Studies on Natural Antisense RNAs and microRNAs

Omid Reza Faridani
STUDIES ON NATURAL ANTISENSE RNAS AND MICRORNAS

Omid Reza Faridani

Stockholm 2007
ABSTRACT

Regulatory RNAs are found in all kingdoms of life and involved in regulation of gene expression at various steps including RNA splicing, editing, stability, modification, export, translation and chromatin remodelling. A large number of regulatory RNAs have been described recently. Natural antisense RNAs are examples of regulatory RNAs that contain complementary sequences to other transcripts. They can be transcribed in cis from opposing DNA strands at the same locus or in trans from separate loci. Some natural antisense RNAs hybridise to their target RNA, forming double-stranded RNA. MicroRNAs are well known examples of trans acting natural antisense transcripts that partially hybridise to targets. To study regulatory RNAs, appropriate molecular tools are needed. Therefore, in this thesis we developed methods to detect physical RNA::RNA interactions, double-stranded RNA and microRNAs. We next applied these methods in a variety of biological systems.

To study physical RNA::RNA interactions in growing cells, we considered the hok/sok toxin–antitoxin plasmid stabilization locus of the R1 plasmid in *Escherichia coli* as a paradigm. We designed anti Sok peptide-peptide nucleic acid (peptide-PNA) oligomers that would inhibit hok mRNA::Sok-RNA interactions and release the hok toxin mRNA for translation. Results showed that anti Sok peptide-PNAs were bactericidal with phenotypes consistent with cell killing by Hok protein, supporting the model for hok mRNA::Sok-RNA interactions. We also observed that peptide-PNAs accumulate in cells and therefore are capable of efficient competition.

Double-stranded RNAs can be created in cells by hybridisation of sense and antisense transcripts or by a fold-back RNA sequence. To study such complexes, we developed a two-step ligation method for enzymatic attachment of adaptors. The method can be applied to any double-stranded RNA fragment in its duplex form without a need for prior sequence or termini information. Using this method, we mapped a double-stranded region in the highly structured hok mRNA in *Escherichia coli* total RNA, providing evidence for its folded structure in cells.

MicroRNAs regulate expression of their target mRNA through translation repression. Because of their short length, microRNAs are difficult to detect. We aimed to develop a method that could profile the expression of several microRNAs simultaneously by combining ligation of size-coded probes and PCR amplification. The new method is able to distinguish between microRNAs that differ by only one nucleotide and also detects only mature microRNAs. Using this method, we generated unique microRNA signatures for various cell and tissue samples.
LIST OF PUBLICATIONS

This thesis is based on the following original articles, referred to in the text by their Roman numerals.

I. **Omid R Faridani**, Abbas Nikravesh, Deo Prakash Pandey, Kenn Gerdes and Liam Good.
   Competitive inhibition of natural antisense Sok-RNA interactions activates Hok-mediated cell killing in *Escherichia coli*.

II. Abbas Nikravesh, Rikard Dryselius, **Omid R Faridani**, Shan Goh, Majid Sadeghizadeh, Mehrdad Behmanesh, Anita Ganyu, Erik Jan Klok, Rula Zain and Liam Good.
    Antisense PNA Accumulates in *Escherichia coli* and Mediates a Long Post-antibiotic Effect.

III. **Omid R Faridani** and Liam Good.
    Specific ligation to double-stranded RNA for analyses of cellular RNA::RNA interactions.

IV. Jafar Kiani, Ehsan Arefian, Masoud Soleimani, Ali M Banaei, Liam Good and **Omid R Faridani**.
    Simultaneous detection of multiple microRNAs using Size-coded Ligation-mediated PCR.

* Corresponding author
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAR</td>
<td>Adenosine deaminase RNA specific</td>
</tr>
<tr>
<td>Apo-B</td>
<td>Apolipoprotein B</td>
</tr>
<tr>
<td>dsRNA</td>
<td>Double-stranded RNA</td>
</tr>
<tr>
<td>hok</td>
<td>Host killing</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IRE</td>
<td>Iron-response elements</td>
</tr>
<tr>
<td>IRE-BP</td>
<td>Iron-response elements-Binding protein</td>
</tr>
<tr>
<td>LNA</td>
<td>Locked Nucleic Acid</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>peptide-PNA</td>
<td>Peptide-peptide nucleic acid</td>
</tr>
<tr>
<td>piRNA</td>
<td>Piwi-associated RNA</td>
</tr>
<tr>
<td>rasiRNA</td>
<td>Repeat-associated small interfering RNA</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA induced silencing complex</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>scaRNA</td>
<td>Small Cajal body RNA</td>
</tr>
<tr>
<td>SL-PCR</td>
<td>Size-coded Ligation-mediated PCR</td>
</tr>
<tr>
<td>snoRNA</td>
<td>Small nucleolar RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>snRNA</td>
<td>Small nuclear rRNA</td>
</tr>
<tr>
<td>sok</td>
<td>Suppression of killing</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Single-stranded RNA</td>
</tr>
<tr>
<td>STR</td>
<td>Simple tandem repeat</td>
</tr>
<tr>
<td>TCR</td>
<td>T-cell receptor</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

