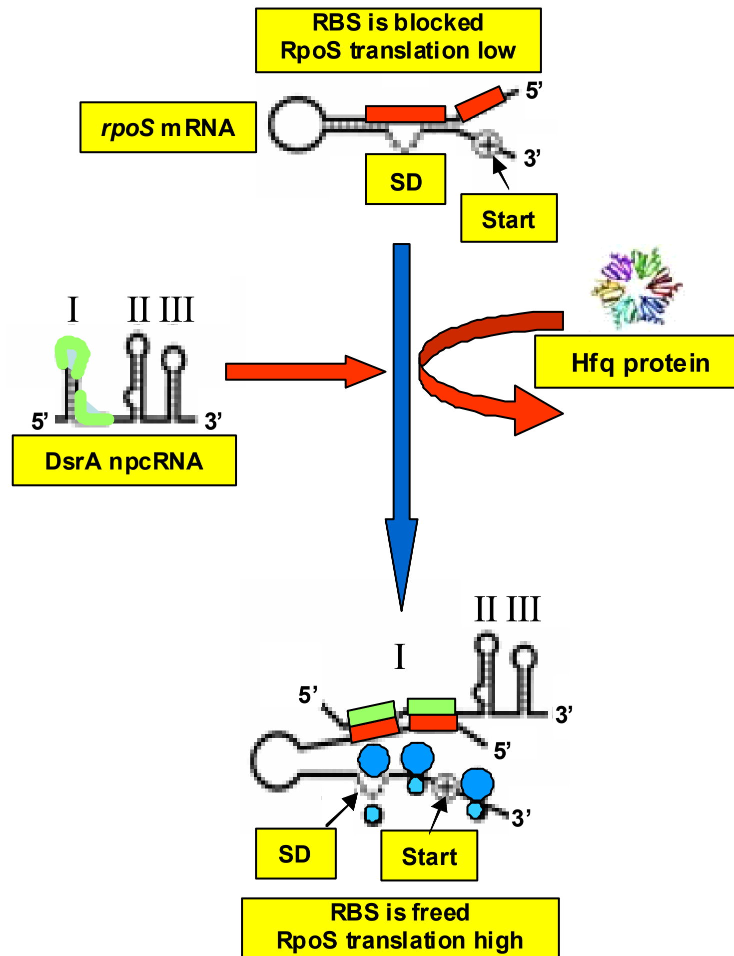

Figure 1.3: Overview and model of DsrA/Hfq/*rpoS*-mediated regulation. In this model, The hexagonal ring, Hfq complex unfolds the *rpoS* 5'-leader at the RBS region, allowing DsrA Domain I to form base-pairing formation with the *rpoS* mRNA. This stabilizes an alternative conformer of r*poS* mRNA that leads to increase translation of RpoS by exposing the RBS. SD: Sine-Delgarno sequence, RBS: ribosomal binding site, Hfq: host factor 1 protein, 5': 5' end, 3': 3' end, Start: Start codon (adapted from Brescia *et al.*, 2003).

developmental regulation (Eddy, 2001; Suzuma *et al.*, 2002; Altuvia, 2004) through several mechanisms such as RNA-RNA or RNA-DNA base-pairing, RNA-protein interactions and intrinsic RNA activity (Altuvia and Wagner, 2000; Wassarman and Storz, 2000). The npcRNAs are also found to involve in other cellular processes such as DNA transcription, chromosome replication, RNA processing, protein synthesis and protein translocation. Some of the known functions of the npcRNAs in prokaryote and eukaryote are briefly described below.

(1.4.1) npcRNA Functions in Post-Transcriptional Gene Silencing

A class of small npcRNAs termed as micro RNAs (miRNAs) and short interfering RNAs (= small interfering RNAs; siRNAs) are found to involve in post-transcriptional gene silencing in eukaryotes. Both miRNAs and siRNAs are generated from long, double-stranded RNAs by a protein complex containing an RNase III-type endonuclease called Dicer (Grishok *et al.*, 2001; Hutvagner *et al.*, 2001; Ketting *et al.*, 2001, Storz *et al.*, 2005; Kawaji and Hayashizaki, 2008).

