Developmentally Regulated Non-coding RNAs in the Social Amoeba *Dictyostelium discoideum*

Andrea Hinas
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Biology
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2007

Acta Universitatis Agriculturae Sueciae
2007:8

ISSN 1652-6880
ISBN 978-91-576-7307-7

Tryck: SLU Service/Repro, Uppsala 2007

Abstract

Hinas, A., 2007. *Developmentally Regulated Non-coding RNAs in the Social Amoeba* Dictyostelium discoideum*.* Doctoral dissertation.
ISSN 1652-6880, ISBN 978-91-576-7307-7

Non-coding RNAs (ncRNAs) are RNAs that do not encode proteins but instead carry out their functions at the RNA level. During the past decade, ncRNAs have been recognized as major regulators in organisms representing all kingdoms of life. Despite their recent discovery, these ncRNAs are involved in vital processes ranging from cell differentiation and organism development to virulence and stress responses. To exert these important functions, ncRNAs affect *e.g.* chromatin modification, RNA destabilization, inhibition of translation, RNA modification, and protein translocation. Their sizes are as diverse as their functions, varying from ~20 nucleotides up to several thousands of nucleotides.

The social amoeba *Dictyostelium discoideum* has for many years been an appreciated model system for *e.g.* organism development, infection, and chemotaxis, but its ncRNAs have been largely unexplored. In this thesis, experimental as well as computational approaches were employed to investigate the ncRNA profile of this genetically tractable model organism. The identified RNAs include RNA classes with homologs in other organisms, such as small nucleolar RNAs, spliceosomal RNAs, and small interfering RNAs, as well as a large number of entirely novel RNAs. Many of the *D. discoideum* ncRNAs are developmentally regulated, indicating a function in organism development.

Although many ncRNAs are now known to play crucial roles in biology, the functions of most ncRNAs are yet to be discovered. In this quest, easily manipulated model organisms will be invaluable. This thesis further establishes *D. discoideum* as an attractive model system for discerning ncRNA function and regulation.

Keywords: Dictyostelium discoideum, non-coding RNA, small nucleolar RNA, spliceosomal RNA, RNA interference, antisense RNA, development

Author's address: Andrea Hinas, Swedish University of Agricultural Sciences, Department of Molecular Biology, Box 590, S-751 24 Uppsala, Sweden

Till Farmor som alltid frågar hur cellerna mår

Contents

Appendix

Papers I-IV

This thesis is based on the following papers, which will be referred to by their Roman numerals:

I. *Aspegren, A., *Hinas, A., Larsson, P., Larsson, A., & Söderbom, F. 2004. Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Research 32*, 4646-4656.

II. *Hinas, A., *Larsson, P., Avesson, L., Kirsebom, L. A., Virtanen, A., & Söderbom, F. 2006. Identification of the major spliceosomal RNAs in *Dictyostelium discoideum* reveals developmentally regulated U2 variants and polyadenylated snRNAs. *Eukaryotic Cell 5*, 924-934.

III. Larsson, P., Hinas, A., Ardell, D., Kirsebom, L., Virtanen, A., & Söderbom, F. Application of Karlin-Dembo statistics concerning maximal scoring Markov-dependent subsequences to find candidate non-coding RNAs in the *Dictyostelium discoideum* genome (Manuscript).

IV. Hinas A., Reimegård, J, Wagner E.G.H., Nellen W., Ambros, V., & Söderbom F. The small RNA repertoire of *Dictyostelium discoideum*: small antisense RNAs that may be derived from longer transcripts and multiple classes of repeat-associated RNAs. (Manuscript).

*Shared first authorship

Additional publications

Popova, B., Kuhlmann, M., Hinas, A., Söderbom, F., & Nellen, W. 2006. HelF, a putative RNA helicase acts as a nuclear suppressor of RNAi but not antisense mediated gene silencing. *Nucleic Acids Research 34,* 773-784.

Hinas, A. & Söderbom, F. 2007. Treasure hunt in an amoeba: non-coding RNAs in *Dictyostelium discoideum*. *Current Genetics, in press* (Review article).

Problems worthy of attack prove their worth by hitting back.

Piet Hein

1. Introduction

1.1. A historical overview of RNA research

1.1.1. RNA placed in the center of protein synthesis

In the mid 1950s, scientists were trying hard to get the major pieces of the genetic puzzle to fit together, but as always in science without a map and without even knowing whether they possessed all the major pieces. Accumulating evidence indicated that DNA was the carrier of genetic information in most organisms, but at the same time it was dispensable for protein production (Avery, MacLeod & McCarty, 1944; Borsook *et al.*, 1952; Brachet & Chantrenne, 1951; Watson & Crick, 1953). The protein synthesis machineries, ribosomes, at that time called microsomes, had been found to consist of both RNA and proteins (Claude, 1943). Several researchers suggested that RNA represented an intermediate between DNA and protein, although hard evidence was lacking [(Rich & Watson, 1954) and references therein]. Considerable efforts were made to model interactions between amino acids and RNA (or even DNA) molecules to postulate how nucleic acids could serve as templates for protein synthesis, with discouraging results. One important missing piece came with the discovery of small adaptor RNAs, transfer RNAs (tRNAs) which were coupled to amino acids and thus could decode the nucleotide sequence of genes into the amino acid chains of proteins (Hoagland *et al.*, 1958). However, the important template RNA which was supposed to function as a blueprint during protein synthesis had still not been identified. As the ribosome was known to contain RNA components, it was suggested that these ribosomal RNAs themselves constituted the template for protein production, although this raised the obvious question how so few RNAs could give rise to the great variety of proteins observed. This question was solved a couple of years later when two independent groups reported evidence for unstable messenger RNA (mRNA) molecules being the true intermediate between DNA and proteins (Brenner, 1961; Gros *et al.*, 1961). With this, the corner pieces of the genetic puzzle were firmly established.

1.1.2. RNA as a catalyst and the RNA world hypothesis

As scientists found it hard to envision how DNA and proteins, two very different molecules, could have co-evolved they turned to RNA for explanation. RNA could carry genetic information as well as complex secondary and tertiary structures, possibly enabling catalytic function, and was therefore early suggested as a possible ancestor of DNA and proteins (Crick, 1968; Orgel, 1968; Woese, 1967). Initially, this controversial hypothesis had little supporting evidence. The breakthrough for the model did not come until more than a decade later when two

contemporary catalytic RNAs, the self-splicing group I intron and the tRNA processing ribonuclease P RNA, were discovered (Guerrier-Takada *et al.*, 1983; Kruger *et al.*, 1982). This gave new fuel to the theory, which was termed the "RNA world hypothesis" (Gilbert, 1986). The researchers Sidney Altman and Thomas Cech were awarded the Nobel Prize in Chemistry in 1989 for their ground-breaking findings of these catalytic RNAs. Since then, catalytic functions have been proven or suggested for many RNAs, of which the rRNAs are probably the most spectacular as the proteins in this gigantic RNA-protein complex are believed to primarily play a structural role while the RNAs catalyze the synthesis of new proteins (Moore & Steitz, 2006). Today, the hypothesis of an early world where RNA played the role of genetic material as well as catalyst is widely accepted, although whether this represented the very first kind of life or developed from an earlier, unknown, life form is still a matter of debate.

1.2. Still an RNA world – the many functions of non-coding RNAs

Although DNA has taken over the role as genetic material in most organisms, and proteins are responsible for most catalytic reactions, RNA continues to play a central role in many cellular processes. The obvious roles of RNA include those of messenger, adaptor, and catalyst during protein synthesis, but these only represent a fraction of the functions of this versatile nucleic acid. Even at the core of the DNA world – DNA replication – the DNA polymerase still relies on RNA primers to be able to initiate replication (Benkovic, Valentine & Salinas, 2001). Also, in many organisms with linear chromosomes, telomerase RNA is required to prevent chromosome ends from getting shorter with each round of DNA replication (Blackburn, 2006).

Non-coding RNAs (ncRNAs) include, as their name implies, all RNAs that are not translated into proteins but instead carry out their functions as RNAs. Their functions can be directly catalytic, as in the case of ribozymes, or indirect, *e.g.* by base-pairing to a target RNA and thereby conferring specificity to an associated catalytic protein (Huttenhofer & Schattner, 2006). ncRNAs play important roles in all organisms by regulating vital processes such as development, virulence and stress responses (Carthew, 2006; Gottesman, 2005; Huttenhofer, Schattner & Polacek, 2005; Romby, Vandenesch & Wagner, 2006). In humans, dysregulation of certain ncRNAs is associated with severe diseases, *e.g.* cancer and Prader-Willi syndrome (Cavaille *et al.*, 2000; Esquela-Kerscher & Slack, 2006; Kishore & Stamm, 2006). In order to exert their functions, ncRNAs employ a wide spectrum of mechanisms, affecting chromatin structure, transcription, RNA stability, RNA splicing, and protein transport, just to mention a few (Halic & Beckmann, 2005; Matzke & Birchler, 2005; Tycowski *et al.*, 2006; Valencia-Sanchez *et al.*, 2006; Zamore & Haley, 2005). Despite the many diverse roles assigned to various ncRNAs, the function of a large number of ncRNAs is still unknown, and many ncRNAs are probably yet to be discovered. In the following sections, the functions of three ncRNA classes especially relevant for this thesis are described.

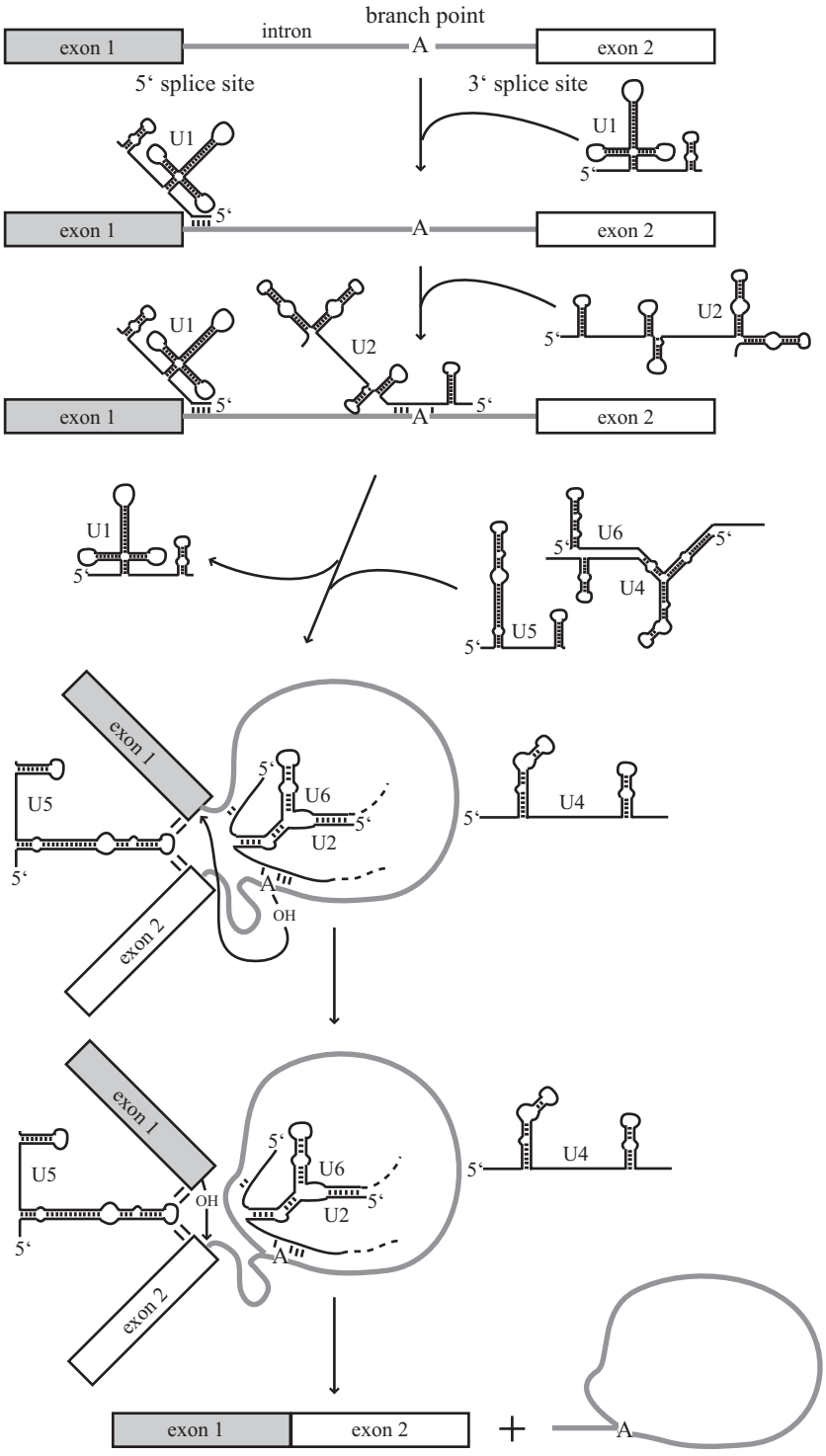
branch point
intron
exon 1
A
exon 2
5‘ splice site
3‘ splice site
U1
5‘
U2
A
OH

Figure 1 (previous page). Pre-mRNA splicing by the major spliceosome. RNA-RNA interactions are shown whereas proteins have been omitted for clarity.