The genome expresses both proteins and RNAs as its functional products. Regulation of gene expression refers to the various cellular controls on the amount, location and timing of the functional gene product appearance. Gene expression regulation occur at several levels; however, studies have traditionally focused on events at the levels of initiation of transcription and also post-translation. More recently, it has become more obvious that a number of events regulate gene expression at the post-transcriptional/translational level.

Translation is modulated through various factors ranging from ribosome entry rate and degradation processes to interactions of proteins and other RNA sequences. Regulation of mRNA translation by other RNA sequences has drawn attention to the new classes of non-coding regulatory RNAs; the RNAs that do not encode proteins but have regulatory functions. In fact, the majority of the genomes of mammals and other complex organisms is transcribed into non-coding RNAs. To date, some of non-coding and regulatory RNA species have been discovered and described including microRNAs and snoRNAs. Here we described major gene expression regulatory processes in three levels with the focus on regulatory RNA families.

1.1 TRANSCRIPTIONAL GENE REGULATION

Gene transcription is resulted from the complex interplay between RNA polymerases, DNA sequences and structures and many other influencing factors. Regulation in this level can be divided into two categories; 1) genetic regulation based on the DNA sequence and 2) epigenetic regulation based on DNA and chromatin chemical and structural modifications.

1.1.1 Genetic-based regulation

Genetic-based regulation of expression involves interactions of protein factors with DNA elements and also, changes in the nucleotide sequences in the genome. The expression of protein factors vary in a tissue-specific manner, causing target genes to be expressed differentially. In addition, DNA sequences can be changed in different cells or individuals by variety of mechanisms, which in turn can alter gene expression.
1.1.1.1 Transcription factors

Transcription factors (TFs) are probably the best appreciated regulators of gene expression. TFs are proteins that bind to specific DNA sequences known as “regulatory elements” using DNA binding domains and regulate function of RNA polymerases using activation domains. TFs can function alone or through interactions with other proteins to enhance (activation) or prevent (repression) gene expression through modulating RNA polymerase start. There are two types of TFs, general and tissue /developmental stage specific. Levels of TFs vary in different biological situations and hence, the genes they control are differentially expressed [1].

TFs can regulate transcription in response to external stimuli through a variety of mechanisms. Some TFs (called nuclear hormone receptors) are induced by binding to a small hydrophobic hormones and morphogens [2]. Some TFs instead are activated by signal transduction in which hydrophilic hormones bind to specific receptors on the cell surface and the signal is passed into the nucleus and to TFs via other molecules. In addition, some TFs (called orphan nuclear receptor) are ligand-independent and their activation pathways are unknown [3].

1.1.1.2 Genetic polymorphism

Genetic polymorphism refers to the occurrence of two or more normal alleles for a single locus. The difference between a polymorphism and a mutation is that a polymorphism occurs commonly and it is associated with a normal phenotype with a frequency of at least 1% of the population.

Some polymorphisms are involved in alteration of gene expression. Microsatellites, also called simple tandem repeats (STRs) are a type of simple sequence length polymorphism comprising tandem copies of, usually, di-, tri- or tetr nucleotide repeat units. Sometimes microsatellites within or in the immediate vicinity of a gene can expand to considerable lengths and affect gene expression, causing disease [4]. Single nucleotide polymorphisms (SNPs) are basically stable point mutations and are major contributors to genetic variation, comprising approximately 80% of all known polymorphisms. Their density in the human genome is estimated to be on average 1 per 1000 base pairs. SNPs can affect the strength of promoters and change the splicing sites [5, 6].
1.1.1.3 Recombination