Dicer cleaves double stranded RNA templates into short siRNAs of 21 to 23 nt that in turn, are able to target certain mRNA regions complementary to the siRNAs resulting in specific cleavage of the mRNA at this site and in loss of function of this mRNA (Bass, 2000; Zamore *et al.*, 2000; Elbashir *et al.*, 2001; Ketting *et al.*, 2001; Nykanen *et al.*, 2001; Zamore, 2001). This mechanism of mRNA degradation has been termed RNA interference (RNAi) (Fire *et al.*, 1998) and is thought to have evolved as a

mechanism for antiviral defense by targeting certain double stranded RNA viruses (Hamilton and Baulcombe, 1999; Tabara *et al.*, 1999; Ketting *et al.*, 1999; Sharp and Zamore, 2000; Wu-Scharf *et al.*, 2000; Elbashir *et al.*, 2001; Waterhouse *et al.*, 2001).

In miRNAs, the RNA precursors are stable stem-loop structures of 60 – 70 nt in length. The 22 nt miRNAs are processed from the stem portion of the precursor (Hutvagner *et al.*, 2001; Grishok *et al.*, 2001) and only one of the strands of the RNA helix is processed and stable. The lin-4 (22 nt) and let-7 (23 nt) are two examples of miRNAs found in *C. elegans*. The common denominator between miRNAs and siRNAs is their size (21 – 23 nt) as well as a similar mechanism of generation *i.e.* from a larger RNA precursor. The generation and mechanism of action of siRNA and miRNA is illustrated in Figure 1.4.

However, it is unclear whether the siRNA-endonuclease complex contains double stranded siRNA. The second difference between miRNAs and siRNAs is their function. While siRNAs are able to degrade specific mRNAs by the RNAi mechanism, miRNA presumably do not degrade their target mRNAs, but apparently exert their function at the level of translation regulation by binding to the 3'-UTRs of mRNAs.

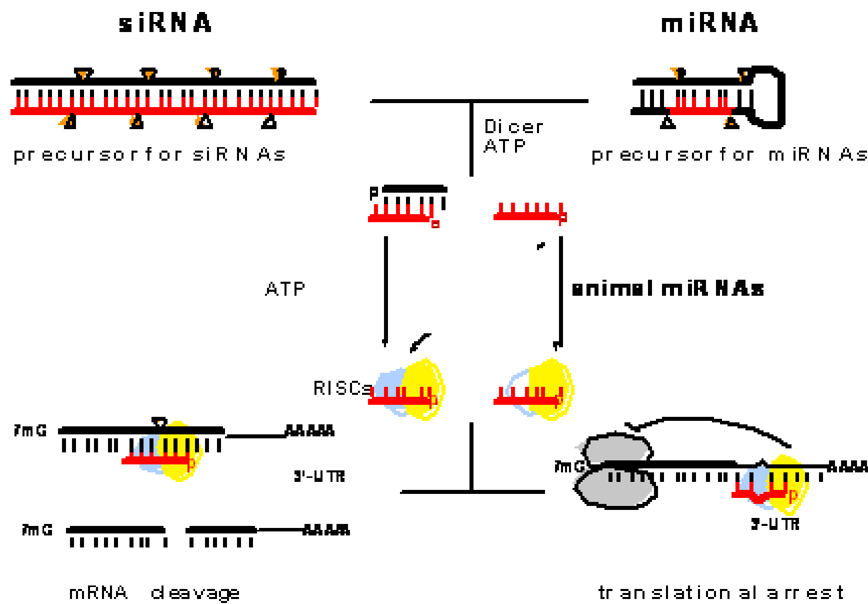

Figure 1.4: Mechanisms of action for siRNA and miRNA. The 21–23 nt long siRNAs or miRNAs are processed from long, bimolecular or intramolecular RNA duplexes by Dicer and ATP; for miRNAs, only one strand appears to be stable while for siRNAs both strands can be detected at this stage; subsequently, miRNAs and siRNAs assemble into RISCs (requiring ATP), that contain a single stranded RNA molecule and target mRNAs either in their ORFs or 3'-UTRs. For miRNAs this leads to translational arrest, while for siRNAs this results in cleavage of the mRNA target at the site of complementarity. As opposed to animal miRNAs, the majority of plant miRNAs have been proposed to function similar to siRNAs, namely degradation of mRNAs. ATP: adenosine triphosphate, miRNA: micro RNA, siRNA: short interference RNA, RISC: RNA-induced silencing complex, ORF: open reading frame, UTR: untranslated region (adapted from Rhoades *et al.*, 2002)