1.2.1. mRNA maturation: spliceosomal RNAs

The majority of eukaryotic mRNAs are dependent on removal of introns and joining of exons to become proper blueprints for translation into proteins. This process, which is called splicing, is carried out by a large ribonucleoprotein (RNP) complex, the spliceosome (Tycowski*, et al.*, 2006; Will & Luhrmann, 2006). The major (or U2-dependent) spliceosome, which is responsible for most splicing, consists of the five spliceosomal RNAs, or small nuclear RNAs (snRNAs), U1, U2, U4, U5, and U6, and more than 200 proteins. During the splicing process, the snRNAs base pair to the pre-mRNA as well as to each other (Figure 1). This enables proper positioning of reactive nucleotides for the two-step splicing reaction and it is likely that the catalysis of the reaction is primarily based on these snRNA-pre-mRNA interactions (Valadkhan, 2005). Many, but not all, eukaryotes also possess a minor (or U12-dependent) spliceosome (Patel & Steitz, 2003; Russell *et al.*, 2006). Apart from U5, which is present in both spliceosomes, the minor spliceosome consists of a different set of RNAs and proteins and recognizes introns with other splicing signals than does the major spliceosome.

1.2.2. RNA modification: small nucleolar RNAs

In organisms from all kingdoms of life, rRNAs, tRNAs and snRNAs carry chemical modifications *e.g.* ribose methylations and pseudouridylations. The positions of the modifications are well conserved and many, in the case of rRNA, are clustered in *e.g.* the peptidyl transfer center, indicating that they may be important for structure and/or function of the modified RNA (Decatur & Fournier, 2003; Maden, 1990). In bacterial rRNAs, which have relatively few of these modified nucleotides, enzymes specific for each position are responsible for the modifications. However, the number of such modifications is considerably higher in eukaryotes and archaea (~100 each of pseudouridylations and 2'-*O*-ribose methylations in vertebrates), probably making it costly to have enzymes specific for each target nucleotide. Instead, one non-specific enzyme for each type of modification achieves specificity by associating with many different RNAs that carry sequences complementary to the target RNA (Decatur & Fournier, 2003; Soderbom, 2006). These guide RNAs, in eukaryotes referred to as small nucleolar RNAs (snoRNAs), are divided into two main classes, box C/D and box H/ACA (Figure 2). Box C/D snoRNAs base pair to their target RNAs via a 10-21 nt long recognition sequence and direct 2'*O*-ribose methylation of the fifth nucleotide in the target RNA, measured from the box D or box D' motif of the snoRNA (Kiss-Laszlo *et al.*, 1996). Box H/ACA snoRNAs, which guide pseudouridylation, also recognize their target through base-pairing, although in this case the target nucleotide is situated between two shorter recognition sequences (Ganot, Bortolin & Kiss, 1997).

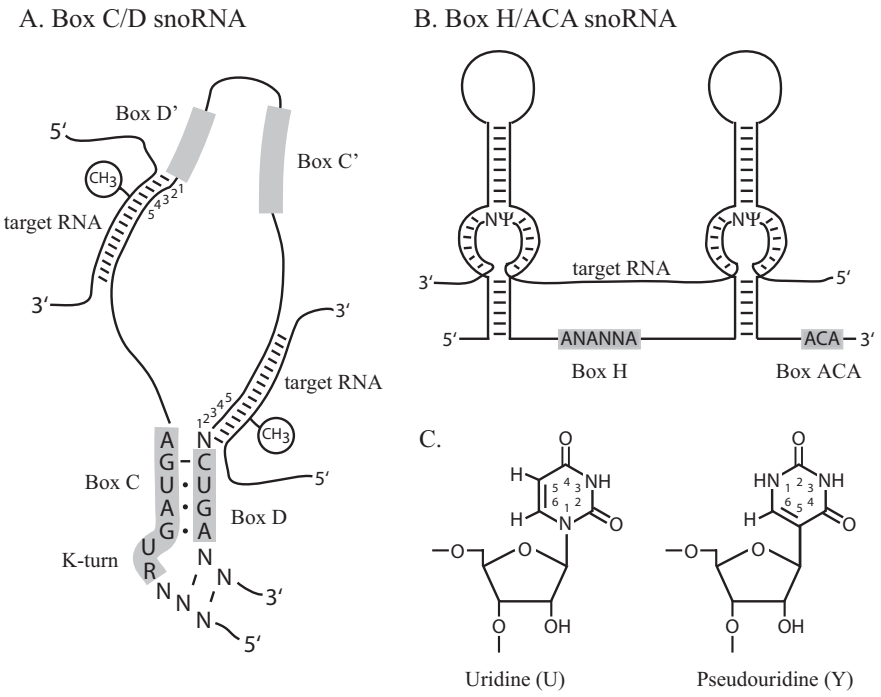

Figure 2. Schematic figure of box C/D snoRNAs **(A)** and box H/ACA snoRNAs **(B)**, which guide 2'-*O*-ribose methylation and pseudouridylation, respectively, of the target RNAs. **(C)** Molecular structure of uridine (left) and pseudouridine (right). Modified, with permission from Springer-Verlag Berlin Heidelberg, from (Soderbom, 2006).

During the last years, a surprising number of eukaryotic snoRNAs has been isolated, several of which are developmentally regulated (Deng *et al.*, 2006; Huttenhofer *et al.*, 2001; Marker *et al.*, 2002; Yuan *et al.*, 2003). rRNAs and snRNAs are classical eukaryotic snoRNA targets but many snoRNAs lack predicted conventional targets and are thus referred to as "orphan" snoRNAs. In mammals, one of these orphan snoRNAs, the brain-specific HBII-52/MBII-52 snoRNA, seems to target an mRNA encoding the serotonin receptor, affecting editing and/or alternative splicing (Cavaille*,* *et al.*, 2000; Kishore & Stamm, 2006). Intriguingly, patients with Prader-Willi syndrome lack this snoRNA, although other factors are also implicated in development of the disease.

1.2.3. Regulation of gene expression: small interfering RNAs and microRNAs

Probably the most significant RNA research discovery and surprise in the past decade is that of numerous small RNAs (20-28 nt) in control of eukaryotic gene expression (Zamore & Haley, 2005). These small RNAs are divided into two main classes, *small interfering RNAs (siRNAs)*, which are derived from longer double-stranded RNAs (dsRNAs), and *microRNAs (miRNAs)*, which originate from

imperfectly base-paired hairpin precursors. A schematic picture of the most common siRNA and miRNA pathways is shown in Figure 3.

The precursors of both miRNAs and siRNAs are processed by the RNase III-type enzyme Dicer into short dsRNAs, which subsequently bind to an Argonaute-Piwi protein (Hammond, 2005). After removal of one of the strands, the single-stranded siRNA/miRNA base pairs to its target RNA, thereby directing the action of the associated protein complex, *e.g.* RNA degradation or inhibition of translation (see below). Besides Dicers, Argonaute-Piwi proteins and RNA-binding proteins, another protein family, RNA dependent RNA polymerases (RdRPs), are also involved in this pathway in *e.g.* nematodes, fungi, and plants, where they are believed to amplify the silencing signal (Wassenegger & Krczal, 2006). Curiously, neither RdRP homologs nor their associated activity have so far been found in mammals or flies.

The first miRNA, *lin-4*, was discovered as a regulator of developmental timing in the nematode *Caenorhabditis elegans* but was for a long time considered a worm-specific oddity (Lee, Feinbaum & Ambros, 1993). This changed abruptly with the discovery of a second *C. elegans* miRNA, *let-7*, which was demonstrated to have homologs in most other animals, including mammals (Pasquinelli *et al.*, 2000; Reinhart *et al.*, 2000). The finding set off large-scale cloning efforts to search for miRNAs in other organisms as well, resulting in identification of a large number of new miRNAs (and other small RNAs) in multicellular organisms; animals and their viruses as well as plants (Aravin & Tuschl, 2005). Although many miRNA targets remain to be investigated, miRNAs have been implicated in many crucial processes, such as apoptosis, insulin secretion, and floral development (Esquela-Kerscher & Slack, 2006; Poy, 2004; Vaucheret, 2006).

Plant miRNAs usually base pair with perfect or near-perfect complementarity to the coding region of their target mRNA, leading to endonucleolytic mRNA cleavage by an associated protein complex (Vaucheret, 2006). This specific cleavage of the mRNA is closely related or identical to the RNA degradation pathway induced by siRNAs (see below). In contrast, animal miRNAs often bind to the 3' untranslated region (UTR) with a higher number of mismatched nucleotides, thereby causing translational arrest instead of mRNA cleavage (Carthew, 2006). Lately, imperfectly base-paired animal miRNAs have been demonstrated to also induce target mRNA degradation, although by an exonucleolytic mechanism rather than the distinct endonucleolytic cleavage caused by perfect base pairing (Valencia-Sanchez*, et al.*, 2006). Based on the low number of base pairs necessary for miRNA-mediated repression of the target gene, it has been estimated that a third of all human protein-coding genes may be regulated by miRNAs (Lewis, Burge & Bartel, 2005).

The discovery of the other main class of small RNAs, siRNAs, was preceded by the finding that exogenously delivered long dsRNAs could induce sequence-dependent repression (termed RNA interference, or RNAi) of a target gene (Fire *et al.*, 1998). For this remarkable finding, Andrew Fire and Craig Mello were awarded the Nobel Prize in Physiology or Medicine in 2006. Shortly after its discovery, the observed RNAi effect was found to be mediated by siRNAs processed from the longer dsRNA precursors (Hamilton & Baulcombe, 1999;

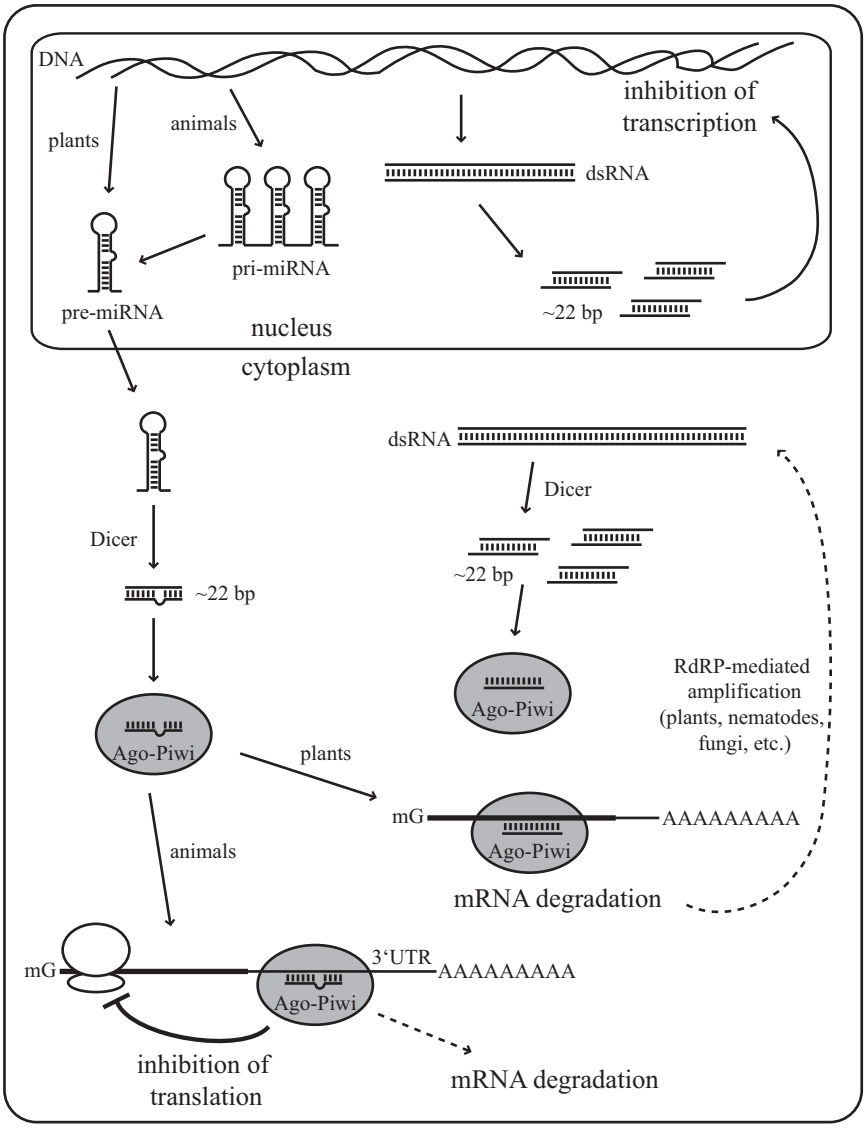

Figure 3. General pathways of small interfering RNAs and microRNAs.

Hammond *et al.*, 2000; Zamore *et al.*, 2000). Since then, siRNAs have been found in most eukaryotes, both unicellular and multicellular, with the notable exception of *Saccharomyces cerevisiae* (Aravin & Tuschl, 2005; Cerutti & Casas-Mollano, 2006). RNAi has been developed as a powerful tool to specifically knock down gene expression in many organisms, and shows great promises for use in treatments of various diseases (Bumcrot *et al.*, 2006).

Besides from being a useful molecular tool, RNAi plays an important cellular role as a defense against viruses and repetitive elements (Aravin & Tuschl, 2005;

Li & Ding, 2005; Voinnet, 2005). The dsRNA precursors of siRNAs can originate from *e.g.* viruses, repetitive elements or be synthesized by RdRPs using a single-stranded RNA (ssRNA) as a template (Wassenegger & Krczal, 2006). The best-studied function of siRNAs is in post-transcriptional gene silencing, where siRNAs guide a protein (often referred as "Slicer") of the Argonaute-Piwi family by base-pairing with perfect complementarity to the target RNA, leading to specific endonucleolytic cleavage and RNA degradation (Baumberger & Baulcombe, 2005; Liu *et al.*, 2004; Meister *et al.*, 2004; Okamura *et al.*, 2004). However, in *e.g.* fungi, plants, and ciliates, siRNAs have also been demonstrated to direct DNA and/or histone methylation, leading to heterochromatin formation and transcriptional silencing or, in the case of ciliates even DNA elimination (Matzke & Birchler, 2005; Mochizuki & Gorovsky, 2004).