Recombination is the process by which a strand of DNA is broken and then joined to the end of a different DNA molecule and the process can be homologous or non-homologous. Homologous recombination involves exchanges between chromatids of homolog chromosomes and generally occurs during meiosis. Non-homologous recombination is a pathway that can be used to repair double-strand breaks in DNA and also can lead to translocation and hence incorrect gene regulation [7]. Non-homologous recombination plays a critical role in gene rearrangement. Gene rearrangement makes the organization of immunoglobulin (Ig) and T-cell receptor (TCR) genes quite unique that an individual B or T lymphocyte produces a single type of Ig or TCR [8]. In addition, chromosome segments undergo many enzymatic “cut and paste” reactions as part of intrachromatid recombination events. Another type of recombination occurs when some large-scale deletions and insertions may be generated by pairing of non-allelic interspersed repeats like Alu repeats, followed by breakage and rejoining of chromatid fragments [9].

1.1.1.4 Alternative promoters

Several mammalian genes have two or more alternative promoters, which can result in tissue-specific, developmental stage-specific and sex-specific gene regulation. Also they can lead to differential subcellular localization or functional capacity [10].

1.1.2 Epigenetic-based regulation

Transcription not only depends on genetic factors such as DNA elements and changes in the factors functioning on DNA sequences but it is also influenced by DNA base modifications and the structure of chromatin (chromatin remodeling). This layer of gene regulation is described as epigenetic [11].

1.1.2.1 DNA methylation

The most characterised epigenetic mechanism is DNA methylation. DNA methylation involves the addition of a methyl group to the cytosine, a chemical modification that can be inherited without changing the DNA sequence. In humans, approximately 1% of DNA bases undergo methylation. In adult somatic tissues, DNA methylation typically occurs in CpG dinucleotides that are found at upstream of many genes (CpG islands); non-CpG methylation is prevalent in embryonic stem cells. Between 60-70% of all CpGs are methylated. DNA methylation may block transcription by physically
impeding the binding of transcriptional proteins to DNA. Also, methylated DNA may be bound by protein factors that recruit additional proteins to modify histones, thereby forming compact, inactive chromatin termed silent chromatin [12].

1.1.2.2 Chromatin remodeling

Genome packaging into highly condensed chromatin (heterochromatin or silent chromatin) makes it inaccessible to the factors required for transcription [13]. The level of condensation in chromatin is believed to be regulated by post-translational modification of histones. There are at least eight distinct types of histone modifications: acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, deimination, and proline isomerization [14-16]. Modification can occur at several residues of the H3 and H4 histone tails that protrude from the nucleosome. The core of the H2A and H3 histones can also be modified. Combinations of modifications are believed to create a code termed the "histone code" [17]. Histone methylation is generally linked to transcriptional repression. However, methylation of certain lysine and arginine residues of histones results in transcriptional activation. Histone acetylation is generally associated with euchromatin formation and transcriptional activation [14, 15].

1.1.2.3 The transcription factory and chromosome territory

In the nucleus, it is believed that active polymerases and transcription factor-DNA complexes are organized into clusters, providing a space for transcription; these clusters are named the transcription factories. According to the model, genes in chromatin can diffuse into factories and get transcribed or they can stay outside and form heterochromatin [18]. Concentrations of RNA polymerases II in factories are 1000 fold higher than the average in the nucleus of HeLa cells. It is also thought that transcription factories are specialized for transcription by each of three RNA polymerases and even further for transcription of particular sets of genes [19].

In addition, individual chromosomes in the nucleus are organized into a distinct space termed the chromosome territory. The organization of the chromosome territory is believed to be flexible and creates a nuclear environment that can be important for gene expression regulation during development [20]. According to one model, there is extensive intermixing of chromatin strings from the margins of chromatin territories [21]. Therefore, genes from different territories can co-localize together in transcription factories.
1.2 POST-TRANSCRIPTIONAL/TRANSLATIONAL GENE REGULATION

After transcription, many factors can still modulate gene expression by altering transcript stability or translation rate [22]. There are two major types of regulatory factors in this level: RNA binding proteins and regulatory RNAs [23, 24]. In many events these two factors function together. RNA binding proteins are responsible for regulation of initiation, elongation and termination of translation, RNA splicing, stability, 3’ End Processing, modification and export [24]. Regulatory RNAs are found across all kingdoms (Eukaryotic, Bacteria, Archaea) and mostly appear to play roles in guiding protein factors to RNA sequences [25]. Translation regulation, RNA splicing, editing and modification is performed with the help of regulatory RNAs. Here we listed a number of phenomena that are involved in post-transcriptional regulation.