(1.4.2) npcRNA Functions in RNA Editing

RNA editing is a co-transcriptional or post-transcriptional process by which a genome-encoded RNA sequence is altered by nucleotide insertions, deletions or base modifications (Figure 1.5). In the mitochondria of trypanosomatid protozoa, the precursors of messenger RNAs (pre-mRNAs) are edited by site-specific insertions and deletions of uridylate (U) residues. Small *trans*-acting, guide RNAs (gRNAs) supplied the genetic information for this RNA editing process. U-insertion editing occurs through a series of enzymatic steps that begins with gRNA-directed, pre-mRNA cleavage. Inserted U residues are derived from free uridine triphosphate and are added to the 3' terminus of a 5' pre-mRNA cleavage product (Kable *et al.*, 1997; Simpson *et al.*, 2000).

(1.4.3) npcRNA Functions in Pre-mRNA Splicing

In Eukarya, small nuclear RNAs (snRNAs) play a central role in the splicing of pre-mRNAs (Will and Luhrmann, 2001). Eukaryal genes are often interrupted by introns in primary transcripts. The mature RNAs are obtained after the introns are removed. In this process, several small nuclear RNAs (snRNA U1, U2, U4, U5 and U6) are involved. Together with their specific protein components, small nuclear ribonucleoprotein (snRNP), they are required for the assembly of a ribonucleoprotein complex termed the eight spliceosome. The spliceosome catalyzes splicing of pre-mRNAs by removing non-coding introns from eukaryal mRNA precursors (Moore and Sharp, 1993) (Figure 1.6).