Even though siRNAs are generally viewed as a defense against viral RNAs and repetitive elements, numerous siRNAs with perfect complementarity to mRNAs (other than mRNAs originating from repetitive elements) have been isolated from *e.g.* plants, *Tetrahymena thermophila* and *C. elegans* (Ambros *et al.*, 2003; Borsani *et al.*, 2005; Lee, Hammell & Ambros, 2006; Lee & Collins, 2006). Together with other unconventional small RNAs such as the plant trans-acting siRNAs (Talmor-Neiman *et al.*, 2006; Vazquez *et al.*, 2004), these findings suggest that the roles of siRNAs are much more diverse than has formerly been anticipated.

1.3. How to find non-coding RNAs

So, having established the importance of ncRNAs, how can they be identified? Only two decades ago, ncRNA explorers were limited to enzymatic and chemical cleavage of gel-purified RNAs to determine their sequences. Apart from being a tedious task, this method depended on large amounts of starting material, meaning that only the most abundant RNAs, *e.g.* spliceosomal RNAs, could be identified. Today, methods to isolate RNAs of varying size and abundance are available, and the number of identified ncRNAs has reached astonishing levels (Aravin & Tuschl, 2005; Huttenhofer, Schattner & Polacek, 2005; Huttenhofer & Vogel, 2006). Some of these methods, experimental as well as computational, are briefly described below.

1.3.1. Experimental methods

The approach to identify novel RNAs by adding known sequences to the 5' and 3' ends of unknown RNAs followed by reverse transcription polymerase chain reaction (RT-PCR) and cloning caused a sudden, massive increase in the numbers of isolated ncRNA. This technique has been applied to organisms belonging to all three kingdoms of life, and has resulted in the isolation of large numbers of RNAs, ranging in size from ~20 nt to >500 nt (Aravin & Tuschl, 2005; Huttenhofer & Vogel, 2006). The various methods used differ mainly in the way the known 5' and 3' sequences are added, by ligation or by *e.g.* poly(C) tailing.

Recently, the development of cloning-independent sequencing methods such as high-throughput pyrosequencing (also referred to as 454 sequencing) and massive parallel signature sequencing (MPSS) has increased the number of sequences that can be handled in a single experiment to more than a million (Brenner *et al.*, 2000; Margulies *et al.*, 2005). Although these methods are currently limited to short sequences, they have proven immensely useful, especially in the research area of miRNAs and RNAi where the increased sensitivity has enabled identification of very rare small RNAs (Berezikov *et al.*, 2006; Lu *et al.*, 2006; Ruby *et al.*, 2006).

Approaching the problem from a different angle, tiling microarrays can give an indication of which regions of a genome that are expressed (Johnson *et al.*, 2005). These arrays are constructed from partially overlapping oligonucleotides, often covering whole chromosomes or even entire genomes, and are probed with labeled RNA. However, there is always a risk of cross-hybridization and biased RNA labeling methods might also affect the liability of the method. Therefore, expression of RNAs identified by this method needs to be verified *e.g.* by RT-PCR.

1.3.2. Computational methods

The development of powerful and relatively cheap computers, together with new search programs, has turned out to be very useful for prediction of ncRNA genes. Prediction based on sequence similarity to RNA classes known from other organisms is useful when searching for evolutionarily well-conserved RNAs, *e.g.* rRNAs. However, ncRNAs are generally more difficult to identify by primary sequence similarity than are proteins, as secondary and tertiary structure is often equally or even more important for ncRNA function than is the primary sequence (Eddy, 2002). Knowledge of sequence and structure motifs important for function of the ncRNA, and hence likely to be conserved, increases the chances of accurate prediction of ncRNA genes.

Many ncRNAs have been identified by searching for highly conserved regions in the genome sequences of closely related species. This approach, termed comparative genomics, has been successful in finding ncRNAs belonging to known classes as well as entirely novel, conserved RNAs (Rivas & Eddy, 2001; Washietl, Hofacker & Stadler, 2005). To reduce the fraction of false positives, the primary sequence alignment is usually combined with secondary structure predictions. When comparing two putative RNAs with similar predicted secondary structures, it is possible to identify compensatory base changes in stem structures, *i.e.* changes in primary sequences but without loss of base pairing. A conserved structure indicates importance for function, hence increasing the probability that the predicted RNAs are expressed.

The classical comparative genomics approach (described above) to finding ncRNA genes is to search for highly *conserved* regions in aligned sequences. Recently, however, the comparative genomics approach has been further developed to instead search for highly *divergent* regions in sequences from very closely related organisms, such as the human and chimpanzee genomes (Pollard *et al.*, 2006). Most intriguingly, one of these super-diverged regions, HAR1 (for

Human Accelerated Region), corresponds to an ncRNA expressed in developing neurons in human brain.

The methods described above are applicable to any organism, provided that certain requirements are fulfilled such as, in the case of comparative genomics, the presence of a closely related organism with a sequenced genome. Other methods are less general but may be very useful for certain organisms. An example of this is the identification of novel ncRNAs in organisms with A/T rich genomes, taking advantage of the fact that known ncRNAs in these organisms frequently have substantially higher G/C content than the genome average (Klein, Misulovin & Eddy, 2002; Schattner, 2002; Upadhyay *et al.*, 2005).

1.4. *Dictyostelium discoideum* as a model organism

1.4.1. Model organisms in biology

Model organisms are truly invaluable in all areas of biological research. With the advanced molecular tools available today, it is easy to forget that most of what we know about basic cellular processes even in humans is originally based on research conducted on viruses and bacteria. Also today, model organisms play crucial roles in basic as well as applied research. If one, as many researchers do, wishes to draw conclusions from studies of model organisms to understand biological functions in *e.g.* humans, there is always a trade-off between experimental power and biological relevance in the choice of model organism. The bacterial work-horse *Escherichia coli* has many advantages as a model organism, *e.g.* high growth rate and a large number of available genetic tools, however, it is not a eukaryote. The historically most popular eukaryotic model system is probably the budding yeast *S. cerevisiae*, which can be manipulated much in the same way as *E. coli* and has been the source of knowledge of many processes now known to be general for eukaryotes. But yeast is not the organism of choice when it comes to studies of organism development, as it is unicellular. Therefore, multicellular model organisms are highly attractive. In this study, the social amoeba *Dictyostelium discoideum* has been explored as a model organism for ncRNA discovery and function during organism development.

1.4.2. Dictyostelium discoideum

The social amoeba *Dictyostelium discoideum* has for a long time been an appreciated organism for studies of many different cellular mechanisms, ranging from cell motility and organism development to host-pathogen interactions (Kessin, 2001; Williams *et al.*, 2006). *D. discoideum* is also interesting from an evolutionary perspective, since it appears to have diverged from the animal-fungi lineage shortly after plants (Figure 4) (Bapteste *et al.*, 2002; Eichinger *et al.*, 2005). Moreover, it has more proteins in common with animals than with plants or fungi (Eichinger, *et al.*, 2005). In nature, *D. discoideum* lives in the soil where it feeds on bacteria. When faced by starvation, *D. discoideum* cells start to produce and secrete cyclic AMP (cAMP), which is recognized by other *D. discoideum* cells in the surroundings. These cells begin moving towards the cAMP source, as well

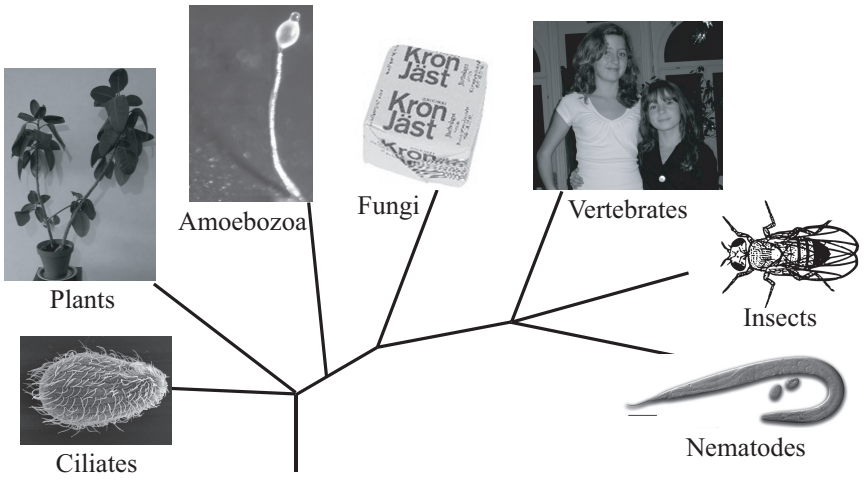

Figure 4. Schematic tree, based on protein sequence comparisons, depicting the approximate evolutionary position of *D. discoideum* (Amoebozoa) relative to other eukaryotes. *Tetrahymena thermophila* (Ciliates) image courtesy of A. Bell; *Caenorhabditis elegans* (Nematodes) image from http://www.wormatlas.org.

as producing their own cAMP, until ~100.000 cells have formed a mound. From this point forward, the cells behave as a multicellular organism and go through a series of well-defined developmental stages (Figure 5). During development, the cells differentiate into two major cell types, prespore and prestalk cells. The development includes a structure called slug (16h post-starvation under laboratory conditions), which is capable of moving towards light and higher temperatures. The developmental phase is completed by the formation of a ball of spore cells supported by a stalk (24h post-starvation). The spores can then survive for several decades awaiting better growth conditions, in contrast to the stalk cells, which vacuolize and die in the development process.

In the laboratory, large numbers of cells can be grown in liquid broth (~8h generation time) and developed in a synchronized manner on nitrocellulose filters, offering an easily-handled model system for cell differentiation and organism development. Furthermore, many molecular tools can be applied to *D. discoideum*, such as gene disruption by homologous recombination, random enzyme-mediated insertion (REMI) mutagenesis and RNAi-mediated knock-down of gene expression (Eichinger & Rivero, 2006). The 34 megabase pair (Mbp) haploid genome has been fully sequenced (Eichinger*, et al.*, 2005) and can be explored through public databases (Chisholm *et al.*, 2006). The *D. discoideum* genome is very A/T rich, with 78% A/T overall and an even higher average in intergenic regions (Eichinger*, et al.*, 2005). The number of protein-coding genes is estimated to 12.500 of which 69% are spliced, containing an average of 1.9 introns (Eichinger*, et al.*, 2005; Szafranski *et al.*, 2005).

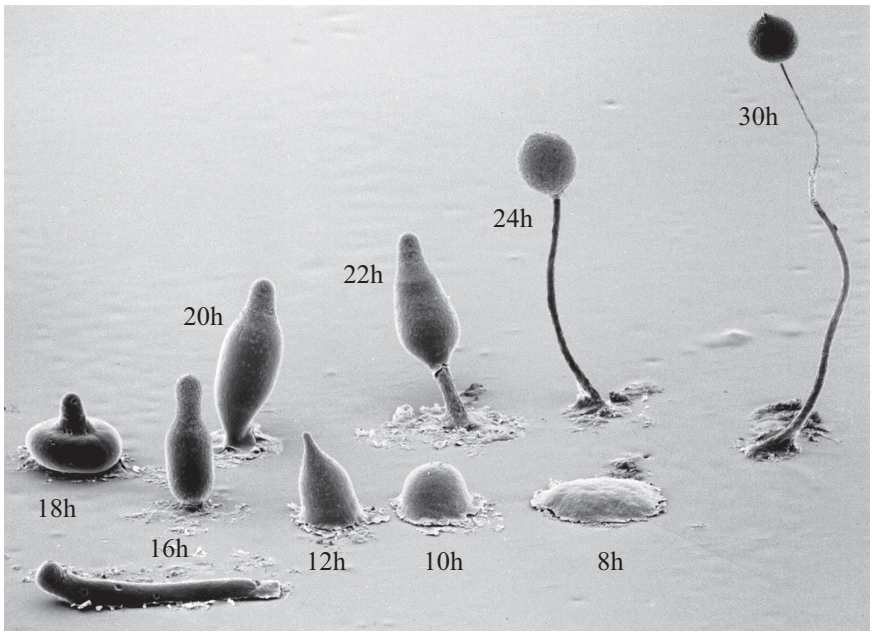

Figure 5. The multicellular developmental phase of *D. discoideum*. Time points designate hours after induction of starvation. Image courtesy of M.J. Grimson & R.L. Blanton, Biological Sciences Electron Microscopy Laboratory, Texas Tech University.

1.5. Non-coding RNAs in *Dictyostelium discoideum*

Prior to this survey, very few ncRNAs except for rRNAs and tRNAs had been identified in *D. discoideum*. Most of these ncRNAs are described briefly below. For a more comprehensive review, see (Hinas & Söderbom, in press).

1.5.1. Small nuclear RNAs

The two nuclear RNAs D1/Dd8 and D2/Dd9 (Kaneda *et al.*, 1983; Takeishi & Kaneda, 1979; Takeishi & Kaneda, 1981; Wise & Weiner, 1980; Wise & Weiner, 1981) were identified by two independent groups following the reports of human small nuclear RNAs immunoprecipitated using serum from systemic lupus erythematosus (SLE) patients (Lerner & Steitz, 1979; Lerner & Steitz, 1981). D2/Dd9 turned out to be a homolog of the U3 snoRNA, which is involved in rRNA maturation (Nazar, 2004; Wise & Weiner, 1980). D1/Dd8, on the other hand, is a 188 nt RNA of unknown function lacking obvious sequence homologs in other organisms (Kaneda*, et al.*, 1983).