1.2.1 Alternative splicing and polyadenylation

A large percentage of human genes undergo alternative splicing where different transcript isoforms derived from the same gene by various combinations of exons. For many genes, alternative splicing results in diversity in coding regions that leads to translation of protein isoforms [26]. For example, several isoforms of proteins involved in neurotransmitter storage and release, neurotransmitter receptors and ion channels are produced by alternative RNA splicing. In some genes, alternative splicing occurs in the untranslated regions and this suggests the involvement of additional regulatory processes [27]. Splicing is performed by small nuclear RNAs (snRNAs) accompanied by proteins (snRNPs).

Furthermore, in many genes, two or more polyadenylation signals are found in the 3’ UTR which result in transcripts with different ends. The alternatively polyadenylation can express tissue specificity [28].

1.2.2 RNA editing and modifications

The genetic code can be altered in mRNA by a phenomenon termed RNA editing. The reaction is catalyzed by ADAR, an RNA specific adenosine deaminase on double-stranded RNA (dsRNA) substrates. Editing of adenosine to inosine (A to I) has been reported in glutamate receptor subunit mRNAs in neurons [29]. Editing of cytosine to uracil (C to U) has been found in mRNA of serum protein apolipoprotein B (Apo-B) in intestinal epithelium [30].
Furthermore, almost all classes of RNAs contain nucleotides that are chemically modified post-transcriptionally with great diversity. There are about 100 different base modifications known including 2'-O-methylation, pseudouridine [31]methylations and uridine isomerization [32]. Post-transcriptional modification does not change the genetic code but provides an approach to avoid unwanted folding of RNA. Modification of RNA is catalysed by modifying enzymes and is guided by small nucleolar RNAs (snoRNA), a large subfamily of about 200 small transcripts that reside in the nucleolus. snoRNA mediate base-specific modifications of Pre-rRNA, pre-tRNA, U snRNAs and mRNAs [32]. Some snoRNA species are expressed in a tissue-specific manner in the brain [33]. A novel family of modification guide RNAs has been described in Cajal bodies named small Cajal body RNAs (scaRNAs) [34, 35].

### 1.2.3 Translation and RNA subcellular localization

Gene expression can be controlled in the translation step itself. Ribosome binding efficiency is affected by the 5' cap, the 5' UTR and 3' UTR in mRNA. The 5' and 3' UTR sequences may interact to enhance translation. The 3' UTR has an important role in translational regulation as well as mRNA stability and localization [36]. For example, translation of the ferritin mRNAs is regulated by binding of a specific RNA-Binding protein (IRE-BP) to iron-response elements (IREs) hairpins at the 5' UTR of the mRNAs. When iron concentrations are low, IRE-BP binds to the IREs and inhibits translation initiation and when iron concentrations are high, IRE-BP does not bind to the IREs and translation can be initiated [37].

Naturally occurring antisense RNAs and microRNAs can also regulate translation in the initiation step [38, 39]. We will describe these regulatory RNAs in the next sections.

Riboswitches are another regulatory elements found in bacterial mRNAs. The described examples regulate translation of mRNA by binding to the small molecules directly [40]. Most reported riboswitches interact with metabolic ligands related to the pathway in which their host mRNA produced protein is involved [41, 42]. So far, riboswitch-mediated control of translation has not been reported in higher organisms.

Furthermore, translational regulation can be fine-tuned by specific subcellular localization. Cytoskeletal elements can localize associated mRNAs to specific regions of the cell cytoplasm. This localization is specified by sequences in the 3' UTR of the mRNA. For example, mRNA localization can be seen in mammalian myoblasts when differentiating to myotubes [43].
1.2.4 RNA stability
Changes in mRNA stability lead to changes in its protein expression. The amount of an mRNA is a function of both its rate of synthesis and its rate of degradation. Most mRNAs of higher eukaryotes have half-lives of hours; however, there are many short-lived mRNAs that contain multiple copies of the sequence AUUUA in their 3′ UTR [44]. Destabilization mechanism of these sequences is not yet understood. As an example, stability of transferrin receptor mRNA is decreased by the effect of iron on iron-response elements (IREs) in the mRNA 3′ UTR [37].

1.3 POST-TRANSLATIONAL GENE REGULATION
The function or activity of a protein can be still modulated after translation is complete. For example, covalent or non-covalent binding of small molecules helps to ensure appropriate folding of protein and/or availability of a functional group in a right situation. Many proteins comprise a binding domain which interacts with other proteins as part of the regulatory process. Here a brief summary of post-translational gene regulation is provided.