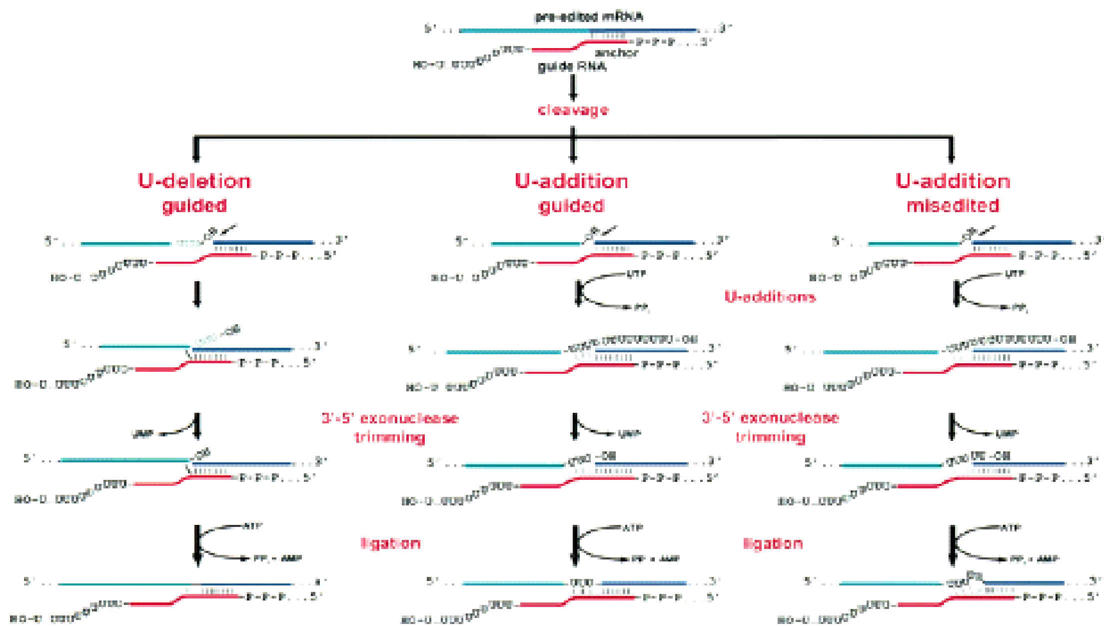

Figure 1.5: Mechanism of U insertion/deletion RNA editing. The vertical lines indicate base pairs. The arrowheads indicate sites of cleavage. Modified enzyme cascade model. The 3´-oligo (U) tail of the gRNA (guide-RNA) is shown as a single-stranded overhang, but it is possible that the tail can interact with the purine-rich pre-edited sequence and the gRNA may have secondary structure. In the U deletion model 3 unpaired Us (in gray) to be deleted are shown as an example. It is possible that the U addition activity adds U's to the 3´-end of the cleavage fragment at the deletion site, which are then trimmed back, but this scenario is not indicated. In the U addition model 13 U's are shown added to the 5´-fragment, but the evidence indicates that the number of added Us is actually heterogeneous. In the 'guided' diagrams the exonuclease nucleotide trimming is complete, yielding the correct –3 or +3 guided products. If trimming is incomplete or excessive prior to ligation, gRNA-dependent misedited products are produced, as shown in the 'misedited' diagram (adapted from Simpson and Emeson, 1996).

