A STUDY OF NON-CODING RNAs (ncRNAs) OF *Salmonella* Typhi DERIVED FROM SALMONELLA PATHOGENICITY ISLANDS UNDER DIFFERENT STRESS CONDITIONS

by

PRISCILLA SAW CHEAH PHEI

Thesis submitted in fulfillment of the requirement
for the degree of
Master of Science (Molecular Medicine)

June 2017

ACKNOWLEDGMENT

I would like to express my deep gratitude to my supervisor Assoc. Prof. Phua Kia Kien, for his patient guidance, advice, and assistance in keeping my progress smooth and to Dr. Khoo Boon Yin, my co-supervisor, for her useful and constructive recommendations on this project.

Moreover, I would like to extend my gratitude to Dr. Suresh, my field supervisor, for his patient guidance, enthusiastic encouragement and useful criticisms for this research work. I would also like to thank Mr. Sivachandran for his help in transcriptome sequencing data analysis and Mr. Siven who guided me in RNA molecular work when I was stationed in AIMST University.

In addition, I would like to extend my sincerest thanks to Dr. Kyoung Soo Kim from Kyung Hee University who provided me the THP-1 cell line for this research project. Furthermore, I would like to thank the technicians of the laboratory and administrative staff of INFORMM for offering resources for this research project.

I wish to acknownlegde the financial support from the Ministry of Higher Education (MoHE) in the form of Fundatmental Research Grant Scheme (FRGS) – 203/CIPPM/6711205 and MyBrain Scholarship. Finally, I wish to thank my parents and friends for their support and encouragement throughout my study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVATIONS

%	Percentage
µg	Microgram
µl	Microlitre
ºC	Degree Celsius
^{32}P	Phosphorus 32
A	Adenine
Approx.	Approximate
APS	Ammonium persulfate
ATCC	American Type Culture Collection
bp	Base pair
C	Cytosine
Cells/ml	Cells per millilitre
cfu/ml	Colony-forming unit per millilitre
CO_2	Carbon dioxide
$CoCl_2$	Cobalt chloride
Csr	Carbon storage regulatory
ddUTP	2',3'- Dideoxyuridine-5'-Triphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
E.coli	Escherichia coli
EDTA	Ethylenediaminetetra acetic acid
FBS	Fetal Bovine Serum
fmol/µl	Fentamol per microlitre
g	Gram
G	Guanine
g/ml	Gram per mol

GalR	Galactose operon repressor
GDPD	Glycerophosphodiester phosphosdiesterase
H_2O	Water
H_2O_2	Hydrogen peroxide
HCl	Hydrochloride acid
LA	Luria agar
LB	Luria Bertani
M	Molar
mfe	Minimum free energy
ml	Millilitre
mM	Millimolar
MOI	Multiplicity of Infection
mRNAs	Messenger RNAs
NA	Nutrient agar
NaOH	Sodium hydroxide
NB	Nutrient broth
NcRNAs	Non-coding RNAs
ng/µl	Nanogram per microlitre
ng/ml	Nanogram per millitre
OD	Optical density
OMPs	Outer membrane proteins
PBS	Phosphate buffered saline
PMA	Phorbol-12-myristate-13-acetate
RBS	Ribosomal binding site
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Revolution per minute
RPMI	Roswell Park Memorial Institute
rRNA	Ribosomal RNA
S. Typhi	*Salmonella* Typhi

S. Typhimurium	*Salmonella* Typhimurium
SDS	Sodium dodecyl sulphate
SPI	*Salmonella* Pathogenecity Island
SSC	Saline-sodium citrate
ssDNA	Single stranded DNA
T	Thymine
T3SS	Type III secretion system
T-75	TPP tissue culture flask 75cm^2
TBE	Tris/Borate/EDTA
TEMED	Tetramethylethylenediamine
THP-1cells	Human monocytic cells
TPR	Tetratricopeptide repeat protein
TSI	Triple Sugar Iron
UTR	Untranslated region
UV	Ultraviolet
V	Voltage
w/v	Weight per volume
g	Gravity force

PENYELIDIKAN TENTANG RNA BUKAN PENGODAN (ncRNA) DARIPADA *Salmonella* Typhi YANG DIPEROLEHI DARIPADA SALMONELLA PATHOGENICITY ISLAND DI BAWAH TEKANAN YANG BERBEZA

ABSTRAK

Bukan pengekodan RNA boleh didapati dalam semua bakteria termasuk *Salmonella* Typhi. *Salmonella* Typhi adalah terkenal kerana ia mampu menjangkiti manusia dan menyebabkan demam kepialu, satu penyakit sistemik yang boleh membawa kematian kepada manusia. Keunikannya ialah ia juga mampu bertahan dalam pelbagai jenis persekitaran termasuk dalam sel makrofaj. *Salmonella* Typhi adalah sama seperti bakteria lain yang juga mempunyai RNA seperti bukan pengekodan RNA sejenis molekul yang berfungsi di peringkat RNA yang memainkan peranan penting dalam laluan biologi. Mereka membantu menyesuaikan fisiologi bakteria sebagai tindak balas terhadap faktor persekitaran seperti pH, osmolariti, suhu dan lain-lain. Oleh itu, kajian ini bertujuan menyiasat bukan pengekodan-RNA Salmonella Typhi yang diekspresi dalam (differentiated) sel THP-1 yang bertukar kepada makrofaj dengan menggunakan penjujukan transkriptom untuk mencari kewujudan bukan pengekodan RNA yang unik terhadap faktor-faktor persekitaran tersebut. Walaubagaimana pun, hasil kajian tersebut tidak memuaskan dan terpaksa dihentikan akibat pencemaran sampel oleh RNA daripada sel THP-1 yang sukar untuk ditapis keluar. Maka, kajian alternatif dijalankan secara spesifik ke atas 10 ncRNAs menggunakan teknik Pemblotan Northern, IntaRNA; untuk meramalkan sasaran