1.3.1 Protein modifications
Proteins often undergo a variety of chemical modification reactions. This can involve covalent addition of simple chemical groups to the side chains of single amino acids or the addition of different types of carbohydrate or lipid groups [45]. A variety of protein modifications has been reported including phosphorylation (PO4-), methylation (CH3) hydroxylation (OH), acetylation (CH3CO), N-glycosylation (complex carbohydrate), GPI (glycolipid) and myristoylation (C14 fatty acyl group).

Reversible phosphorylation of proteins is mediated by activities of kinases and phosphatases. In signal transduction pathways one kinase makes a second kinase active by phosphorylation and this goes on as a signal from the cell surface to target proteins within the cell [46].

Proteins may also irreversibly be cleaved to generate a smaller mature product. Plasma proteins, polypeptide hormones and growth factors are some examples [47, 48].
1.3.2 Protein subcellular localization

Some proteins are exported to specific intracellular locations such as the nucleus (histones, DNA and RNA polymerases and transcription factors), the mitochondrion and peroxisomes [49]. Localization of proteins requires specific signals in the sequence (signal sequence or leader sequence) or specific types of modifications [50].

1.3.3 Protein-protein and protein-small molecule interactions

Many proteins consist of interacting multiple subunits which can be identical or different polypeptide chains [51-53]. Interactions between the subunit polypeptide chains are important in regulation of protein activity. Many eukaryotic transcription factors function as activators or repressors via protein-protein interactions with components of the transcription machinery [54]. Protein-protein interactions, which can themselves be regulated by the binding of small molecules and phosphorylation, are important in many biological processes [55].

Non-covalent interactions of proteins with specific cofactors (for example, divalent cations, amino acids, nucleotides, ATP and NAD\(^+\)) or ligands can regulate their conformation and function [56, 57]. This type of regulation is common in controlling metabolic pathways by feedback inhibition.

1.3.4 Protein stability

Protein abundance in cells is regulated by differential rates of both protein synthesis and degradation. The half-lives of proteins in cells vary from minutes to several days. Many short-lived proteins function as regulatory molecules, such as transcription factors.

The major pathway of selective protein degradation in eukaryotic cells uses ubiquitin as a marker that targets proteins for rapid proteolysis [58]. Ubiquitin is a 76-amino-acid polypeptide that is highly conserved in all eukaryotes. Polyubiquinated proteins are recognized and degraded by a large multisubunit protease complex, called the proteasome. Many proteins involved in gene expression and cell proliferation are regulated by ubiquitination and proteolysis.

The other major pathway of protein degradation in eukaryotic cells is mediated by lysosomes [59]. Lysosomes contain several proteases and digest extracellular proteins in the endocytosis as well as the cytoplasmic organelles and cytosolic proteins.
1.4 THE RNA WORLD AND REGULATORY RNAs

The central dogma of molecular biology (DNA makes RNA makes Protein) implies that RNA is just a messenger. However, the “RNA world” hypothesis holds that RNA molecules were the only biological compound at the origin of life, capable of storing information and doing enzymatic activities [60]. It is becoming clear that the functional influence of RNA molecules in modern world is still extensive. More and more non-coding RNAs are being discovered to have well-adapted, specialized biological roles. In fact, many non-coding RNAs have roles in which RNA is a more optimal substance than protein. Regulation of gene expression can be achieved simply by base-pairing of a regulatory RNA to an mRNA target post-transcriptionally without a need for the more sophisticated and the more evolved catalytic ability of proteins [61].

In addition, previous studies suggested that in some conditions, post-transcriptional regulation by non-coding RNAs provide the fastest response to external stimuli compare to transcriptional regulation by regulatory proteins and post-translational regulation by protein-protein interactions [62, 63]. These characteristics of non-coding RNAs may describe their prevalence in all kingdoms of life.

Among the newly discovered non-coding RNAs, microRNAs and natural antisense RNAs are the most admired species. Other notable small regulatory RNAs include Piwi-associated RNAs (piRNAs) [64, 65] and repeat-associated small interfering RNAs (rasiRNAs) [66], which were discovered in germline cells recently.

1.4.1 Natural antisense RNAs

Natural antisense RNAs are transcripts that contain complementary sequences to other transcripts. They can be transcribed in cis from opposing DNA strands at the same locus and hence have perfect complementarities, or in trans from separate loci and therefore can have partial complementary sequences. Short transcripts of microRNAs are the best example for trans acting natural antisense transcripts that are hybridised to their target with partial complementarities. We will discuss microRNAs in more details in next section. Other RNAs like snoRNAs and snRNAs are also known to interact with their targets in trans (see previous sections).