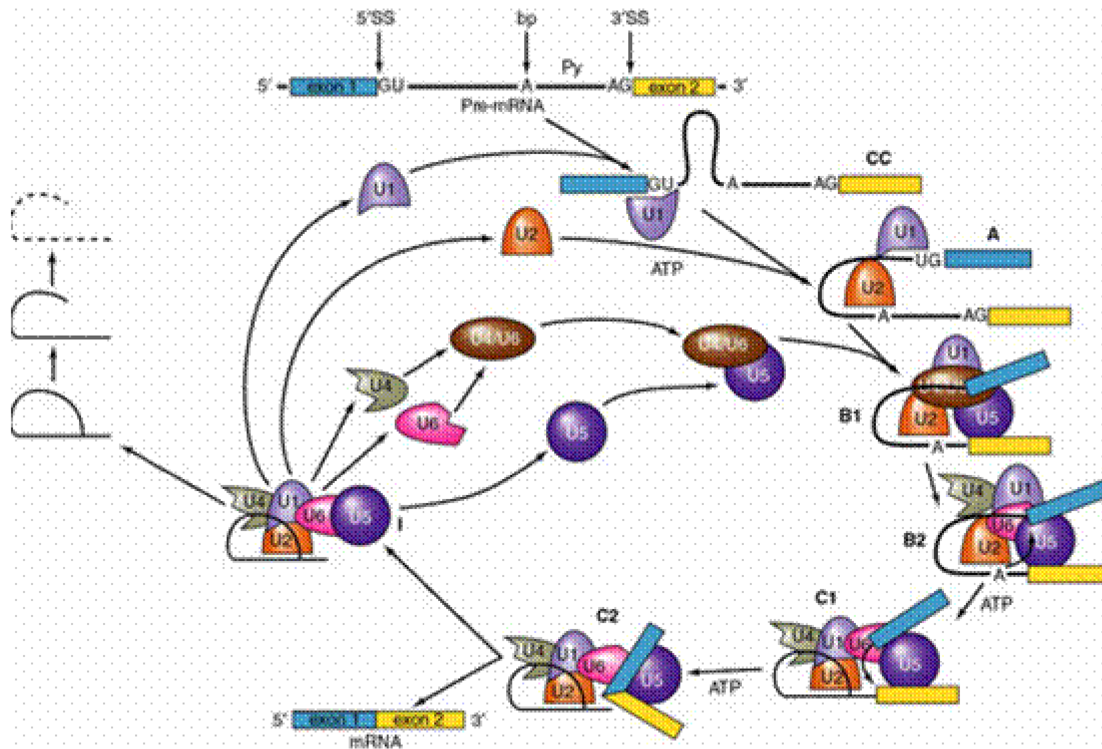

Figure 1.6: Spliceosome assembly. As the newly transcribed mRNA is emerging from the RNA polymerase II, U1 and U2 snRNPs bind through RNA:RNA base pairing to the 5′ splice and branch sites, respectively. U2 snRNA binding is assisted by the U2AF, which interacts with the polypyrimidine tract and the 3′splice site. U5 snRNP enters as a part of the tri-snRNP complex with U4 and U6 snRNAs and base-pairs with the last two nucleotides of exon 1 and the first two nucleotides of exon 2. This brings together the two phosphodiester bonds that must be broken and reformed during splicing reaction. Acting like a chaperone, U4 initially base-pairs extensively with U6. This interaction is then disrupted so that U6 can interact with U2 near the activated adenosine branch point (bp). As U5 and U6 form these interactions with the pre-mRNA, U1 leaves, allowing U6 access to the 5′ splice site. Through this trans-esterification reaction, the phoshodiester bond at the 5′ splice site is broken, and eventually exon 1 is joined to exon 2, resulting in the mature mRNA product and the lariat intron (adapted from http://www.designeduniverse.com/articles/Nobel_Prize/spliceosome_cycle.jpg, 2005).

(1.4.4) npcRNA Functions in Protein Stability

Another unique bacterial npcRNA termed transfer-messenger RNAs (tmRNAs) have been shown to affect protein stability. The ssrA RNA and 10Sa RNA are example of tmRNAs. They varied in size between 95 nt in *S.* Typhimurium LT2 to 466 nt in *E. coli* K12. The tmRNA is recognized as both tRNA and mRNA by stalled ribosomes (Gillet and Felden, 2001). When a ribosome is stalled, tmRNA is delivered to the site of the stalled ribosome and the nascent polypeptide is then transferred to the alanine-charged tRNA portion of tmRNA. The transcript is then placed by the mRNA portion of tmRNA, which encodes a tag for degradation of the stalled peptide. The mechanism and function of tmRNA is illustrated in Figure 1.7.

(1.4.5) npcRNA Functions in Nucleotide Modification of RNA

The biogenesis of eukaryal ribosomes in the nucleolus involves the processing of the rRNA primary transcript. Extended spacer regions from precursor rRNA are removed and mature ribosomal RNAs are generated. Before its cleavage by endonucleases and exonucleases, the nascent rRNA undergoes a pattern of nucleoside modifications. The two prevalent types of modification are 2‘-O-ribose methylation or pseudouridylation. Each of these modifications is found at about 50 to 100 sites per eukaryal ribosome in contrast to the *E. coli* ribosome, which contains only four ribose-methylated nucleotides and 10 pseudouridines (Maden, 1990; Cavaille and Bachellerie, 1998).

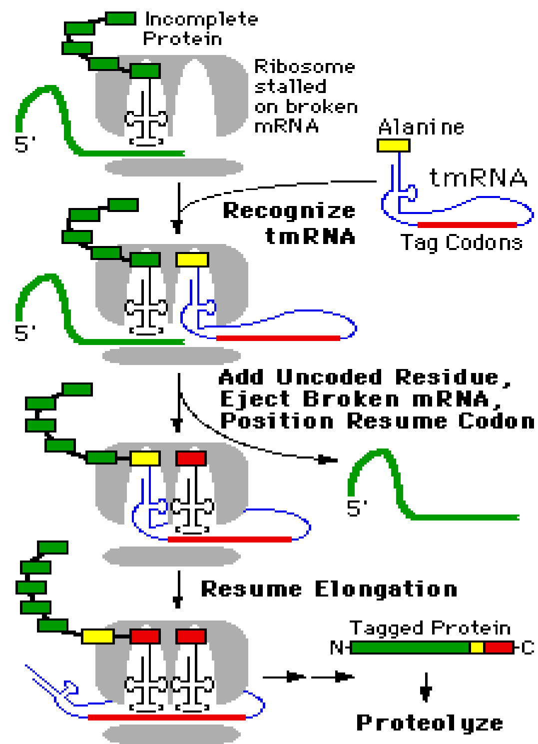

Figure 1.7: Mechanism and function of tmRNA. A tmRNA is delivered to the site of a stalled ribosome and the nascent polypeptide (green boxes) is then transferred to the alanine-charged (yellow box) tRNA portion of tmRNA. The transcript is then placed by the mRNA portion of tmRNA, which encodes a tag (red boxes) for degradation of the stalled peptide (adapted from http://www.indiana.edu/~tmrna/images/model.gif).