mRNA, dan RNAfold untuk menentukan struktur sekunder. Semua ncRNA tersebut disiasat dalam keadaan stres yang berbeza seperti berasid, beralkali dan keadaan oksidatif untuk dibandingkan dengan keadaan pertumbuhan sel dalam keadaan normal. Berdasarkan keputusan kajian, Styr-9 dan Styr-161 telah didapati terlibat dalam mekanisme patogenik berdasarkan ramalan dari perisian (IntaRNA), menunjukkan mereka berinteraksi dengan rfbI dan mRNA fur masing-masing. Styr-381 pula didapati berinteraksi dengan rbsA mRNA yang terlibat dalam pembentukan pengangkut ABC yang mungkin secara tidak langsung terlibat dalam mekanisme patogenik. Manakala StyR-103 dan Styr-137 terlibat dalam proses selular di mana pembentukan sampul luar sel dan pembahagian sel masing-masing berlaku. Hasil kajian ini menunjukkan kewujudan ncRNA yang berpotensi mempunyai peranan dalam mekanisme patogenik oleh *S.* Typhi terhadap manusia. Maka kajian in perlu diteruskan lagi untuk lebih memahami peranan patogenik *S.* Typhi.

A STUDY OF NON-CODING RNAs (ncRNAs) OF *Salmonella* Typhi DERIVED FROM SALMONELLA PATHOGENICITY ISLANDS UNDER DIFFERENT STRESS CONDITIONS

ABSTRACT

Non-coding RNAs can be found in all bacteria including *Salmonella* Typhi. The *Salmonella* Typhi is infamous for infecting humans and causing typhoid fever, a systemic disease that may be mortal to humans. It is able to survive in various environments including in macrophags. *Salmonella* Typhi is similar to other bacteria by consisting of non-coding RNAs that play crucial roles in its biological pathways in response to environmental. One of the aims in this study is to investigate non-coding RNAs of *Salmonella* Typhi in differentiated THP-1 cells to find novel non-coding RNA expressions during invasion of macrophages via transcriptome sequencing. However, the investigation was later suspended due to the challenge of extracting *S.* Typhi RNA away from the large RNA pool of the differentiated THP-1 cells, hampering the transcriptome analysis of *S.* Typhi's RNAs. Therefore alternative approached was studied where the expressions of 10 specific ncRNAs under different stress were investigated using Northern blot technique, as well as IntaRNA software to predict their target mRNAs, and RNAfold software to determine their secondary structures. The various stresses were conditions to compare with cell growth under normal conditions. Based on the results, StyR-9 and StyR-161 were found to be involved in *S.* Typhi's virulence mechanism, by interacting with rfbI and fur mRNA, respectively. StyR-381 was found to

interact with rbsA mRNA that is involved in ABC transporter formation, which may be indirectly involved with the virulence mechanism. StyR-103 and StyR-137 were involved in cellular processes, specifically cell envelope formation and cell division, respectively. The study thus showed that computational approaches are convenient, less time-consuming and more preferable compared to conventional 'wet lab' methods for analysis of RNA molecules. However, the outcomes of this study will need to be further studied for their significance in understanding the pathogenic *S.* Typhi bacteria.

CHAPTER 1

INTRODUCTION

1.1 Typhoid fever

Typhoid fever is a globally common enteric fever caused by consuming food and water contaminated with *Salmonella enterica* serovar Typhi (*S.* Typhi). This food-borne pathogen has recorded approximately 16 million cases of infection,with 600,000 related deaths worldwide annually (Kidgell *et al.*, 2002). *S.* Typhi is not easy to eradicate as some infected individuals may remain as asymptomatic carriers and continuously shed the bacteria in their faeces, thus further spreading the infection unknowingly (Buchwald & Blaser, 1984).

1.2 *Salmonella* Typhi

S. Typhi is a gram negative bacillus bacteria that belongs to the *Enterobactericea* family and is known to be able to survive in a wide range of conditions in terms of transmission, pathogeneses and survival in the environment. This organismis thus capable of surviving in contaminated food or water, different bodily sites with different micro-environments and environmental reservoirs. This is because it has the ability to change its genotype and phenotypes to adapt to different conditions andundergo significant biological changes during transitions from external environments to hosts, between different locations within a host and between different hosts.