Natural antisense RNAs in cis (also called overlapping transcripts) were first detected in viruses [67], then in prokaryotes [68] and finally in eukaryotes [69-71]. They seem to be widespread and have been reported in many genomes. It is estimated
that over 20% of all total transcripts are involved in sense/antisense overlaps in human cells [72].

Antisense RNA in *cis* can be categorized according to their relative orientation; head-to-head (5' to 5'), tail-to-tail (3' to 3') or fully overlapping (Figure 1). Studies indicate that tail-to-tail sense/antisense transcripts are the most abundant form and show more conservation, suggesting that such an orientation is more likely to have a regulatory function [73]. Overlapping transcripts might comprise two protein encoding genes, one protein-encoding and one non-encoding gene, or two non-encoding transcripts.

![Figure 1](image_url)

**Figure 1.** Orientations of sense and antisense genes in the genome. (A) Head-to head orientation. (B) Tail-to-tail orientation seems to be the most abundant and conserved form. (C) Fully overlapping orientation.

The rules by which natural antisense RNAs govern gene regulation are clearer in bacteria. In the most cases, hybridisation of the complementary antisense RNA to an mRNA blocks recognition of the initiation codon and binding of the 30S ribosomal subunit to the Shine-Delgarno sequence, thereby preventing initiation of translation. In eukaryotes there are several models for mechanisms of action (see also Figure 2). (A) Transcriptional interference, two massive RNA polymerase II complexes on opposite strands may crash with and stall one another. Also, concurrent transcription from opposite strands may introduce topological constraint to DNA molecules. The interference results in either transcription repression or transcription in one direction (sense or antisense) only.
(B) Competing for transcription factors, antisense RNAs in cis may regulate expression of sense RNA by competing for transcriptional factors. It may happen for head-to-head orientated overlapping transcripts.

(C) Direct hybridisation of sense/antisense, antisense RNA transcript can hybridise to its target mRNA and directly modulate the expression by forming double-stranded RNA (dsRNA). This can affect splicing, export, stability or translational control of target mRNA. For instance, one report has shown that the antisense masks a splice site on the sense pre-mRNA sequence [74]. This prevents a given splice variant from being formed and favours for variants that do not require splicing of the masked region.

(D) Double-stranded RNA-dependent mechanisms, sense and antisense RNAs may form a double-stranded RNA that can itself be a substrate for RNA editing and RNA interference.

(E) Chromatin remodelling, transcription of non-coding antisense transcripts is involved in monoallelic gene expression through chromatin remodelling, including genomic imprinting, X-inactivation and clonal expression of lymphocyte genes most likely through the recruitment of histone-modifying enzymes.

**Figure 2.** Models for sense/antisense mechanism of action in eukaryotes. (A) Transcriptional interference caused by bidirectional transcription from a same locus. (B) Competing for transcriptional factors. (C) Direct hybridisation of sense/antisense transcripts and RNA masking. (D) Activation of double-stranded RNA-dependent mechanisms such as RNA Editing and RNAi. (E) Chromatin remodelling as the result of antisense RNA expression.
Although gene expression can be regulated indirectly by antisense RNAs, the formation of dsRNA through sense/antisense hybridisation is essential to known antisense regulatory mechanisms. Therefore, factors that influence intermolecular base pairing are likely to be important for antisense regulation. Cellular, genetic and biochemical studies have provided sufficient evidence for base-pairing between antisense and sense RNAs for a variety of small RNAs, including bacterial antisense RNAs (Paper I), microRNAs and guide RNAs such as snoRNA and snRNAs. However, little is known about the formation of dsRNA from long overlapping RNAs.

In general, since RNA is covered by proteins during transcription, protein-RNA complexes may affect RNA::RNA interactions and dsRNA formation. Also, helicases are associated with RNA metabolism and appear to reshape both RNA::RNA and RNA::protein interactions. Many proteins also mediate RNA::RNA interactions. However, in some cases, the cellular response to dsRNA such as editing and siRNA production gives clear evidence for dsRNA formation. Further insight into mechanisms of antisense regulation will require a better understanding of biogenesis and metabolism as well as behavior of the sense and antisense transcripts including dsRNA formation.

To elucidate the extent of dsRNA formation, we designed and developed an enzymatic method that can detect any dsRNA fragments from the cells through ligation of dsRNA adaptors without a prior knowledge about sequence or termini (paper III).