1.5.2. An mRNA-like non-coding RNA

Another *D. discoideum* ncRNA to which no function has been assigned is the dutA RNA (for *D. discoideum* untranslatable RNA). The ~1300 nt dutA RNA is

cytoplasmic and, interestingly, expressed in a distinct subset of prestalk cells in developing *D. discoideum* (Maeda *et al.*, 2003; Yoshida, Kumimoto & Okamoto, 1994). The dutA RNA is capped and polyadenylated and thus falls into the enigmatic class of mRNA-like non-coding RNAs also observed in other organisms (Erdmann *et al.*, 2000; Mehler & Mattick, 2006).

1.5.3. Gene regulation by a natural antisense RNA

Microarray analyses and expressed sequence tag (EST) libraries indicate an abundance of antisense transcripts in various organisms and it has been suggested that these antisense transcripts are involved in regulating accumulation of their complementary RNA (Munroe & Zhu, 2006; Werner & Berdal, 2005). However, only few cases of such antisense regulation have been reported (Asa *et al.*, 2001; Borsani*, et al.*, 2005; Heard, 2005; Hildebrandt & Nellen, 1992; Kimelman & Kirschner, 1989; Plath *et al.*, 2002). Interestingly, the *D. discoideum* prespore gene *psvA* has been demonstrated to be regulated by an antisense transcript, which covers a large part of the mRNA (Hildebrandt & Nellen, 1992). The *psvA* mRNA is not detected in growing cells, but accumulates in developing cells, whereas the antisense RNA displays the opposite expression pattern. The expression of the antisense RNA is mainly controlled at the level of transcription. The mRNA transcription, on the other hand, is essentially unchanged during development, and the mRNA is stabilized upon inhibition of transcription, suggesting post-transcriptional regulation mediated by the antisense RNA.

1.5.4. RNA interference

No endogenous RNAi-related small RNAs had been identified in *D. discoideum* prior to the present study. However, as mentioned earlier, RNAi can be used to knock down gene expression in this organism (Martens *et al.*, 2002). The *D. discoideum* genome contains several genes encoding putative RNAi machinery proteins, such as the genes coding for two Dicer-like proteins, *drnA* and *drnB*, and three RNA-dependent RNA polymerases, *rrpA*, *rrpB*, and *rrpC*. Of these, only *rrpA* is required for transgenic RNAi (Martens*, et al.*, 2002). Another *D. discoideum* gene with a demonstrated role in RNAi is *helF*, which encodes a nuclear putative RNA helicase. Intriguingly, deletion of this gene increases the effect of transgenic RNAi (Popova *et al.*, 2006). Furthermore, the *D. discoideum* genome contains five complete and one partial gene predicted to encode Piwi-like proteins (Cerutti & Casas-Mollano, 2006).

2. Present investigation

Since we believed that *D. discoideum* would be an attractive model system to study the functions of ncRNAs during organism development, we were initially facing one major obstacle – the almost complete lack of previously known ncRNAs in this organism. We therefore employed experimental and computational approaches in order to find new *D. discoideum* ncRNAs (see Figure 6). The identified ncRNAs were subsequently subjected to characterization regarding *e.g.* similarities to ncRNAs from other organisms, genomic organization, expression pattern during growth and development, and cellular localization.

2.1. Experimental identification of non-coding RNAs

2.1.1. Cloning of 50-500 nt RNAs (Paper I)

In order to identify *D. discoideum* RNAs, we constructed full-length cDNA libraries representing RNAs in the size range of 50-500 nt. To increase the chances of finding ncRNAs with functions during development, we used RNA from *D. discoideum* cells developed for 16 hours. The RNA was divided into two fractions, 50-150 nt and 150-500 nt, respectively, and a poly(C) tail was added to the 3' ends of the RNA. The C-tailed RNA was subsequently treated with tobacco acid pyrophosphatase (TAP) to convert any 5' triphosphates, characteristic of primary transcripts, to monophosphates, followed by ligation of an RNA oligo to the 5' ends. The RNA, now with known 5' and 3' ends, was converted to cDNA, amplified by RT-PCR, and cloned. To reduce the number of clones derived from rRNA, colony hybridization with radioactively labeled probes against 5S and 5.8S rRNA was performed. Also, 150-500 nt clones with inserts that contained certain restriction sites were cleaved and discarded to lower the number of clones containing 5.8S and 17S rRNA. After these screening steps, sequencing demonstrated that ~14% of the remaining clones (36 unique sequences) represented new *D. discoideum* ncRNAs. These RNAs are further described in sections 2.3 and 2.4.

2.1.2. Cloning of 18-26 nt RNAs (Paper IV)

To get a handle on the role that RNAi plays in the life cycle of *D. discoideum*, we constructed two cDNA libraries representing 18-26 nt RNAs. Both libraries were constructed from the same starting material, *i.e.* RNA pooled from growing as well as developed cells (16 and 24h), but differed in the treatment of the 5' ends of the RNA. In the first library, RNA oligos with known sequences were ligated to the 3' and 5' end of the size-fractionated RNA, followed by RT-PCR amplification and cloning. Since the RNA was not subjected to TAP treatment prior to 5' ligation (compare to method described in section 2.1.1.), only RNAs with a natural 5' monophosphate (*e.g.* Dicer products) would be ligated to the oligo. In the second library, an RNA oligo was ligated to the 3' end of the size-fractionated RNA, followed by reverse transcription. During this process, the reverse transcriptase will add a few untemplated C residues to the 3' end of the cDNA strand.

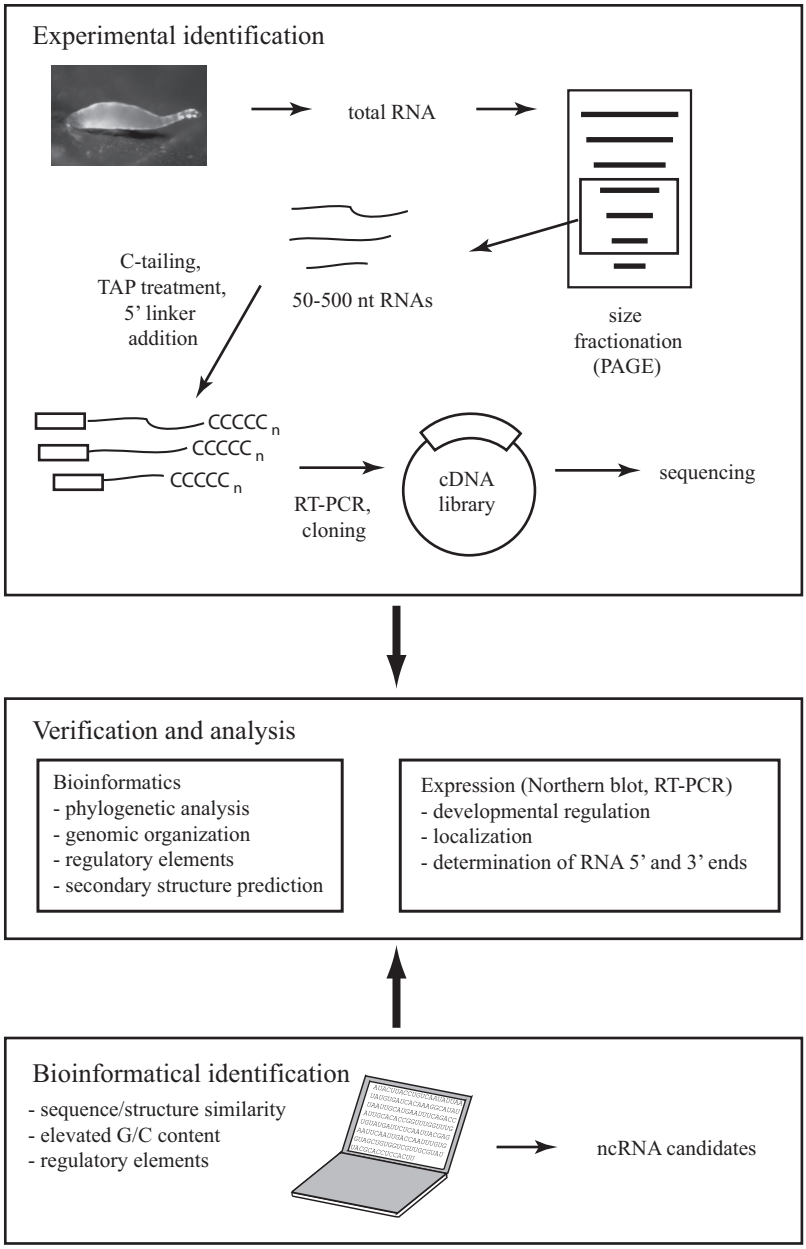

Figure 6. Experimental and computational methods used to identify *D. discoideum* ncRNAs >50 nt in length. Modified, with permission from Springer-Verlag, from (Hinas & Söderbom, in press). For isolation of 18-26 nt RNAs, similar experimental methods were used.

This enzymatic property was utilized in the next step, when a DNA oligo with three terminal G residues was hybridized to the cDNA, followed by PCR and cloning. By this approach, also RNAs which do not carry 5' monophosphates can be cloned. The RNAs isolated in the two cDNA libraries are described in section 2.5.

2.2. Computational identification of non-coding RNAs

2.2.1. Computational identification of the major spliceosomal RNAs by sequence and structure similarities (Paper II)

Although the majority of the *D. discoideum* protein-coding genes contain at least one intron (Eichinger*, et al.*, 2005), no spliceosomal RNAs (small nuclear RNAs, snRNAs) had been identified in this organism prior to our investigations. We isolated the U2 snRNA in the 150-500 nt cDNA library (section 2.1.1) whereas the other snRNAs (U1, U4, U5, and U6) were predicted computationally. In the bioinformatics approach, we took advantage of the conserved sequence and secondary structure motifs as well as RNA-RNA interactions that have been observed for spliceosomal RNAs from other organisms. Two of the *D. discoideum* snRNAs, U5 and U6, could be predicted by overall sequence similarity to their *A. thaliana* counterparts. For U1 and U4, short motifs that in other organisms are known to be important for snRNA-snRNA as well as snRNA-pre-mRNA interactions were used to search the *D. discoideum* genome. Secondary structure predictions and other conserved motifs were then utilized in order to recognize the best candidate sequences, leading to the identification of the complete set of major spliceosomal RNAs from *D. discoideum*. The identified spliceosomal RNAs are described in section 2.3.3.

2.2.2. Computational identification of novel ncRNAs by nucleotide composition bias (Paper III)

During our efforts to find *D. discoideum* ncRNAs, we noticed that most of the isolated RNAs had a much higher G/C content (~45%) than the genome average (~22%). Furthermore, all of our isolated ncRNAs were located between protein-coding genes (Paper I, II). From this point forward, the term "intergenic" will be used to describe regions between known genes (both RNA genes and protein-coding genes). In the intergenic regions, the average G/C content is even lower (~14%) than the genome average. Utilizing this nucleotide composition bias between ncRNA genes and surrounding regions, we searched the *D. discoideum* genome for stretches of nucleotides with a composition that resembled that of a number of *D. discoideum* ncRNAs, *e.g.* tRNAs, and other ncRNAs isolated by us (Paper I, II). To refine the search, we did not only score each nucleotide independently, but took into account also the previous nucleotide. This is based on the assumption that in a functional ncRNA, neighboring nucleotides contribute to structure not only by base pairing, but also by *e.g.* stacking interactions. Moreover, this should increase the probability of predicting the correct orientation of the ncRNA genes, since nucleotides on the opposite strand should not be dependent on each other to the same extent as those on the ncRNA strand.

This search generated 1731 predicted ncRNA genes. Since oligo/tiling microarrays are not yet available for *D. discoideum*, the expression of such a high number of ncRNA candidates could not be analyzed easily. When similar methods for ncRNA discovery have been employed in other organisms with A/T rich genomes, comparative genomics have been used to narrow down the number of candidates (Klein, Misulovin & Eddy, 2002; Schattner, 2002; Upadhyay*, et al.*, 2005). Presently, there are no sequenced genomes for *D. discoideum* relatives close enough for comparative genomics approaches. However, we had previously noticed that *D. discoideum* ncRNA genes frequently occur in multiple, identical or similar, copies (Papers I, II). Therefore, the computationally predicted ncRNA candidates were grouped according to sequence similarity as we believed that groups of similar sequences would be more likely to be expressed than single predicted sequences. In another screening step (separate from the grouping of sequences), we searched for the putative ncRNA promoter [A/T]CCCA[A/C/T]AA previously identified by us (Papers I and II) upstream of the predicted ncRNA genes. Both DNA strands of the regions corresponding to predicted ncRNA genes from the two screens were subjected to Northern blot analysis in order to investigate expression. Identified ncRNAs are described in section 2.4.2.

2.3. Isolated ncRNAs (>50 nt) also found in other organisms

2.3.1. Small nucleolar RNAs (Paper I)

Most small nucleolar RNAs (snoRNAs) can be divided into two classes based on sequence and secondary structure motifs; box C/D and box H/ACA snoRNAs which guide 2'*O*-ribose methylation and pseudouridylation of RNAs, respectively (see section 1.2.2.). Presently, only a few such modifications have been described for the *D. discoideum* rRNAs (McCarroll *et al.*, 1983). In other organisms, the modifications guided by box C/D and box H/ACA snoRNAs are introduced by the methyltransferase Nop1/fibrillarin and pseudouridylase Cbf5, respectively, both of which have highly conserved homologs in *D. discoideum* (Reinders *et al.*, 2006; Söderbom, 2006).