Passage from the external environment into the host is associated with a number of cellular stresses such as changes in temperature, osmolarity (intestine), shifts in pH (stomach), exposure to toxic reactive oxygen species and nitrogen intermediates, as well as starvation for iron and other nutrients. Hence to survive in these different conditions, *S.* Typhi must be capable of changing gene expressions rapidly within the host.

Unfortunately *S.* Typhi's stress response for survival in mammalian hosts also triggers the virulence response that causes the systemic disease of typhoid fever in humans that leads to mortality. Itspathogenic virulence mechanisms are controlled by environmental factors such as oxidative stress, pH, osmolarity, iron limitation and others (Chowdhury *et al.*, 1996) by altering it genotype and phenotype expressions a sophisticated mechanism for sensing external condition and respond. The outer membrane proteins (OMPs) of *S.* Typhi are the first to be exposed to environmental stimuli,acting as an interface for the cell, which then send signals to the bacteria for both rapid survival and virulence responses (Hamid & Jain, 2008).

1.3 Overview of non-coding RNA (ncRNA)

For bacteria to regulate their responses accurately based on specific stress conditions, ncRNAs must play their roles. NcRNAs in prokaryotes are not well known until recently, when a huge number of these RNAs were discovered in E. coli. Unlike prokaryotes, the ncRNAs in eukaryotes have been well documented. They are also known as micro RNAs or short interfering RNAs. They act as novel regulators for the development of

eukaryotic cells and cell death, and may even suppress specific gene expressions by binding to the target mRNAs or proteins. Just as in eukaryotic cells, prokaryotic ncRNAs play significant roles in manipulating gene expressionsat post-transcriptional level by suppressing or initiating translation, or by stabilising the folding of their target mRNAs to facilitate translation. They are mainly found in extrachromosomal elements (Herbig & Nieselt, 2011) and located in intergenic regions of the chromosome (Argaman *et al.*, 2001).

During stress response and virulence mechanism, ncRNAs in bacteria have been identified to play many crucial roles (Gottesman, 2004; Storz *et al.*, 2004; Storz *et al.*, 2005). Due to vast and rapid exposure to different environmental stresses, bacterial responses must also be rapid in order to survive and replicate. Thus the existence of ncRNAs are vital as they are small molecules that require only small amounts of energy to be transcribed and are functional at RNA level.

For examples, ncRNAs in E.Coli and *S.* Typhimurium were found to be highly regulated during many stress conditions. As such, their initial response to environmental stimuli is to regulate the synthesis of specific ncRNAs. When the level of ncRNAs are increased, thesebacteria can rapidly initiate their regulatory network to work on stress responses. Bacterial ncRNAs are unique in a sense that it binds to mRNA with imperfect base-pairing interaction (Waters & Storz, 2009). NcRNAs that are involved in the regulatory system may have different effects on their target mRNAs. Most ncRNAs usually bind at or near to the ribosomal binding site (RBS) of their

target mRNA, preventing RNA polymerase from binding to the mRNA, thus inhibiting the translation process. Once the mRNA is not accessible, degradation will occur with the help of RNAse E. On the other hand, some ncRNAs bind to their target mRNAs to unwind or increase the accessibility to their ribosomal binding sites, further stabilizing the mRNAs for translation process to occur.

Prokaryotic ncRNAs may also bind to proteins such as the CsrB in E. Coli (Babitzke & Romeo, 2007). Past research showed that ncRNAs bind to proteins to prevent or activate their functions. This shows that ncRNAs are able to optimally coordinate the regulatory system by affecting changes at all synthesis levels in a prokaryotic cell. Furthermore, a number of ncRNAs are able to bind and regulateseveral target mRNAs, such as RybB, making them quite an efficient regulator of gene expressions (Massé *et al.*, 2005). Most ncRNAs consist of a rho-independent terminator, and a poly (U) at the 3'-end protectingthem from 3'-exonucleases (Storz *et al.*, 2011) that are essential for Hfq action for ncRNA stabilization (Otaka *et al.*, 2011).

Hfq is a protein that acts as a chaperone that mediates RNA-RNA interactions, especially between ncRNAs and their target mRNAs, either for repressing or facilitating the translation process. Hfq binds to both ncRNAs and their target mRNAs to bring them together,though the mechanisms are yet to be fully understood (Aixia Zhang *et al.*, 2003; Storz *et al.*, 2004; Sittka *et al.*, 2008). Thus, Hfq is another essential component in facilitating stress and virulence response of gram negative bacteria for their growth and survival, especially inside mammalian / human hosts. However not all ncRNAs require Hfq in regulating different responses.

1.4 Examples of non-coding RNAs and theirroles in bacterial stress responses

One of the typical responses of a host's innate immune response upon pathogen entry is to produce reactive oxygen species (ROS), which are highly toxic and reactive molecules present in macrophage cells. These macrophages usually engulf and destroy pathogens by phagocytosis (Andrews, 2012). However, most enteric pathogens are capable of surviving under this oxidative stress. OxyS, an ncRNA in *Salmonella*, is highly expressed when *Salmonella* enters macrophages. FhlA, a transcriptional activator for formate metabolism, is repressed by OxyS during the oxidative stress. It was found that OxyS bind to the RBS of the fhlA to repress and degrade the mRNA through a 'kissing' complex formation between them that form a stable antisense target complex. The formation is found to be sufficient to repress the translation process of the fhlA (Argaman *et al.*, 2001). As the results, the kissing formation could repress fhlA translation.