1.4.2 MicroRNAs

MicroRNAs (miRNAs) are short antisense RNAs that reside in RISC (RNA-induced silencing complex) and regulate their target in trans through translational arrest [75-77]. Recently, more than 5000 miRNA genes from viruses, plants and animals have been registered (http://microrna.sanger.ac.uk/) and even more predicted using various algorithms [78-80]. Although miRNAs are first transcribed as hairpins, the functional form of the molecule is a 21-23 mer single-stranded RNA. MicroRNAs are involved in multiple biological processes, including stem cell differentiation, organ development, disease and response to environmental stresses [77].

To study miRNA function it is necessary to determine the expression profile of multiple miRNAs. However, miRNA detection is difficult with standard methods due to the short length of the molecule. In paper IV we sought to develop a method to profile the expression of several miRNAs simultaneously.
2 RESULTS AND DISCUSSIONS

To study regulatory RNAs function and mechanism of action appropriate molecular tools are required. In particular, it seems necessary to apply tools that can detect molecules or molecular interactions in vivo. Therefore, the aims of the thesis were to provide a method to detect:

- RNA::RNA physical interactions between sense/antisense RNAs in vivo (paper I & II)
- Double-stranded RNA structures in conditions close to in vivo (paper III)
- Multiple microRNAs simultaneously from total RNA (paper IV)

2.1 SOK-RNA REPRESSIONS HOK RNA EXPRESSION THROUGH BASE PAIRING

In bacteria, many natural antisense RNAs regulate the expression of target mRNAs through dsRNA-mediated translation inhibition. A paradigm for sense/antisense RNA pairing is the hok/sok toxin–antitoxin plasmid stabilization locus of the R1 plasmid in Escherichia coli [81]. In this system hok (host killing) mRNA that encodes a membrane toxin is neutralized by antisense Sok-RNA (suppression of killing) binding that blocks translation of hok mRNA [82]. Pairing between Sok and hok transcripts is supported by in vitro and phylogenetic studies [83, 84]. Sok-RNA is highly expressed but very unstable. In contrast, the full-length hok transcript is stable and inaccessible to either ribosome initiation or Sok-RNA binding. Slow 3’ end processing of hok mRNA results in mature transcripts that are accessible for translation as well as Sok-RNA binding. In the presence of a hok/sok-carrying plasmid, Sok-RNA binds to the mature form of hok mRNA and represses translation [85, 86]. If plasmid is lost, the Sok-RNA is decayed rapidly and this frees hok mRNA for translation and toxin production kills the cell [87].

To test the model, we designed anti Sok peptide-peptide nucleic acid (peptide-PNA) oligomers that would inhibit hok mRNA::Sok-RNA interactions by out competing Sok-RNA and therefore, release the hok toxin mRNA for translation (Paper I). As expected, anti Sok peptide-PNAs were bactericidal and induced phenotypes consistent with cell killing by Hok protein. The results support the sense/antisense model for hok mRNA repression by Sok-RNA and demonstrate that antisense agents can be used in bacteria to examine RNA::RNA interactions in vivo. Also, the results suggest that antitoxin antisense depletion can provide a novel antimicrobial strategy by activating suicide in bacteria.
2.2 PEPTIDE-PNA ANTISENSE ACCUMULATES IN BACTERIAL CELLS

Anti Sok peptide-PNAs showed the ability to compete with the native interactions between natural antisense Sok-RNA and hok mRNA (paper I). We then compared the binding kinetics of peptide-PNA::RNA and RNA::RNA interactions in vitro. Surprisingly, the in vitro results showed that interaction between peptide-PNA and Sok-RNA is slower compared to interaction between hok mRNA and Sok-RNA in the same concentrations. These results suggest that efficient competition for Sok-RNA in cells requires an excess of peptide-PNA relative to hok mRNA. Such an excess is likely to exist as hok genes are driven by weak promoters, and PNA appears to accumulate in cells and is not removed by drug efflux pumps (Paper II). These results help to explain several observations made within Paper I.

2.3 A METHOD TO AMPLIFY CELLULAR DOUBLE-STRANDED RNA

As shown for the hok/sok system, natural antisense transcripts may regulate gene expression by binding to target RNAs and forming dsRNA. Other regulatory motifs in mRNA sequences can also make hairpins or structures with dsRNA regions.