2.3.1.1. Box C/D snoRNAs

In the cDNA library representing *D. discoideum* RNAs sized 50-150 nt, 17 individual sequences were found to represent putative box C/D snoRNAs. All of these were detected by Northern blot analysis and some were, interestingly, expressed at different levels in cells undergoing development (16 and 24 hours after induced starvation) compared to growing cells. Among the developmentally regulated snoRNAs, two were up-regulated and one was down-regulated during development.

In other organisms, the specificity of the box C/D snoRNAs is mediated by a 10-21 nt recognition sequence immediately upstream of the box D and/or D' sequence motifs. Therefore, we searched the *D. discoideum* rRNAs and snRNAs for regions complementary to sequences upstream of the putative D and D' boxes of the

isolated snoRNAs in order to identify putative targets. This approach yielded candidate rRNA targets for two-thirds of the snoRNAs, and 2'-*O*-methylation was experimentally verified for one of the predicted targets by a modified primer extension assay. Several of the other rRNA targets predicted in *D. discoideum* are 2'-*O*-methylated in yeast, plants, and animals.

2.3.1.2. Box H/ACA snoRNA

One putative box H/ACA snoRNA, with the predicted hallmark double hairpin structure and sequence motifs, was isolated in the 50-150 nt cDNA library and expression during growth and throughout development was verified by Northern blot analysis. Box H/ACA snoRNAs targets are difficult to predict without prior knowledge of the pseudouridylation sites of *e.g.* rRNAs and snRNAs. Therefore, and because of the high A/U content of the putative target recognition sequence, no targets for the *D. discoideum* box H/ACA snoRNA could be confidently predicted.

2.3.2. Signal recognition particle RNA (Paper I)

The signal recognition particle (SRP) is involved in localization of membrane proteins to their proper membrane (Halic & Beckmann, 2005). It is present in organisms from all kingdoms of life and consists of an RNA molecule and one to six proteins, depending on the organism. The SRP binds to the signal peptide of nascent proteins and stalls the ribosome until the signal peptide is inserted into the membrane, after which release of SRP allows for translation to proceed.

In the 150-500 nt cDNA library, we were able to identify a *D. discoideum* RNA with sequence and structure characteristics typical for SRP RNA. The RNA was expressed at similar levels in growing and developed cells, and an additional, highly similar, SRP RNA gene was found in the genome.

2.3.3. Spliceosomal RNAs (Papers I, II)

The spliceosomal RNAs (snRNAs) are involved in splicing of eukaryotic pre-mRNAs (see section 1.2.1.). One of the *D. discoideum* snRNAs, U2, was represented in the 150-500 nt cDNA library and the remaining RNA members of the major spliceosome, U1, U4, U5, and U6, were predicted computationally from the genome sequence based on conserved sequence and structure motifs. In total, 18 snRNA genes were predicted and the expression of 17 of these was confirmed by Northern blot and/or RT-PCR analysis. In accordance with other eukaryotes, all of the *D. discoideum* snRNAs except for U6 were shown by immunoprecipitation to carry a trimethylated 5' cap. By subcellular fractionation and subsequent Northern blot analysis, the snRNAs were demonstrated to be localized in the nucleus. Recently, we have confirmed the nuclear localization of the U6 snRNA in growing cells by fluorescent *in situ* hybridization (FISH) (Figure 7, unpublished data).

2.3.3.1. Developmentally regulated U2-like RNAs in the cytoplasm

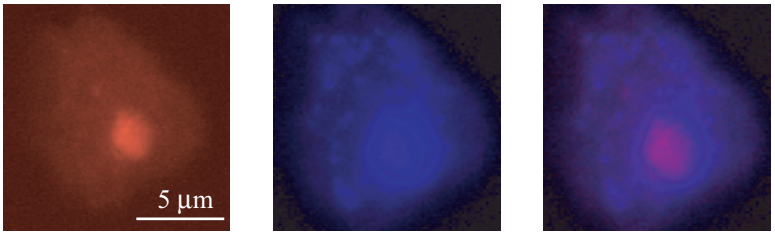

Figure 7. Fluorescent *in situ* hybridization (FISH) of the U6 snRNA from *D. discoideum*. A Cy3-labeled DNA oligonucleotide was used in order to detect U6 snRNA (left). DAPI staining of DNA (middle) and an overlay image (right).

During the course of investigation of the snRNAs, we discovered a subclass of U2-like RNAs (four out of the total seven U2 RNAs) with unexpected, novel characteristics. These RNAs contain the common motifs typical for U2, such as sequences predicted to base pair to the pre-mRNA branch point and U6 snRNA, respectively. However, the U2-like RNAs deviate substantially in other parts. The most striking difference is the presence of an extended 5' end of 30-40 nt which is predicted to base pair, thus forming a stem structure. Remarkably, and in contrast to the conventional U2 snRNAs, the U2-like RNAs are enriched in the cytoplasm and down-regulated significantly during development.

2.3.3.2. Polyadenylated, cytoplasmic spliceosomal RNAs

Another unexpected feature of the *D. discoideum* snRNAs was that a fraction of all snRNAs are polyadenylated. Traditionally, eukaryotic polyadenylation has been connected to stabilization and translation of mRNAs (Wickens, Anderson & Jackson, 1997). During the course of our studies, however, evidence for polyadenylation of eukaryotic ncRNAs has accumulated, especially in *S. cerevisiae* [(Anderson, 2005) and references therein]. In *S. cerevisiae*, polyadenylation of *e.g.* aberrant tRNAs by the poly(A) polymerases Trf4 and Trf5 leads to RNA degradation by the nuclear exosome. This resembles the destabilizing role that polyadenylation has long been known to play in bacteria (Sarkar, 1997).

Polyadenylation of the *D. discoideum* snRNAs was demonstrated by RNase H/ oligo-dT cleavage (removes poly(A) tails) followed by Northern blot analysis, as well as by RT-PCR and subsequent cloning and sequencing. Only polyadenylated U1 RNA accumulated to a level detectable by Northern blot analysis, whereas polyadenylation of the remaining snRNAs could only be detected by RT-PCR. Interestingly, and in contrast to the findings of ncRNA polyadenylation in yeast, the polyadenylated *D. discoideum* U1 RNAs appears to be exclusively cytoplasmic.

2.4. *Dictyostelium discoideum*-specific non-coding RNAs (>50 nt)

Besides the many ncRNA classes conserved in most eukaryotes (some even in archaea and bacteria), species-specific ncRNAs are frequently identified when investigating the RNA repertoire of various organisms (Huttenhofer, *et al.*, 2001; Marker, *et al.*, 2002; Tang *et al.*, 2002; Vogel *et al.*, 2003; Yuan, *et al.*, 2003). In *D. discoideum*, we found several such novel ncRNAs by experimental and computational methods. These ncRNAs are described below.

2.4.1. Cytoplasmic Class I and Class II RNAs (Paper I)

A surprisingly large fraction of the clones in the 50-150 nt cDNA library represented 55-65 nt RNAs without apparent sequence homologs in other organisms. Based on sequence similarity, these RNAs were further divided into two classes, Class I and Class II, with 14 and 2 members, respectively. For Class I, 24 additional genes were predicted from the genome sequence. The Class I RNAs have conserved 5' (16 nt) and 3' (8 nt) sequences which were predicted to form a stem structure. The existence of this stem structure has recently been experimentally verified *in vitro* (Avesson & Söderbom, unpublished). The two Class II RNAs are almost identical to each other and are predicted to form a stem structure similar to that of the Class I RNAs. Furthermore, the Class I and Class II RNAs share an 11 nt sequence motif.

Interestingly, the Class I RNAs are down-regulated slightly at 16 and 24 hours of development compared to growing cells. Both Class I and Class II RNAs are abundant and almost exclusively cytoplasmic, as shown by subcellular fractionation followed by Northern blot analysis.

2.4.2. Novel computationally identified non-coding RNAs (Paper III)

Based on the nucleotide composition bias of *D. discoideum* ncRNAs, in particular their elevated G/C content compared to the surrounding genomic regions, we predicted 1731 ncRNA genes (section 2.2.2.). After further screening steps (grouping of similar sequences and presence of upstream DUSE motif, respectively) expression of candidated ncRNA genes was analyzed.

Of the 89 identified groups, seven were selected for Northern blot analysis. RNA expression was verified for four of these groups, and the sizes estimated from the hybridization signals (160-270 nt) corresponded reasonably well to the predicted sizes. These groups were named drg223, drg229, drg232, and drg233 for *Dictyostelium* RNA group. For drg229, drg232, and drg233, RNA expression was only analyzed in growing cells. However, in the case of drg223, for which RNA levels were analyzed also in developed cells, the detected hybridization signal was weak in growing cells but increased during development (16 and 24 hours).

The 5' and 3' termini of another group, drg232, were determined by RACE analysis. Interestingly, this confirmed expression of at least three different members of the group. The majority of the predicted drg232 members are also preceded by the DUSE.

The second screening approach, presence of an upstream DUSE, resulted in a number of 77 candidate ncRNA genes. Three of these were analyzed by Northern blot, demonstrating expression of all three ncRNAs. These ncRNAs, termed drd189, drd190, and drd191 (*Dictyostelium* RNA with DUSE) were all transcribed from the predicted DNA strand. Interestingly, these RNAs were shown to be cytoplasmic (data not shown). drd189 corresponds to a recently reported selenocysteine tRNA (Shrimali *et al.*, 2005) whereas drd190 and drd191 represent RNAs without apparent sequence homologs in other organisms.

2.5. Isolated ~21 nt RNAs

2.5.1. Multiple classes of repeat-associated small RNAs (Paper IV)

2.5.1.1. Small RNAs derived from the DIRS-1 retrotransposon

More than half of the clones in the 18-26 nt cDNA library represented 21 nt RNAs derived from the abundant DIRS-1 retrotransposon (Kuhlmann *et al.*, 2005). The DIRS-1 element, which is flanked by inverted long terminal repeats (LTRs), encodes a 4.5 kb long mRNA with three open reading frames (ORFs) and a heat-shock induced antisense RNA, transcribed from the opposite DNA strand, covering ~900 nt of the mRNA 3' end (Glockner *et al.*, 2001; Rosen, Sivertsen & Firtel, 1983; Zuker *et al.*, 1984). Clusters rich in partial and complete DIRS-1 elements are found at one end of each chromosome, and have been suggested to constitute centromeres (Eichinger*, et al.*, 2005). Even though DIRS-1 is the most abundant retrotransposon in the *D. discoideum* genome, the number of cDNA library clones representing DIRS-1 small RNAs is much higher than would be anticipated. The cloned DIRS-1 small RNAs are derived from both strands of the retrotransposon, also from regions which are not predicted to form dsRNA by base-pairing between mRNA and antisense RNA or between the LTRs. DIRS-1 mRNA expression has been reported to be up-regulated during development (Rosen, Sivertsen & Firtel, 1983), and we observed the same tendency by Northern blot analysis of a randomly chosen DIRS-1 small RNA.

2.5.1.2. Small RNAs from a partial fragment of the Skipper retrotransposon

A small number of 21 nt RNAs originating from another abundant retrotransposon, Skipper (Leng *et al.*, 1998), were also identified in the cDNA library. These small RNAs (eight sequences) all matched a partial Skipper element on chromosome 2, although some of them could also potentially be derived from complete or partial Skipper elements present at multiple genomic loci. In contrast to the DIRS-1 retrotransposon, no antisense transcript has been reported for Skipper, raising the question of how the small RNAs may be produced. However, there are at least two possibilities for this. First, the partial Skipper fragment containing all eight small RNA loci has the potential to base pair, thus forming a ~300 bp stem from which the small RNAs could be processed. Second, the region to which six of the eight small RNAs mapped could fold into a hairpin precursor similar to the pre-microRNA of multicellular organisms (Bartel, 2004). More experiments will be required to find out which, if any, of these two models is correct.

Similar to the DIRS-1 small RNA (see above), the Skipper small RNAs were up-regulated during development. However, no developmental regulation of the Skipper mRNA has been reported (Kuhlmann*, et al.*, 2005). As previously mentioned, some of the Skipper small RNAs represented in the library had perfect matches in other Skipper sequences in the genome. However, some of the small RNAs are also almost perfectly complementary to non-Skipper mRNAs (data not shown), suggesting that the expression of these mRNAs may be regulated by the Skipper small RNAs.

2.5.1.3. Expression of DIRS-1 and Skipper small RNAs in RNAi knockout strains

The expression of the DIRS-1 and Skipper small RNAs in a number of strains depleted of putative RNAi machinery components, *e.g.* Dicer and RdRP homologs, was analyzed by Northern blot. The levels of the tested DIRS-1 small RNA were not significantly altered in any of the strains, indicating that its biogenesis relies on presently unknown factors or that the proteins involved in the RNAi pathway have redundant functions. The Skipper-derived small RNA, on the other hand, was up-regulated almost 40-fold in a strain where one of the genes encoding an RdRP homolog, *rrpC*, had been knocked out. The same small RNA was also up-regulated, although not to the same extent, in a strain depleted of one of the Dicer-like proteins, encoded by *drnA*. The observed discrepancy in expression patterns between the small RNAs originating from DIRS-1 and Skipper suggests that these small RNAs are produced by different pathways.