Moreover, the presence of pathogenic bacteria within a host may also affect the host’s physiological system by causing an imbalance of the levels of pro-oxidant and anti-oxidant molecules. Pro-oxidants are molecules such as reactive oxygen species (ROS) that oxidize proteins and lipids, possibly causing cell death, whereas anti-oxidants are molecules that counteract the reactive toxicity of pro-oxidantsthat are damaging the cells (Berghoff *et al.*, 2013). In E. coli, the SOS response leads to the activation of TisB gene, producing a toxic protein that, when overexpressed,causes cessation of growth and may lead to cell death via programmed response to DNA

damage (Dörr *et al.*, 2010). To counteract this effect, istR-1 (75nt) is highly expressed during the oxidative stress to inhibit TisB mRNAs by blocking their RBS (Vogel *et al.*, 2004). IstR-1 is an ncRNA that is regulated by LexA, a repressor enzyme that repress genes related to SOS response (Butala *et al.*, 2008).

Apart from that,prior to reaching the intestines, enteric pathogens need to survive through the host's stomach, which has a highly acidic environment. Most pathogens,however, prefer to survive inenvironments with neutral pH, thereforeadaptations are required upon entry into hosts in order to survive in the stomach (Darwin, 2013). Bacteria responds to an acidic condition such as the stomach by remodelling its envelope structure once the σ^E factor system is activated by the stress. GadY, an ncRNA in E. coli, is accumulated in an acidic environment,positively affecting GadX mRNAs expression. GadX mRNA is a transcriptional activator in the acidic response mechanism, whereby it aids the expression of GadA and GadB, and glutamate decarboxylase (Opdyke *et al.*, 2004). Glutamate decarboxylase is an enzyme involved in glutamate metabolism, that is needed to convert glutamate into ɣ-aminobutyric acid to allow survival of E. coli in low pH conditions (Castanie-Cornet *et al.*, 1999). The binding of GadY ncRNA preventsdegradation of GadX mRNA by RNAse E by blocking the recognition site of RNAse E thereby increasing the level of GadX mRNAs available in the cells.

Bacteria may also face nutrient-limited conditions, especially in batch cultures wherenutrient substrates eventually deplete and force the bacteria

to enter stationary phase. In the last stage of survival with limited nutrients, bacterial growth will be halted and their metabolic system will utilize the nutrients in a pulsed manner until cell death occur. Thus, during that stage, bacteria maintain at the maximum growth at certain rate before cell death occurred (Bren *et al.*, 2013).

RprA is an ncRNA that is triggered and overexpressed during stationary phase and biofilm formation, which in turn activates the RpoS mRNA, a global regulator. RprA positively regulates RpoS mRNA by an antisense mechanism that relieve the intramoleculesof the target mRNA to initiate translation process (Majdalani *et al.*, 2002; McCullen *et al.*, 2010). During biofilm formation,RprA activates RscC/RcsD/RscB, a three-component system, involved in signal transduction for biofilm maturation and capsule synthesis in E. coli (Majdalani *et al.*, 2005; Mika *et al.*, 2012).

Many pathogens such as E. coli and *S.* Typhimurium remodel their envelopeonce encounteror interact with the host. The remodelling ofenvelope and outer membrane structures are also responsible for the activation of the virulence mechanisms in E. coli and *S.* Typhimurium. These pathogenic pathways lead to the formation of Type Three Secretion System (T3SS) and pili. T3SS is a supracomplex protein structure that is located at the inner and outer membrane of the bacterial envelope. This system is usually found in gram negative bacteria and it has a needle-like structure that is able to puncture the host's cell and release pathogenic proteins called effectors inside the host's cell. The effectors are the virulent proteins that will manipulate the host's system by interfering with its cytoskeleton, allowing the pathogen to colonize and multiply within the

host cell by subversion of the host's immune system (Coburn *et al.*, 2007). The relationship between the virulence mechanisms and stress response in these pathogens are important in order to initiate host invasion, and the regulation of these mechanisms involve ncRNAs. In *S.* Typhimurium,for example,an ncRNA called IsrJ is found to influence its T3SS expression and is regulated by HilA, a global regulator for invasion. HilA responds to stress signals such as osmolarity, pH, oxygen limitation and others, which in turn will trigger IsrJ expression. Mutation of the IsrJ has shown that the T3SS effector could not be transferred to the host cytoplasm, supporting the conclusion that IsrJ ncRNA is part of T3SS regulatory system (Padalon-Brauch *et al.*, 2008).

1.5 *Salmonella* Typhi non-coding RNAs and transcriptome sequencing

Most findings on ncRNAs are from studies using *S.* Typhimurium and E. coli, which are enteric pathogens, similar to *S.* Typhi. *S.* Typhimurium, in particular, has been the model to study typhoidal diseases as it commonly infects mammals such as humans and mice, and cause systemic diseases, though mortality is uncommon in humans. *S.* Typhi, on the other hand, is a host-specific pathogen with unique virulence mechanisms, causing typhoid fever in humans that may cause mortality. Due to *S.* Typhi's specificity and uniqueness, there remains a challenge in fully understanding and eradicating typhoid fever.