Both types of modification are directed by small nucleolar ribonucleoprotein particles (snoRNPs), each composed a small nucleolar RNA (snoRNA) and a set of proteins (Maxwell and Fournier, 1995; Smith and Steitz, 1997; Weinstein and Steitz, 1999). The specificity of modification is achieved by base pairing of snoRNA with pre-rRNA across the site to be modified. The snoRNAs are also involved in nucleotide modification in spliceosomal snRNAs in vertebrates (Tycowski *et al.*, 1996; Hüttenhofer *et al.*, 2001), tRNAs in Archaea and probably in eucaryotic mRNAs as well (Cavaille *et al.*, 2000).

The snoRNAs can be grouped into two major classes based on their structure and conserved sequence motifs (Balakin *et al.*, 1996). The C/D box family which guides 2'-O-ribose methylation and the H/ACA box family which guides pseudouridylation (Tollervey and Kiss, 1997; Smith and Steitz, 1997; Bachellerie *et al.*, 2002) snoRNA represent the most well studied npcRNAs. The structural feature of C/D and H/ACA box snoRNAs is illustrated in Figure 1.8 and Figure 1.9.

(1.4.6) npcRNA Functions in mRNA Stability and Translation

In bacteria, npcRNAs can also be involved in regulation of translation by base-pairing upstream of the initiator AUG start codon where they either block translation or compete with an inhibitory *cis*-acting secondary structure to open up the translation initiation site (Wassarman *et al.*, 1999; Wagner *et al.*, 2002). In *E. coli*, examples of these npcRNAs are OxyS RNA, DsrA RNA, and MicF RNAs. The mechanisms of action of these npcRNAs are briefly illustrated in Figure 1.10.

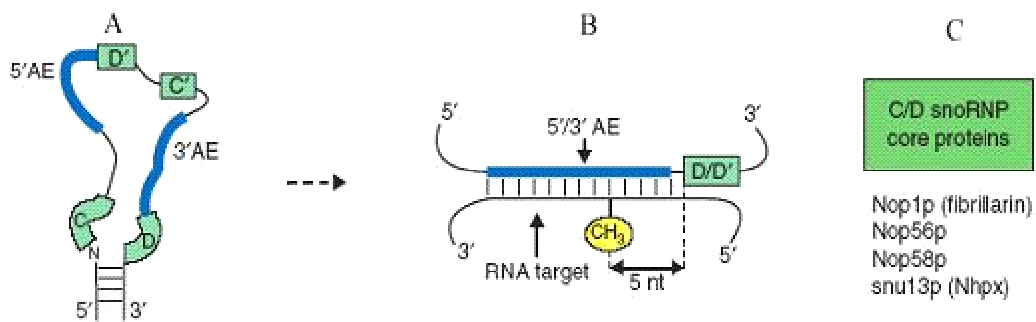

Fig. 1.8: Structural features of C/D box snoRNAs. (A) Schematic secondary structure of the eukaryotic C/D box snoRNAs. Conserved motifs are boxed and sequence tracts complementary to the cognate RNA target (antisense elements) are depicted by thick blue lines. (**B)** Canonical structure of the C/D guide RNA duplex. Indicated is the region of complementarity to target rRNA that is positioned 5' to the D or D' box; 2'-O-ribose methylation is directed to the nucleotide in the rRNA that participates in a Watson–Crick base pair five nucleotides upstream of the D' or D box. **(C)** Example of evolutionary conserved essential core proteins (adapted from Hüttenhofer *et al.*, 2004).