2.5.2. Small RNAs complementary to mRNAs may originate from longer antisense RNAs

The 18-26 nt cDNA library also contained a number of small RNAs with antisense complementarity to mRNAs. Three of the corresponding genes were selected for further analysis; *hatA*, *rsmF*, and *DDB0230011*. *hatA* encodes an actin-binding protein which is important for hyperosmotic stress response (Pintsch, Zischka & Schuster, 2002). *rsmF* is predicted to encode a small GTPase, although there was no evidence for its expression previous to this study. *DDB0230011* is predicted to encode a cysteine-rich protein homologous to *e.g. Xenopus laevis* dorsalizing factor kielin (Matsui *et al.*, 2000). In *D. discoideum*, the DDB0230011 mRNA has previously been demonstrated by *in situ* hybridization to be expressed in a subset of prestalk cells during development (Maeda*, et al.*, 2003).

To investigate the possibility that the small RNAs originate from longer antisense RNAs, the three genes were subjected to RT-PCR analysis. This experiment clearly demonstrated the presence of longer antisense RNAs, in itself an exciting result. Moreover, and most unexpectedly, in the case of *hatA,* for which the primers used for RT-PCR were located in two different exons, the antisense RNA was perfectly complementary to the spliced mRNA and not the unspliced pre-mRNA. Sequences corresponding to such "intron-less" hatA antisense RNAs were also found in *D. discoideum* EST libraries (http://www.dictybase.org). Taken together, this indicates that the antisense RNA amplified by RT-PCR is not simply transcribed from the opposite DNA strand but synthesized by an RdRP, using the mature mRNA as a template.

2.5.2.1. Target mRNAs are developmentally regulated (Paper IV)

Northern blot analysis of *D. discoideum* growing cells and cells developed for 16 hours demonstrated that all three target mRNAs are strongly developmentally regulated. The mRNA expression of one of these genes, *hatA*, was shown to be high in growing cells, but low in 16h cells. This is in accordance to previously published *hatA* expression data (Pintsch, Zischka & Schuster, 2002). In contrast to hatA, rsmF and DDB0230011 mRNA could only be detected in 16h cells. The hatA antisense RNA was readily detected and followed the regulation of its mRNA. Intriguingly, the hybridization signal for the hatA antisense RNA suggested a size of ~1100 nt, substantially larger than the corresponding mRNA (~470 nt). No Northern blot signals could be detected for the antisense RNAs of *rsmF* and *DDB0230011*, although these RNAs were easily detected by RT-PCR analysis.

2.5.2.2. RdRP-dependent regulation of the hatA antisense RNA (unpublished)

Since an antisense RNA was only detected by Northern blot analysis for one of the three genes, *hatA*, this gene was selected for further studies. The corresponding protein, hisactophilin, has previously been demonstrated to be post-translationally modified and localize to the cell membrane when cells are subjected to hyperosmosis. Furthermore, cells lacking hisactophilin are hypersensitive to this type of stress (Pintsch, Zischka & Schuster, 2002). By Northern blot analysis, we observed that the hatA mRNA is expressed at a lower level in a strain lacking the *rrpC* gene, which is predicted to encode an RdRP, than in the wildtype strain (Figure 8). The same tendency was observed for a strain where the *drnA* gene, which encodes a Dicer homolog, had been disrupted but not in knockout strains of the *drnB*, *rrpA*, *rrpB*, or *helF* genes (data not shown). We decided to focus on the *rrpC* knockout strain and investigate whether this strain is more sensitive to hyperosmotic stress than the wildtype. This was performed by subjecting cells to hyperosmotic stress (2M sorbitol) (Pintsch, Zischka & Schuster, 2002) followed by plating treated cells on a bacteria lawn to assay cell viability. Unfortunately, the results from this experiment, which was only performed once, were inconclusive. However, Northern blot analysis of RNA isolated from untreated cells and cells subjected to hyperosmotic stress showed that both the hatA mRNA and antisense RNA levels decreased substantially in wildtype cells upon stress. Interestingly, cells lacking the *rrpC* gene failed to down-regulate the antisense RNA, whereas the mRNA followed the pattern of the wildtype cells (Figure 8). It should be noted that the mRNA and antisense RNA were down-regulated already after one hour of incubation in phosphate buffer, before exposure to the hyperosmotic buffer. This suggests that the down-regulation is not specific for the hyperosmotic stress response, but may be more general.

2.5.2.3. Abundant antisense RNAs in *Dictyostelium discoideum* (unpublished)

The discovery of long antisense RNAs in *D. discoideum* (Paper IV) prompted us to investigate the prevalence of antisense RNAs on a genomic level. We therefore searched *D. discoideum* EST databases for clones representing RNAs with antisense complementarity to annotated mRNAs (http://www.dictybase.org).

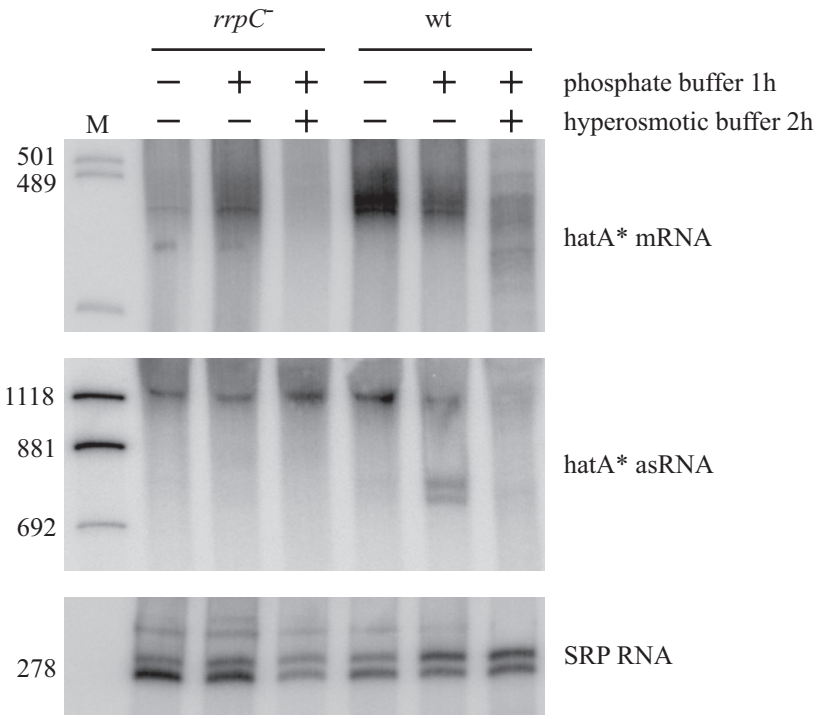

Figure 8. *D. discoideum* cells lacking the *rrpC* gene, which encodes an RdRP homolog, fail to down-regulate the hatA antisense RNA upon stress. Northern blot analysis of hatA mRNA and antisense RNA in wildtype cells (wt) and *rrpC*- cells without prior treatment, after 1h incubation in phosphate buffer, and after 2h exposure to hyperosmotic conditions, respectively. The asterisk indicates that the in vitro transcribed hatA probes most likely cannot distinguish between the highly similar genes *hatA*, *hatB*, and *hatC*.

Intriguingly, a large number of EST clones represented antisense RNAs, corresponding to approximately a third of all annotated protein-coding genes (data not shown). Although these antisense ESTs have not been fully investigated yet, initial analysis indicates that many of the genes for which most antisense EST clones were identified are spatially and/or temporally regulated. Few or no antisense EST clones were found for genes encoding *e.g.* actin and ribosomal proteins, for which no developmental regulation has been reported (data not shown).

2.5.3. Intergenic small RNAs – potential microRNAs or degradation products (Paper IV; unpublished)?

A fraction of the cDNA library sequences mapped to locations in the *D. discoideum* genome for which no expression has been reported previously. Most of these originate from a 27 nt region downstream of the 26S rRNA gene on the extrachromosomal rDNA palindrome. This region has previously been referred to

as non-transcribed (Sucgang *et al.*, 2003). When expression of this region was analyzed by Northern blot, no hybridization signals representing RNAs shorter than ~40 nt were observed (data not shown). It is therefore possible that the cDNA library sequences derived from this region represent degradation products of a longer, previously unknown, RNA.

Other small RNAs represented in the cDNA libraries were derived from chromosomal intergenic regions. Interestingly, the length of these RNAs seems to be biased towards 19-21 nt, indicating that they may be produced by Dicer cleavage. Moreover, preliminary analysis of the regions surrounding some of these RNAs suggests the formation of hairpin structures similar to the microRNA precursors of multicellular organisms (data not shown). However, the intergenic *D. discoideum* RNAs represented in the cDNA libraries could also be degradation products from longer RNAs which have not yet been identified.

2.6. Genomic organization of non-coding RNAs and discovery of a non-coding RNA promoter (Papers I-III)

2.6.1. A putative Dictyostelium discoideum *non-coding RNA promoter*

When analyzing the sequences from our 50-500 nt cDNA library (Paper I), we also searched for conserved upstream sequence elements in the corresponding part of the genome. This identified an 8 nt motif, [A/T]CCCA[A/C/T]AA, termed *Dictyostelium* Upstream Sequence Element (DUSE), which is present upstream of the majority of the ncRNA genes. Intriguingly, this putative promoter sequence is located at a fixed position ~63 nt upstream of the ncRNA transcription start sites.

During the analysis of the computationally predicted spliceosomal RNAs (Paper II), we noted that the same conserved element, with some minor variations, was present at the same position upstream of all of these RNA genes. For the U1, U2, U4, and U5 RNAs, multiple gene copies are expressed. At least one of each corresponding gene was preceded by the motif ACCCATAA, whereas some of the other genes had substitutions in the motif. Differences in the number of clones isolated in RACE experiments suggest that deviation from the consensus motif results in lower expression of the corresponding RNA. In particular, conservation of the three C residues seems to be important for efficient expression. This is especially striking for one of the predicted U1 genes, *U1E*, for which RT-PCR analysis failed to detect any expression. Accordingly, two of the three C residues in the consensus sequence are missing in the *U1E* DUSE (A**AT**CATAA).

Since the majority of our identified RNAs were preceded by the DUSE, we used this sequence motif as a screening criterion in the search for novel ncRNAs by nucleotide composition bias (Paper III). Five predicted ncRNA genes, which were preceded by the DUSE, were tested for expression by Northern blot analysis. Strikingly, expression was verified for all five predicted RNA genes, further corroborating the importance of the sequence motif. 5' RACE analysis of one of these RNAs demonstrated that the fixed position of the DUSE, 63 nt upstream of the first nucleotide of the RNA, was conserved also for this ncRNA gene.

2.6.2. Organization of Dictyostelium discoideum *non-coding RNA genes*

The genomic organization of certain classes of ncRNA genes differs between organisms. For example, snoRNAs are frequently encoded in introns in animals, whereas in *e.g.* plants and budding yeast they are usually transcribed from intergenic regions (Soderbom, 2006). In *D. discoideum*, the genes for all ncRNAs identified to date, including snoRNAs, are located in intergenic regions (Figure 9). We could show that some snoRNAs are co-transcribed (Paper I), also similar to the situation in *e.g. A. thaliana* and *S. cerevisiae* (Brown, Echeverria & Qu, 2003; Chanfreau *et al.*, 1998). Interestingly, in *D. discoideum*, two of the cotranscribed snoRNA gene pairs each consisted of one developmentally up-regulated snoRNA gene and one constitutively expressed snoRNA gene.

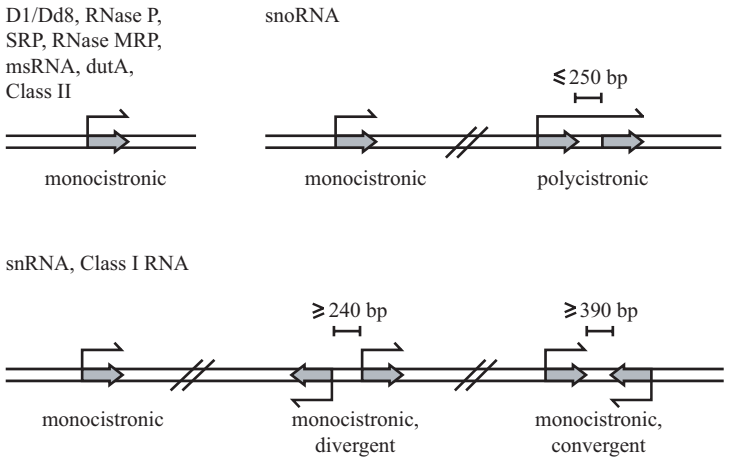

Figure 9. Organization of *D. discoideum* ncRNA genes. Modified, with permission from Springer-Verlag, from (Hinas & Söderbom, in press).

Many of the *D. discoideum* ncRNA genes were present as multiple, identical or similar, copies in the genome. This included *e.g.* Class I RNAs (Paper I), spliceosomal RNAs (Paper II), and novel RNA groups identified by nucleotide composition bias (Paper III). The ncRNA genes are often located closely together, both with RNA genes of the same class and of other classes. Several of the spliceosomal RNA genes are encoded in pairs, either convergent, divergent or in tandem (Paper II). A similar gene organization has previously been observed for *D. discoideum* tRNA genes (Eichinger*, et al.*, 2005).

3. Discussion

The main aim of this study was to identify and characterize *D. discoideum* ncRNAs with possible roles in organism development. In order to identify these RNAs, we employed experimental as well as computational tools, yielding a large number of new ncRNAs of which many are developmentally regulated. In the following sections, I have attempted to put the characteristics of the *D. discoideum* ncRNAs in context of what is known about similar RNAs and processes in other organisms. Furthermore, some advantages and disadvantages of the different identification methods are described.