With the advance technology, discovery of RNA has become much more approachable to study transcriptome profiling using advance RNA-sequencing. The transcriptome sequencing is the study or analysis of complete set of transcripts and their expression of a particular cell, for a specific developmental stage or physiological condition. Understanding the transcriptome is essential for interpreting the functional elements of the RNA and revealing the molecular constituents of cells and tissues, and also for understanding development and disease. The key aims of transcriptomics are: to catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs; to determine the profiling, transcriptional structure of genes, in terms of their ribosomal binding site, folding, splicing patterns and other post-transcriptional modifications; and to quantify the changing expression levels of each transcript during development and under different conditions or/and disease (Perkins *et al.*, 2009; Srikumar *et al.*, 2015).

Although transcriptome sequencing is still a technology under active development, it offers several key advantages over existing technologies. For example, unlike hybridization-based approaches, transcriptome sequencing is not limited to detecting transcripts that correspond to existing genomic sequence but it also able to reveal the precise location of transcription boundaries, to a single- base resolution. Furthermore, 30-bp short reads from transcriptome sequencing give information about how two exons are connected, whereas longer reads or pair-end short reads should reveal connectivity between multiple exons. In addition, transcriptome sequencing can also reveal sequence variations (for example, SNPs) in the transcribed regions (Nicole Cloonan & Grimmond, 2008) and possible to find novel RNA that are expressed particularly in specific conditions (Weirick *et al.*, 2016).

The aims of transcriptome sequencing are not only to catalogue all mRNA transcript but also non-coding RNA (ncRNA). Many successful research using transcriptome sequencing found novel ncRNAs in many species (Argaman *et al.*). With the advance sequencing in the market such as Illumina sequencing that is capable of sequencing ncRNA provided 1µg of high purity of total RNA as input with RNA integrity of 8.

More studies on the RNA-omics of *S.* Typhi have been emerging over the past few years, and based on the latest paper published by Suresh V. Chinni in 2010, 97 novel ncRNAs have been found in *S.* Typhi by transcriptome sequencing (Chinni *et al.*, 2010).

Because of these advantages, trascriptome sequencing will undoubtedly be valuable for understanding transcriptomic dynamics during development and normal physiological changes, and in different conditions, where it will allow robust comparison under normal growth and different conditions, as well as the ncRNAs expression of typhoid disease.

1.6 Understanding non-coding RNA using Northern blot technique

Analysis of RNA expressions using Northern blots under various growth conditions is frequently the first step followed by computational approaches to identify and understand the physiological importance of ncRNAs in bacterial regulatory system (Massé & Gottesman, 2002; Johansen *et al.*, 2006). Northern blot is still commonly used to identify, validate and study the expression profile of ncRNAs because it is a relatively inexpensive technique for RNA studies. It is a useful technique that allows comparisons between different samples, such as from different growth phases and stress conditions,showing any target ncRNAs that are overexpressed in the samples. Furthermore, this method can guide researchers to uncover new ncRNA functions under different stress conditions for further individual studies. The other advantage of this method is that the membrane used in Northern blot can be re-probed and stored for many years.

In the past, the Northern blot technique usually refers to a radioactive blot technique where nucleotides labelled with ^{32}P enable highly specific probe activities, allowing the detection of as low as 10 fg (femtogram) of target RNA. However, the radioactive^{32}P has now been replaced with fluorescence or chemiluminescent dyes for non-radioactive alternatives. The

chemiluminescence method, where CDP-Star is used as substrates, is a good alternative as it is rapid, non-destructive, and is able to detect as low as 50 fg target RNA (J. Osborn, 2000). Whereas nucleotides tagged with fluorescein are detected with highly anti-fluorescein alkaline phosphate (AP) enzyme.

1.7 Computational approach for non-coding RNAs

Despite the numerous number of prokaryotic ncRNAs found in recent years, only a small portion of them has been thoroughly studied. One of the most important aspects in studying an ncRNA is its target-binding. The sequence-specific binding is important to achieve stability, followed by facilitation of translation or suppression of the target (Pain *et al.*, 2015).

In addition, bacterial ncRNAs are thought to regulate more than one target mRNAs. For example, RyhB ncRNA that is bound by Hfq protein to increase interaction with target mRNAs were found to be also highly interactive with target mRNAs in iron homeostasis, such as napF, soda, cysE, yciS, acpS, nag Z and dadA (Tjaden *et al.*, 2006). Over the past decade, many sophisticated algorithms have been developed with improved accuracy in predicting possible targets of ncRNAs. Subsequently, computational methods are now widely used to predict the target mRNAs of novel ncRNAs and their secondary structures with high accuracy, reduced time and labos. As a result, many prediction models have been designed for these purposes, as depicted in Table 1.1 (below) (Tjaden *et al.*, 2006; Tjaden, 2008; Wenqian Zhang *et al.*, 2015).