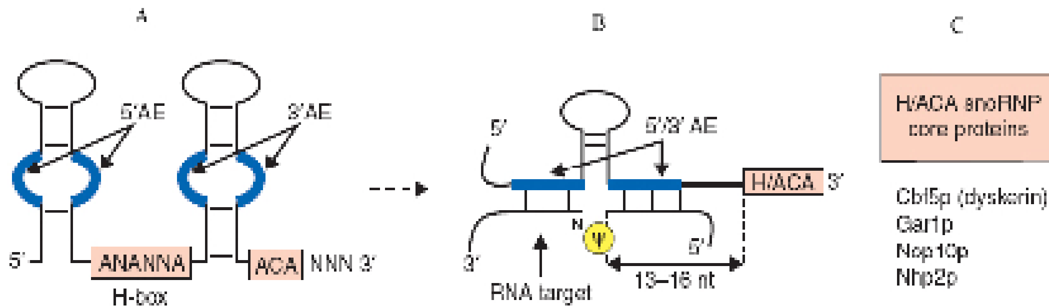

Fig. 1.9: Structural features of H/ACA box snoRNAs. (A) Schematic secondary structure of eukaryotic box H/ACA snoRNAs. Conserved motifs are boxed and sequence tracts complementary to the cognate RNA target (antisense elements) are depicted by thick blue lines. (**B)** Canonical structure of the H/ACA guide RNA duplex. Indicated is the region containing one or two sequence tracts complementary to rRNA that are located within the bulge region of the 5' or 3' helices; base pairing to rRNA positions the uridine nucleotide to be modified in the pseudouridylation pocket between the regions of rRNA-snoRNA complementarity. The target uridine is located 14 to 16 nt upstream from the ACA or H box. N: any nucleotide. **(C)** Example of evolutionary conserved essential core proteins (adapted from Hüttenhofer *et al.*, 2004).

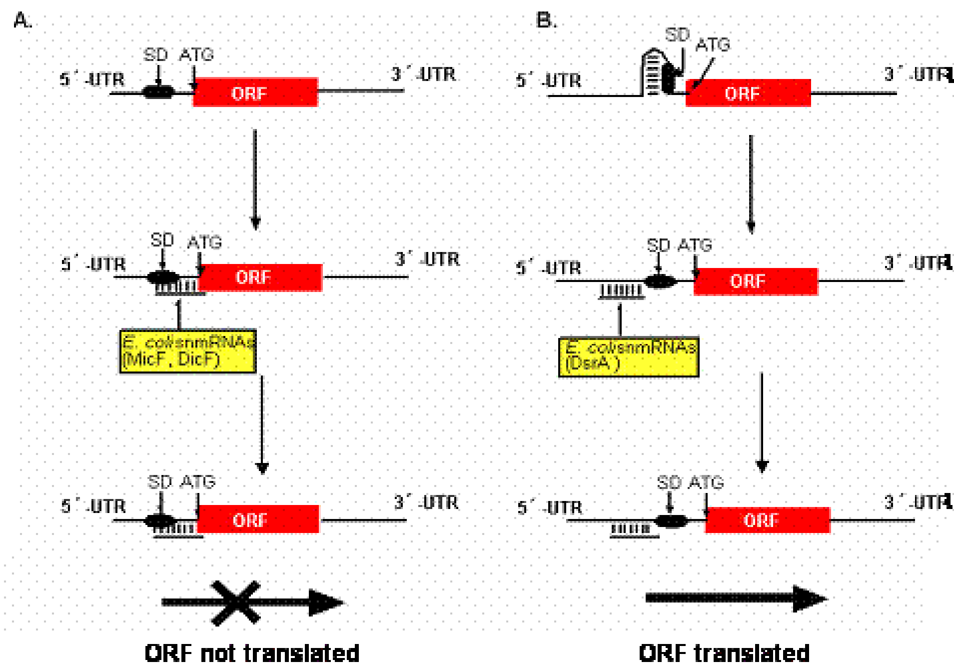

Fig. 1.10: Mechanism of translational inhibition/activation in *E. coli* directed by npcRNAs. (**A**) npcRNA binds to Shine-Dalgarno sequence and/or initiation codon, thereby inhibiting translation of the ORF. (**B**) npcRNA disrupts the double-strand RNA structures between 5'-UTR and SD thus allowing translation to occur. SD: Shine-Dalgarno sequence, ATG: initiation codon, ORF: Open reading frame, 5'-UTR: 5'-untranslated region, 3'-UTR: 3'-untranslated region (adapted from Tang *et al*., 2002).