3.1. Non-coding RNA discovery – a method comparison

Since the first ncRNAs were discovered several decades ago, many tools have been developed in order to identify new such RNAs. When choosing a method for ncRNA identification, there are several factors to consider. For example, what kind of RNAs is the aim; *e.g.* RNAs belonging to a certain class or novel RNAs? Do they contain specific sequence or structure motifs, associate with certain proteins, or display a specific expression pattern? Are high-throughput methods like tiling microarrays and deep sequencing analysis available for the identification or would a method with lower throughput be acceptable? Is the genome of the organism in question fully sequenced? Is there a possibility to use comparative genomics? In the following sections, I discuss the advantages and disadvantages with various experimental and computational approaches available for ncRNA identification, with emphasis on the methods used in this study.

3.1.1. Experimental approaches: cDNA library construction

The method used for construction of a cDNA library is, naturally, reflected in its outcome. For example, in our cDNA libraries representing 50-500 nt RNAs (Paper I), we found no clones representing tRNAs even though this RNA class is known to be very abundant. This is most likely due to the tight secondary structure of tRNAs, and indicates that we may also fail to isolate other highly structured ncRNAs by this method. On the other hand, the lack of tRNAs in the library could also be regarded as an advantage, since the aim in our case was to isolate ncRNAs that had not previously been identified. The major reason for the cloning bias towards less structured RNAs is most likely steric hindrance during 5' end ligation of the RNA oligo, as tRNAs have been isolated from several organisms using a similar method but without the 5' ligation step (Huttenhofer*, et al.*, 2001; Marker*, et al.*, 2002; Yuan*, et al.*, 2003). The difficulties in isolating highly structured RNAs may also explain the low number of box H/ACA snoRNAs (which have a well-defined secondary structure) represented in our cDNA libraries compared to the previously mentioned studies in other organisms. However, the sequences identified in our cDNA libraries have the advantage of representing full-length RNAs, enabling more reliable bioinformatics analyses and saving time otherwise needed to individually determine the 5' terminus for each new RNA. Furthermore,

isolation of full-length RNAs makes our approach suitable also for organisms for which the genome sequence is not yet available, which was the situation for *D. discoideum* when we started our studies.

The characteristics of the 5' ends of the RNAs proved important also when constructing the cDNA libraries representing 18-26 nt RNAs. In this size range, RNA secondary structure is most likely a minor problem. However, cloning efficiency varied between different RNA classes, depending on whether or not the method relied on ligation of an RNA oligonucleotide to the 5' end of the small RNA (Paper IV). The same phenomenon has previously been observed by others (Ambros*, et al.*, 2003; Aravin & Tuschl, 2005; Lim *et al.*, 2003; Pak & Fire, 2007; Ruby*, et al.*, 2006). If the 5' end of the RNA is a monophosphate, a characteristic of *e.g.* Dicer-processed small RNAs, it can be directly ligated to an RNA oligo without any prior treatment. However, RNA degradation products often have a hydroxyl group at the 5' end, which is a poor substrate for 5' ligation (Aravin & Tuschl, 2005). More importantly, certain classes of small RNAs, such as secondary siRNAs in *C. elegans*, seem to have a di- or triphosphate at the 5' end, which renders them less likely substrates for the ligation reaction and hence underrepresented in this kind of cDNA libraries (Pak & Fire, 2007; Ruby*, et al.*, 2006; Sijen *et al.*, 2007). To bypass the problem with 5' ligation, a method can be used where addition of the 5' linker relies on hybridization rather than ligation, see section 2.1.2. (Ambros*, et al.*, 2003).

We used a 5'-ligation-dependent and a 5'-ligation-independent method in parallel to construct our cDNA libraries of *D. discoideum* 18-26 nt RNAs (Paper IV). The most prominent difference between the sequences within the two libraries was the much larger portion of ncRNAs such as rRNAs and tRNAs, at the expense of DIRS-1 small RNAs, represented in the 5'-ligation-independent library. This indicates that the DIRS-1 small RNAs have 5' monophosphates. Although we used two different approaches to clone small RNAs, we cannot exclude that some RNAs might be underrepresented or completely absent in our libraries. In plants, for example, the last nucleotide of many small RNAs has been demonstrated to be 2'-*O*-methylated, significantly decreasing cloning efficiency (Ebhardt *et al.*, 2005; Li *et al.*, 2005a; Yu *et al.*, 2005). It is possible that similar modifications occur in *D. discoideum*, and that these modifications may prevent cloning of certain small RNAs.

3.1.2. Computational approaches

Computational identification of ncRNA genes can be roughly divided into two groups – prediction of ncRNA genes already known from other organisms and prediction of novel ncRNAs. Our study of the *D. discoideum* ncRNAs included both of these approaches.

The spliceosomal snRNAs are highly conserved in eukaryotes and, importantly, very well studied (Tycowski*, et al.*, 2006). Specific sequence and structural motifs, which are important for RNA-RNA as well as RNA-protein interaction, are present in most, if not all, snRNAs. These conserved motifs were used to search the *D. discoideum* genome for snRNA genes (Paper II). By this approach,

we were able to predict 18 snRNA genes and expression was verified for 17 of the predicted genes.

Although sequence and structure similarity can be used to identify many ncRNAs, this approach is of limited use for RNAs that are less well-conserved and for which very limited information concerning important sequence and structure motifs is available. And clearly, this method cannot be used to find entirely novel ncRNAs. The most common computational approach to find novel ncRNAs today is to use comparative genomics, *i.e.* to align sequences from closely related species and search for conserved regions (Eddy, 2002; Washietl, Hofacker & Stadler, 2005). At present, there are no available genome sequences for organisms closely related to *D. discoideum*, although several are underway (Hinas & Soderbom, 2006). Instead, we made use of our previous knowledge of *D. discoideum* ncRNAs to successfully predict ncRNA genes *de novo* (Paper III). Specifically, we had noticed that all of the *D. discoideum* ncRNA genes isolated by us and others are located between protein-coding genes and that the great majority have a significantly elevated G/C content compared to the surrounding regions. We therefore searched the intergenic regions of the *D. discoideum* genome for sequences with a base composition similar to that of the previously isolated ncRNAs. This was done in two different ways. In the first approach, nucleotides were analyzed individually, whereas in the second, we also took into account the context of each nucleotide, *i.e.* the nature of the previous nucleotide. This approach is expected to include contributions to RNA structure and function from *e.g.* stacking interactions, which are also likely to be of importance.

Nucleotide composition bias has previously been employed to search for ncRNAs in other organisms with A/T rich genomes (Klein, Misulovin & Eddy, 2002; Schattner, 2002; Upadhyay*, et al.*, 2005). In these studies, comparative genomics have been used to narrow down the number of ncRNA candidates but since this was not an option in the case of *D. discoideum*, we had to find alternative ways of identifying the regions most likely to be expressed. We selected candidates based on two separate criteria. The first one was to search for the putative *D. discoideum* ncRNA promoter, DUSE, upstream of the predicted ncRNA genes. This method proved extremely successful, with confirmed expression of all tested candidates and, importantly, with correct prediction of from which of the two DNA strands the RNA is transcribed. In the second screen, sequence similarity was used to divide the candidates into groups, assuming that sequence conservation is connected to function and that multiple sequences thus will be more likely to represent true ncRNAs than will single sequences. This resulted in verified expression of four out of seven tested groups, a comparable or even higher success rate than for previous studies in other organisms where comparative genomics were used.

By using several computational methods, and, importantly, through experimental validations and subsequent adjustments of the methods, we have been able to successfully predict a large number of *D. discoideum* ncRNAs. However, some previously known, experimentally isolated *D. discoideum* ncRNAs have escaped our computational screens. These include *e.g.* the A/T rich dutA (Yoshida, Kumimoto & Okamoto, 1994) and Class I RNAs (Paper I),

however, these RNAs would be possible to find by using the DUSE as the sole search criterion (unpublished). In contrast, some box C/D snoRNA genes, which are not identified in the computational searches, do not have the DUSE. Thus, it is highly likely that some *D. discoideum* ncRNAs are yet to be discovered.

3.2. Developmentally regulated non-coding RNAs

Our search for *D. discoideum* ncRNAs resulted in a large number of RNAs with homologs in other organisms, as well as many completely new ones. Many of the isolated ncRNAs were demonstrated to be developmentally regulated, suggesting that they function during development. The medium-sized RNAs (>50 nt) displaying developmental regulation are discussed in the following section. Smaller RNAs are addressed in section 3.3.

3.2.1. Small nucleolar RNAs

Among the 17 isolated box C/D snoRNAs, two were found to be up-regulated and one down-regulated during development (Paper I). The same phenomenon has been observed in *e.g. C. elegans*, *D. melanogaster* and mammals, where many snoRNAs are specifically expressed during certain developmental stages or in distinct tissues (Deng et al.2006; He et al., 2006; Yuan et al., 2003; Hüttenhofer et al 2001). In mammals, for example, a considerable number of snoRNAs are expressed exclusively in brain tissue (Cavaillé et al, 2000; Hüttenhofer et al). One of these snoRNAs, although displaying the characteristics of a box C/D snoRNA, seems to be involved in editing and/or alternative splicing of the serotonin receptor mRNA (Cavaillé et al 2000; Kishore and Stamm 2006). Intriguingly, this snoRNA is not expressed in Prader-Willi syndrome patients. In contrast, the developmentally regulated *D. discoideum* snoRNAs are predicted to guide 2'-*O*-ribose methylation of conventional rRNA targets (Paper I). The two up-regulated snoRNAs are very similar and share a putative rRNA target, corroborating the importance of differential modification of this specific nucleotide during development. Developmental regulation of snoRNAs with conventional rRNA and snRNA targets have also been observed in *D. melanogaster* and *C. elegans*, although its implications have not been investigated (Deng et al., 2006; Yuan et al., 2003). The exact role of nucleotide modifications in rRNA remains enigmatic, but they are likely to alter RNA-RNA as well as RNA-protein interactions (Helm NAR 2006; Decatur and Fournier JBC 2003). The modifications are highly conserved, and are enriched in *e.g.* the ribosome peptidyl transfer center. Nevertheless, only few cases have been described where loss of a certain snoRNA has a detectable effect on the organism (Badis, Fromont-Racine & Jacquier, 2003; Li *et al.*, 2005b). In the hyperthermophilic archaeon *Sulfolobus solfataricus*, the number of 2'-*O*-ribose methylations in rRNA increases with higher growth temperatures, possibly indicating a role in structure stabilization (Noon et al., 1998). In *C. elegans*, expression of many snoRNAs is increased in response to starvation (He et al., 2006; Deng et al., 2006). However, provided that snoRNA expression levels are reflected in the degree of target modification, our observation of a developmentally down-regulated snoRNA is inconsistent with a

view of increased rRNA modification as simply a general starvation response (Paper I).

3.2.2. Spliceosomal-like RNAs

A surprising discovery during the investigation of the *D. discoideum* spliceosomal RNAs was that of a subclass of U2-like RNAs (Paper II). These U2 variants display the main characteristics of conventional U2 snRNAs, however some divergent regions are present where an extended 5' end, predicted to form a stem structure, is the most prominent. They are considerably down-regulated during development, and, surprisingly, enriched in the cytoplasm in contrast to the conventional, nuclear, spliceosomal RNAs. To our knowledge, this is the first report of cytoplasmic localization of a spliceosomal RNA. Presently, it is not known what causes the developmental regulation or the cytoplasmic enrichment of the U2-like RNAs. It is tempting to speculate that the predicted 5' stem structure is involved in either, or both, of these processes. However, the developmental regulation could also be exerted on the transcriptional level, since the putative promoter (DUSE) of the U2-like RNA genes differs from that of the conventional, nuclear, *D. discoideum* U2 snRNAs.

Shortly after the publication of our results (Paper II), the interesting observation of a similar variant U2 gene was identified in the genome of the amoeba *Acantamoeba castellanii*, although so far its expression has not been verified (Russell*, et al.*, 2006). Even though no homologs in more distantly related eukaryotes have been identified so far, there are several reports of developmentally regulated spliceosomal RNA variants in *e.g.* plants, insects and mouse (Chen *et al.*, 2005; Hanley & Schuler, 1991; Lund, Kahan & Dahlberg, 1985; Sierra-Montes *et al.*, 2005). In most of these reports, the cellular localization of the snRNAs has not been investigated and it is therefore not known whether the cytoplasmic enrichment that we observed for the U2-like RNAs is unique for *D. discoideum*. In *D. melanogaster*, developmentally regulated snRNAs were found to co-sediment with conventional snRNAs and proteins, suggesting a function in splicing (Chen*, et al.*, 2005). The role of developmentally regulated snRNAs remains elusive, but it is possible that they are somehow involved in regulating pre-mRNA splicing *e.g.* by influencing alternative splicing.

3.2.3. Novel non-coding RNAs

An unexpectedly large fraction of the 50-150 nt cDNA library represented novel RNAs without apparent homologs in other organisms (Paper I). These RNAs were divided into two classes, Class I and Class II, based on sequence and predicted secondary structure. Both classes are cytoplasmic and the Class I RNAs are down-regulated during development. Another class of developmentally regulated novel ncRNAs was identified in the computational search for RNAs based on nucleotide composition (Paper III). In this study, the expression pattern during development was only analyzed for one of the predicted ncRNA classes, which was demonstrated to be up-regulated during development.