Table 1.1: Major types of computational tools for prediction of bacterial ncRNAs and their target mRNAs (Li *et al.*, 2012)

Type	Tool	Main features
Comparative genomic-based models for sRNA prediction	QRNA	Sequence and secondary structure; suitable for two sequence alignment
	sRNAPredict	Sequence and Rho-independent terminators
	NAPP	Phylogenetic profiling of nucleic acid fragments; cluster analysis
Prediction model for general RNA-RNA interaction	RNAhybrid	Extension of minimum energy folding algorithm to sequence; neglecting intra-molecular base-pairings and multi-loops
	RNAplex	Extension of minimum energy folding algorithm to two sequences; running faster
	RNAup	Consideration of accessibility of binding sites
Prediction models for sRNA-targt mRNA interactions	TargetRNA	Hybridization; not consider structures from sRNA or mRNA
	IntaRNA	Accessibility of binding sites; user-specified seed
	sRNATarget	Sequence and RNA secondary structure profile; naïve Bayes method
	RNApredator	Target site accessibility; RNAup

1.7.1 NcRNA secondary structure prediction using RNAfold

Bacterial ncRNAs are dynamic gene expression regulators, with functions applicable for environmental sensing to pathogenesis. Predicting ncRNA secondary structures is therefore essential because interactions with their target mRNAs are based on their secondary structures. NcRNA folding is based on canonical base-pair whereby the structure able to show that each nucleotides is paired or not (David H. Mathews *et al.*, 1999). Recently ncRNAs has showed that it is associated with virulence mechanism in host induced expression in *S.* Typhi. For example, StyR-234 ncRNA, was found to be involved in multi drug resistance regulatory system in *S.* Typhi (Balakrishnan *et al.*, 2014). The function of ncRNA is strongly based on its secondary structure and prediction of ncRNA structure has become a major consideration in order to understand the mode of mechanism as we know this regulators play its part at RNA level. Interaction that is governed by secondary structure is stronger compare to tertiary structure because it determine the next interaction (Banerjee *et al.*, 1993; Jaeger *et al.*, 1993; Mathews *et al.*, 1997). Misfolding of their secondary structures may hinder the binding to their target mRNAs. Therefore, secondary structure is determined independently from tertiary structure.

The initial step of understanding ncRNA mechanism is to determine its secondary structure based on a set of canonical base pairings (i.e.: AU, GC and GU) (Gutell *et al.*, 2002). The three base pairs are considered as stabilizing the folding structure of the RNA. Secondary structures of ncRNAs are thus formed based on these canonical base pairs, and pairing were largely establishes as the results of free energies that represent by

thermodynamics of the RNA folding. In addition, RNA secondary structures are well conserved throughout evolution, many computational algorithms have been constructed based on these conserves. The RNA folding is based on free energy minimization because it can determine the target sequence for interaction (Mathews *et al.*, 1997) and it is used to interpret single RNA molecule (Alkan *et al.*, 2006).

The secondary structure of ncRNAs can be predicted based on thermodynamic algorithms (David H. Mathews & Turner, 2006). Free energy minimization is the common method that is based on the sequence approximation interaction to achieve stability with minimal free energy structure or commonly known as nearest-neighbour model (SantaLucia & Turner, 1997). The priority in RNA folding is to achieve the lowest free energy structure to form a stable RNA secondary structure. Based on this algorithm, prediction of single sequence secondary structure is reasonably accurate and many software has been programmed based on this thermodynamic formula. One of the computational approaches that has been established based on this formula is the RNAfold software. The RNAfold software is thus used in this experiment (Zuker & Stiegler, 1981) It is a free software from Vienna RNA Package (Hofacker *et al.*, 1994).

1.7.2 NcRNA target prediction (IntaRNA)

IntaRNA is a new computational approach that is used to predict specific ncRNA interaction with mRNAs using a combined energy score of the interaction that is calculated as the sum of the free energy of hybridization and the free energy required for making the interaction sites accessible. Additionally, IntaRNA is able to predict the exact locations of the RNA–RNA interactions. The IntaRNA is thus a valuable and free tool to facilitate the search for possible ncRNA targets, and its prediction accuracy is close to more complex methods like RNAup or biRNA. Furthermore, it is less time consuming to make prediction analysis compared to RNAplex or RNAup (Eggenhofer *et al.*, 2011). Hence, IntaRNA will be used in this research to predict and study the predicted targets, in comparison to prediction results from other softwares such as TargetRNA, RNAhybrid, RNAplex and RNAup (Busch *et al.*, 2008).

CHAPTER 2

STUDY OBJECTIVES

The study is generally aimed to find and develop a fundamental understanding of novel ncRNAs from *S.* Typhi invading macrophage cells. Intial purpose of this study is to detect any possible *S*. Typhi novel ncRNAs and its expression in infected THP-1 cells.The details objectives as below:

1. To determine the multiplity of infections (MOIs) of *S*. Typhi on differentiated THP-1 cells at the optimal concentration.