To study dsRNA content of transcriptome there is a need to develop methods that enable amplification, profiling, mapping and cloning of dsRNA fragments derived from cells. Therefore, we developed a two-step ligation method for enzymatic attachment of various dsRNA adaptors to any dsRNA fragment in its duplex form without a need for prior sequence or termini information (paper III). Next, we labelled dsRNA specifically with fluorophore-conjugated adaptors. Furthermore, we attached adaptors to dsRNA to provide primer sites and then amplified dsRNA using RT-PCR. As an example, we mapped a dsRNA region in the highly structured hok mRNA in Escherichia coli total RNA.
2.4 A METHOD TO PROFILE SEVERAL MICRORNAS SIMULTANEOUSLY

MicroRNAs (miRNAs) are considered to be the most important class of natural antisense RNAs that can act on their target in trans. Detection of miRNA is difficult with standard methods due to the short length of the molecule. To overcome the problem, new approaches have been developed such as various microarray platforms, bead-based array, Northern blotting with LNA probes [88], real-time stem-loop RT-PCR [75] and Padlock probes/rolling circle amplification [89]. Many of these techniques require specialized and expensive equipments/reagent and some involve laborious steps and radioactive materials. Also, techniques based on surface hybridisation face inherent specificity limits that are problematic when distinguishing similar miRNA sequences. Finally discrimination between mature miRNA and precursors remains an important challenge.

Therefore, we sought to develop a reliable method to profile the expression of several miRNAs simultaneously (paper IV). The Size-coded Ligation-mediated PCR (SL-PCR) method takes advantages of liquid hybridisation kinetics of two size-coded DNA probes on a target miRNA, T4 DNA ligation and PCR amplification. The new method is able to distinguish between miRNAs that differ by only one nucleotide and also detect only mature miRNAs, not precursors. The procedure proved to be simple, rapid and requires only inexpensive materials and instruments. We successfully detected miRNA expression signatures in various biological samples.
3 FUTURE PERSPECTIVES

To have a correct view on gene expression regulation it is important to consider all levels of regulation in cells: transcription, post-transcription and post-translation. Recently, post-transcriptional gene regulation has drawn much attention and new molecular players have been revealed including natural antisense RNAs and microRNAs. Despite growing interest, natural antisense RNA-mediated regulation in eukaryotes is poorly characterised. Their mode of action seems more diverse from what has been described in prokaryotes. Possible RNA::RNA interactions between sense and antisense RNA candidates can be investigated by competitive inhibition of a proper synthetic anti-antisense agent (paper I). Furthermore, a stable dsRNA formed by any form of RNA::RNA interactions can be detected and mapped using two-step ligation method (paper III). By adaptor ligation to all dsRNA fragments in the cell it becomes possible to make a cDNA library that can be used in discovery of new RNA::RNA interactions (sense/antisense pairs or regulatory structures). Also, sense/antisense candidates predicted by bioinformatics can be verified by searching in the dsRNA cDNA libraries.

MicroRNAs, the other important players in the post-transcriptional gene regulation, need to be profiled in a variety of biological studies. Profiles of several miRNAs can be determined simultaneously and easily using Size-coded Ligation-mediated PCR (SL-PCR) in various samples (paper IV). The technique has the ability to reveal the expression signature of several miRNAs in a quantitative manner. As an example, such signatures can provide biomarkers in applications including diagnosis of cancer stages or determination of cell differentiation states.

This thesis provides experimental tools to detect interesting classes of regulatory RNAs and further investigate their mode of action.
4 ACKNOWLEDGEMENTS

I would like to thank:

My supervisors:
Liam Good, for supervising me and for teaching me how to supervise.
Fredrik Söderbom, for always being helpful.

Friends and collaborators:
Abbas Nikravesh, for his company and helps, Rikard dryselius for all his helps,
Mohammad Ali Faghihi for all his helps, Shan goh, Hamid Darban, Parviz
Kokhaei, Mohsen Karimi, Siamak Akbari Kamranvar, Masoud Soleimani, Ehsan
Arefian, Jafar Kiani, Alimohammad Banaei, Ramakrishna Rao Sompallae,
Shahryar Kiaii, Nobutaka, Natalia, Agné, Marcella, Camilla, Carsten, Abhi.

Family:
My mother Akram and my late father Mohammad Reza, for all they have offered me,
my sister Mojdeh and my brothers Ehsan and Iman.

And finally ...
I would like to thank my dear wife Firoozeh for all her supports and my little princess
Bahar for bringing sweetness to our family.
5 REFERENCES