2. To determine *S*. Typhi novel ncRNAs and its expression in differentiated THP-1 cells using transcriptome sequencing.

However, the investigation was later suspended due to the challenge of extracting *S.* Typhi RNA away from the large RNA pool of the differentiated THP-1 cells, hampering the transcriptome analysis of *S.* Typhi's RNAs. Following the initial objective of the study, altenative approached was carried where the expressions of ncRNAs derived from *Salmonella* Pathogenicity Island (SPI) under different stress conditions are examined. The detailed objectives are as below:

1. To determine SPI derived ncRNA expression of *S*. Typhi in various stress conditions.

2. To postulate possible binding interactions between each ncRNA and target mRNAs using IntaRNA.

3. To determine ncRNA secondary structures using RNAfold.

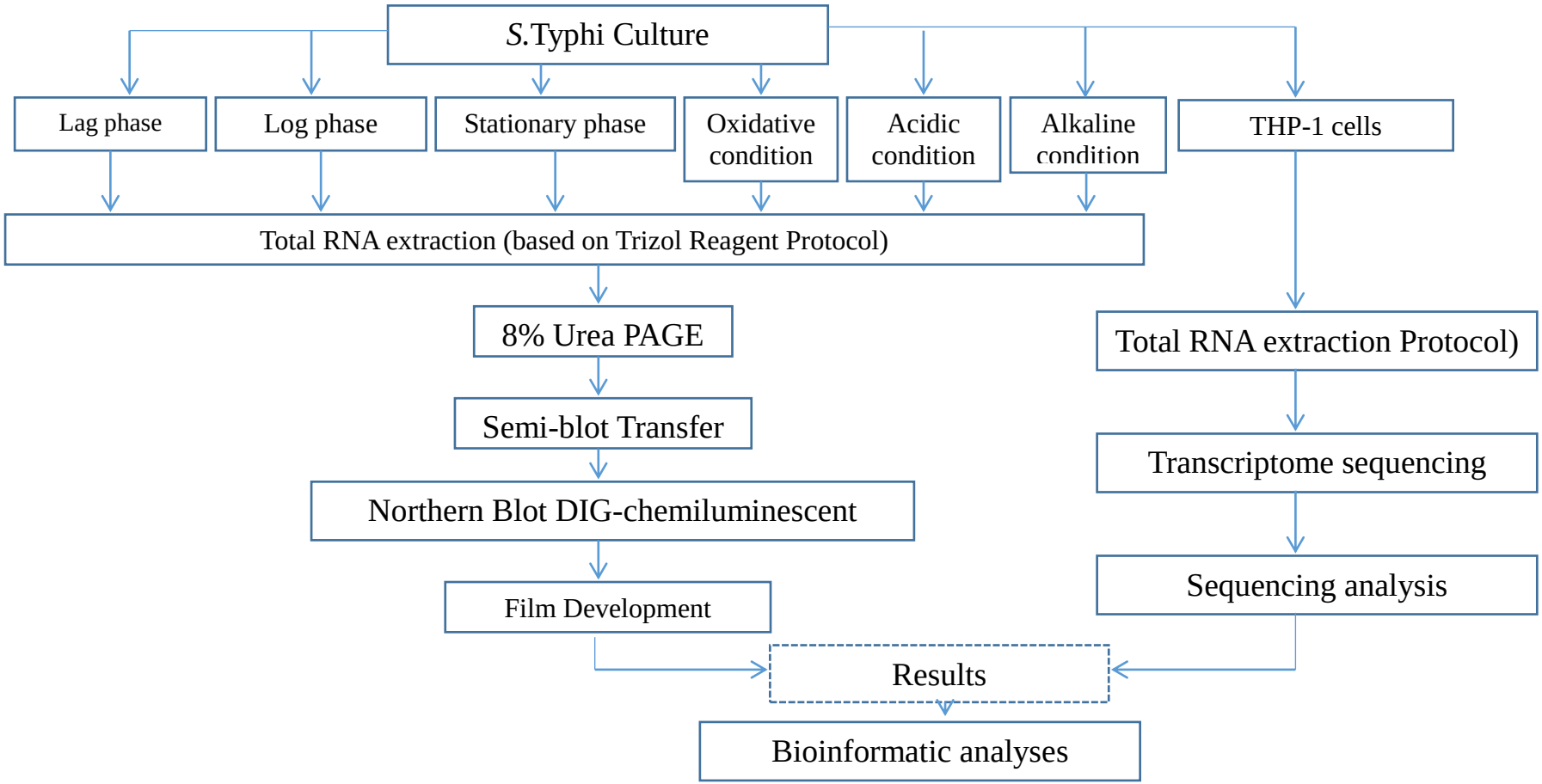

Figure 2.1: Flow diagram of study

CHAPTER 3

MATERIALS

1. **Strain:** *Salmonella* Typhi (ATCC 7251)
2. **Luria Bertani (LB) broth, HIMEDIA, India**

   25 g of LB broth powder was added to 1000 ml of distilled water. The medium was mixed and autoclaved.
3. **Nutrient Agar (NA), HIMEDIA, India**

   28 g of NA was mixed with 1000 ml of distilled water. The mixture was then autoclaved, then left aside until its temperature drop to 70°C to 80°C. Then it was poured into petri plates.
4. **Glycerol Stock**

   <u>50% Stock solution</u>

   5 ml of 100% glycerol stock is mixed with 5 ml of double distilled water (ddH_2O).

   <u>10% Working solution</u>

   0.2 ml of 50% glycerol stock solution is added into 0.8 ml of LB broth.

5. Biochemical Test Preparation

Simmon's citrate Agar (approx. 72 test tubes) – 1.5 ml/test tube

Simmon's citrate agar	3 g
dH_2O	100 ml

Simmon's citrate agar was boiled till completely dissolved in dH_2O, then aliquoted by 1.5 ml into test tubes. The test tubes were then autoclaved and left to cool diagonally for 24 hours at room temperature. Once solidified, the test tubes were kept at 4°C.

Triple Sugar Iron (TSI) Agar (approx. 125 test tubes) – 4 ml/test tube

Triple Sugar Iron agar	32.5 g
dH_2O	500 ml

The TSI agar was boiled till completely dissolved and aliquoted by 4 ml per test tube. The test tubes were then autoclaved and left to cool diagonally at 37°C for 24 hours. Once solidified, they were kept at 4°C.

Salmonella Shigella (SS) Agar

SS Agar	30 g
dH_2O	500 ml

The *Salmonella Shigella* Agar was boiled till completely dissolved and autoclaved. Once autoclaved, the medium was poured into petri plates. The medium petri plates were left to solidify, then kept at 4°C.

6. R2A Agar

18.12 g of agar powder was dissolved in 1000 ml of distilled water. The mixture was mixed and dissolved by heating it for 2 min with agitation. The medium was then autoclaved at 121°C for 15 min. Once the mixture was cooled to 60°C, it was poured into petri plates.

7. H_2O_2 preparation (oxidative condition)

Formula weight = 34.01 g/mol 30% w/v (100vol)

30% = 30 g in 100 ml of water,

Thus 300 g is 34.01 g/mol = 8824 mM

5000 mM preparation

M1V1=M2V2

5000(V1) = 1000 (20 ml)

= 11.322 ml of 8824 mM H_2O_2 + 8.678 ml of double distilled H_2O

1000 mM preparation

5000(V1) = 1000(20 ml)

= 4 ml of 5000 mM H_2O_2 + 16 ml of double distilled H_2O

10 mM preparation

1000(V1) = 10 (20 ml)

= 200 µl of 1000 mM H_2O_2 + 19.8 ml of double distilled H_2O

0.38 mM preparation

10(V1) = 0.38 (20 ml)

= 760 µl of 10 mM H_2O_2 + 19.24 ml of LB broth

The LB broth with 0.38 mM of H_2O_2 was autoclaved at 121°C for 15 min.

8. pH 4 LB broth

LB broth with pH 4 was prepared by adding HCl dropwise into the LB broth until it reached pH 4 using pH meter. The broth was then autoclaved at 121°C for 15 min.

9. pH 8.6 LB broth

LB broth with pH 8.6 was prepared by adding NaOH dropwise into the LB broth until it reached pH 8.6 using pH meter. The broth was then autoclaved at 121°C for 15 min.

10. THP-1 Cell Culture

Type of Cells: THP-1 cells (human monocytic cells/ monocytic leukemia cells)

RPMI 1640 Medium (500 ml per bottle), GIBCO, USA

44.5 ml of RPMI 1640 medium was aliquoted into a 50 ml centrifuge tube. The medium was stored at 4°C.

Fetal Bovine Serum (FBS) (100 ml per bottle), GIBCO, USA

The FBS was placed in a water bath for 30 min at 56°C to be heat inactivated. Then 10 ml was aliquoted into a 15 ml of centrifuge tube. The medium was stored at -20°C.

Penicillin-Streptomycin (10,000 U/ml) (100ml per bottle), GIBCO,USA

Prior to use, the antibiotics were kept at 4°C. 1 ml of the antibiotics was aliquoted into a 1.5 ml centrifuge tube and kept at -20°C.

Complete growth medium

The complete growth medium was the combination of RPMI 1640, FBS and antibiotics as below:

44.5 ml of RPMI 1640

5 ml of FBS

0.5 ml of antibiotics

Once prepared, it was kept at 4°C.

Cell freezing medium

14 ml of RPMI 1640

4 ml of FBS

2 ml of DMSO

The ingredients were mixed together with 20 ml of freezing medium. The medium was then kept at 4°C.

Phorbol 12-myristate 13-acetate (PMA) – In powder form, SIGMA, USA

1 mg was mixed with 200 µl DMSO

1 mg/200 µl = 1000000 ng/200 µl = 5000 ng/µl

2000 ng/ml of PMA

2000 ng/ml: M1V1=M2V2

= (2000/1000) x 10 x (1/5000)

= 4 µl

Then, 4 µl of PMA (5000 ng/µl) was added to 9996 µl double distilled H_2O.

Complete medium with 30ng/ml of PMA

45.5 ml of RPMI medium was mixed with 5 ml of FBS and 0.5 ml of penicillin/streptomycin. 770 µl of 2000 ng/ml of PMA was then added to the mixture.

Medium 2

45.5 ml of RPMI medium was mixed with 5 ml of FBS only. Antibiotics were not added included in this medium.

Gentamicin preparation (10mg/ml), GIBCO, USA

Stock = 10 mg/ml, or10000 µg/ml

1000 µg/ml preparation:

M1V1 = M2V2

10000 (V1) = 1000 (50 ml)

V1 = (1000 x 50 ml)/ 10000