What appears to be species-specific ncRNAs are frequently identified in *e.g.* cDNA libraries from different organisms (Deng*, et al.*, 2006; Huttenhofer*, et al.*, 2001; Marker*, et al.*, 2002; Tang*, et al.*, 2002; Vogel*, et al.*, 2003; Yuan*, et al.*, 2003). Analogous to the *D. discoideum* novel ncRNAs, the expression of these ncRNAs is often regulated, both temporally and spatially, indicating that they may play important roles during cell differentiation and development. However, so far the functions of these ncRNAs remain to be elucidated. In many cases, the species-specific ncRNA genes are present as single copies in the genomes. In contrast, many of the *D. discoideum* novel ncRNAs form distinct classes. Also in *C. elegans*, a class of novel ncRNAs termed stem-bulge RNAs (sbRNAs) was recently identified (Deng et al. 2006). Interestingly, the sizes and predicted secondary structures of the *C. elegans* sbRNAs resemble those of the *D. discoideum* Class I RNAs, although the functional importance of this similarity is not yet known. It is important to note, however, that even though a specific ncRNA does not appear to have any *sequence* homologs in other organisms, this does not necessarily imply a lack of *structural* and/or *functional* homologs. It is possible that several of the identified, apparently unique, ncRNAs from different organisms will eventually turn out to have similar functions.

3.3. RNAi and natural antisense RNAs

3.3.1. Repeat-associated small RNAs

Analysis of the cDNA libraries representing 18-26 nt RNAs from *D. discoideum* showed that a large number of 21 nt sequences match the abundant retrotransposon DIRS-1 (Kuhlmann*, et al.*, 2005). Repeat-associated small interfering RNAs (rasiRNAs) are commonly represented in similar cDNA libraries from other organisms and are implicated in silencing of the corresponding repeats (Aravin & Tuschl, 2005; Aravin, 2003; Reinhart & Bartel, 2002; Watanabe *et al.*, 2006). In many organisms, *e.g. D. melanogaster*, *A. thaliana*, and *S. pombe*, the repeat-associated small RNAs are longer than other siRNAs and miRNAs; 23-28 nt (Aravin, 2003; Hamilton *et al.*, 2002; Reinhart & Bartel, 2002). However, the lengths of the mouse rasiRNAs resemble that of *e.g.* miRNAs; ~22 nt (Watanabe*, et al.*, 2006).

In *D. discoideum*, DIRS-1-rich clusters have been suggested to constitute centromeres (Eichinger*, et al.*, 2005). The DIRS-1 small RNAs might thus be analogous to the centromeric siRNAs found in *e.g. S. pombe* and *A. thaliana* (May *et al.*, 2005; Reinhart & Bartel, 2002; Volpe *et al.*, 2002), which seem to be involved in silencing of the centromeric repeats.

Intriguingly, the DIRS-1 and Skipper small RNAs exhibited different expression patterns in a number of strains where RNAi-related protein genes had been deleted. The accumulation of DIRS-1 small RNAs was not significantly altered in any of the mutant strains, whereas the Skipper small RNAs were considerably up-regulated in strains lacking the Dicer homolog DrnA and the RdRP homolog RrpC. This suggests that the small RNAs originating from the two retrotransposons are produced through at least partly different pathways.

There is no clear correlation between the levels of the Skipper and DIRS-1 retrotransposon small RNAs and mRNAs in the various mutant strains. The Skipper mRNA was previously found to be up-regulated in all RNAi knockout strains tested, except for a strain where the *helF* gene, which encodes a putative RNA helicase, had been deleted (Kuhlmann*, et al.*, 2005). The expression of the DIRS-1 mRNA, on the other hand, was unaffected in all strains except for *rrpC*$^{-}$.

Furthermore, the Skipper mRNA was up-regulated in a strain where the gene encoding the Dnmt2-type DNA methylase DnmA had been disrupted, whereas the DIRS-1 mRNA was unaffected in this knockout strain (Kuhlmann*, et al.*, 2005). The DIRS-1 retrotransposon has been demonstrated to undergo DNA methylation by DnmA, but so far there is no evidence linking the DIRS-1 siRNAs to this DNA methylation (Kuhlmann et al. 2005; Katoh et al. 2006).

3.3.2. Natural antisense RNAs

Besides the repeat-associated small RNAs, mining of the 18-26 nt cDNA libraries also identified a number of small RNAs with antisense complementarity to annotated protein-coding genes (Paper IV). Intriguingly, we could demonstrate expression of longer antisense RNAs from the three loci that were analyzed. All of the corresponding mRNAs are strongly developmentally regulated. For two of the genes, *rsmF* and *DDB0230011*, expression of the antisense RNA could only be detected by RT-PCR. For the third gene, *hatA*, which encodes hisactophilin, a protein involved in hyperosmotic stress response (Pintsch, Zischka & Schuster, 2002), the antisense RNA could also be detected by Northern blot analysis. Interestingly, the antisense RNA displayed the same down-regulation during development as the mRNA. The regulation of both the *hatA* mRNA and antisense RNA seems to be primarily transcriptional.

Gene regulation mediated by cis-encoded antisense RNA is a well-studied phenomenon in bacteria (Gottesman, 2005). In eukaryotes, analysis of data from genome-wide approaches such as EST libraries and microarrays indicates that cis-antisense transcription is common (Munroe & Zhu, 2006; Werner & Berdal, 2005). The overlapping transcripts can consist either of one mRNA and one ncRNA, two mRNAs or two ncRNAs. So far, relatively few cases of eukaryotic gene regulation involving cis-transcribed antisense RNAs have been described (Hildebrandt & Nellen, 1992; Munroe & Zhu, 2006; Werner & Berdal, 2005). These examples, however, point to multiple regulatory mechanisms. For example, expression of the *D. discoideum* prespore *psvA* mRNA seems to be post-transcriptionally regulated by an antisense RNA (Hildebrandt & Nellen, 1992) and two examples of antisense regulation in *A. thaliana* involve RNAi-mediated RNA degradation (Borsani*, et al.*, 2005; Katiyar-Agarwal *et al.*, 2006). However, in *S. cerevisiae*, which lacks the RNAi machinery, transcription of an antisense RNA was recently demonstrated to inhibit the transcription of an mRNA from the opposite DNA strand, a phenomenon known as transcriptional interference (Hongay *et al.*, 2006).

At this point, we do not know whether the *D. discoideum* small antisense RNAs identified in the cDNA library are derived from the longer antisense RNAs, *e.g.*

generated by specific Dicer cleavage, or if they represent general degradation products. It is possible that the sense-antisense regulation occurs by a mechanism distinct from RNAi, as has been observed in *S. cerevisiae* (Hongay*, et al.*, 2006). However, the fact that an RdRP seems to be required for rapid down-regulation of the hatA antisense RNA (section 2.5.2.2., unpublished) suggests that the RNAi machinery is indeed involved in the antisense RNA-mediated regulation of, at least, *hatA*.

3.4. Biogenesis and stability of non-coding RNAs

3.4.1. A versatile non-coding RNA promoter

During the course of our investigation of the *D. discoideum* non-coding RNAs, we discovered a putative promoter element, DUSE (Papers I-III). This 8 nt motif, [A/T]CCC[A/T/C]AA, is present ~63 nt upstream of the transcription start sites (TSS) of the majority of our isolated of ncRNAs, as well as of previously known *D. discoideum* ncRNAs (Hinas & Söderbom, in press). It resembles the upstream sequence element (USE) of spliceosomal RNA genes in *e.g. A. thaliana* (Waibel & Filipowicz, 1990; Vankan, Edoh & Filipowicz, 1988), *D. melanogaster* (Mount *et al.*, 2007), and *T. thermophila* (Orum, Nielsen & Engberg, 1992), as well as plant SRP RNA (Heard *et al.*, 1995; Yukawa *et al.*, 2005) and RNase MRP RNA genes (Kiss, Marshallsay & Filipowicz, 1992).

In other eukaryotes, the U1-U5 snRNAs are transcribed by RNA polymerase II and acquire a trimethylated 5' cap ($m^{2,2,7}G$) during biogenesis, whereas the U6 snRNA is transcribed by RNA polymerase III and has a γ-methylated triphosphate at the 5' end (Will & Luhrmann, 2001). When the same USE is present upstream of RNA polymerase II- and RNA polymerase III-transcribed snRNA genes, additional promoter sequences, e.g. TATA boxes, and/or differential positioning of the USE determine which RNA polymerase will transcribe the gene (Schramm & Hernandez, 2002). In *D. discoideum*, the distance between the DUSE and the transcription start site is the same for all snRNA genes, including U6 (Paper II). Nor have we been able to detect any TATA boxes with confidence due to the high A/T content of the genome. However, we could demonstrate that the *D. discoideum* U1-U5 snRNAs, but not U6 snRNA, carry the ($m^{2,2,7}G$) cap, indicating that the RNA polymerases dedicated to transcription of the different snRNA gene in *D. discoideum* are the same as in other eukaryotes.

At present, we have not carried out any systematic mutational analysis of the DUSE to investigate the contribution of the individual nucleotides to expression of ncRNA. In spite of this, a wealth of information is already available, especially for the snRNA genes where many variations of the DUSE are represented (Paper II). Notably, the only predicted snRNA gene for which no expression could be detected has a degenerate DUSE with only one C residue instead of the conserved three. It is also possible that the down-regulation during development observed for the U2-like RNAs is due to their slightly deviating DUSE sequence. Further experiments are required to fully investigate the characteristics of this putative promoter.

3.4.2. Polyadenylation of non-coding RNAs

In bacteria, polyadenylation of both mRNA and ncRNA leads to RNA degradation (Sarkar, 1997). In contrast, polyadenylation in eukaryotes is generally connected to mRNA stabilization and increased translation efficiency (Wickens, Anderson & Jackson, 1997). For a long time, it was believed that these contradictory effects of polyadenylation on RNA stability represented a fundamental mechanistic difference between bacteria and eukaryotes. However, in recent years, sporadic studies have reported the presence of polyadenylated ncRNAs, *e.g.* rRNAs, snoRNAs, and spliceosomal RNAs, in *S. cerevisiae*, but only in strains lacking certain RNA processing enzymes (Abou Elela & Ares, 1998; Allmang *et al.*, 1999; van Hoof, Lennertz & Parker, 2000).

During our analysis of the *D. discoideum* spliceosomal RNAs (Paper II), we noted that a small fraction of these RNAs is polyadenylated. To our knowledge, this was the first report of polyadenylated snRNAs in a wildtype organism. Interestingly, our findings of polyadenylated snRNAs in *D. discoideum* coincided with a number of reports, mainly from *S. cerevisiae*, of ncRNA polyadenylation leading to RNA degradation by the nuclear exosome [(Anderson, 2005) and references therein]. For example, misfolded tRNAs were demonstrated to be specifically polyadenylated and degraded, pointing to a function in RNA quality control (Kadaba et al. 2004). The enzyme responsible for the polyadenylation, Trf4/Trf5, is different from the conventional poly(A) polymerase involved in mRNA maturation. Whether the polyadenylation of snRNAs in *D. discoideum* is associated with RNA degradation or is part of the normal snRNA maturation pathway remains to be investigated.

In *S. cerevisiae* and *S. pombe*, the Trf4/Trf5 poly(A) polymerases, as well as the exosome that degrades the polyadenylated ncRNAs, are nuclear (Allmang*, et al.*, 1999; Huh *et al.*, 2003; Win *et al.*, 2006). However, in *D. discoideum* we could only detect the polyadenylated snRNAs in the cytoplasm. This may indicate a mechanistic difference in ncRNA polyadenylation between fungi and *D. discoideum*, although the possibility that the snRNAs are polyadenylated in the nucleus prior to their export to the cytoplasm cannot be excluded at this point.

Lately, exonucleolytic RNA degradation has also been implicated in RNAi/miRNA pathways, primarily by degrading the 5' and 3' products resulting from si-/miRNA-mediated mRNA cleavage (Valencia-Sanchez*, et al.*, 2006). Most intriguingly, a very recent report demonstrated polyadenylation and subsequent exosome degradation of such RNAi cleavage products in the green alga *Chlamydomonas reinhardtii*, ultimately connecting the different RNA degradation pathways (Ibrahim et al. 2006). Taken together, these results establish a new role of eukaryotic polyadenylation in RNA degradation, which has much more in common with its bacterial counterpart than previously anticipated.

4. Conclusions and future perspectives

My thesis project has largely focused on the isolation and basic characterization of *D. discoideum* ncRNAs, with the long-term goal to identify ncRNAs with roles in cell differentiation and development. Both experimental and computational methods have been successfully employed in this search, which has resulted in a large number of new ncRNAs of sizes ranging from ~21 nt repeat-associated small RNAs to >1000 nt antisense RNAs. Classes of ncRNAs already known from other organisms as well as novel ncRNAs have been identified, many of which are developmentally regulated. Now remains the tedious, but extremely exciting task of finding out the functions of the novel ncRNAs, as well as elucidating the specific roles of the developmentally regulated ncRNAs belonging to already known classes. This will be approached by using some of the numerous molecular tools applicable to *D. discoideum*. Specific gene knockouts, as well as RNA over-expression, can give clues of which ncRNAs, if any, are required for proper organism development. Fluorescent *in situ* hybridization will be used to determine ncRNA localization/expression patterns during the unicellular as well as multicellular stages of the life cycle. Furthermore, isolation of ncRNA-interacting proteins may give invaluable hints on function, and might also reveal functional homologs of the investigated ncRNA in other organisms which cannot be identified solely by sequence comparisons.

Some points of particular interest:
i) We identified two Box C/D snoRNAs that are up-regulated during development and predicted to guide 2'-*O*-ribose methylation of the same rRNA nucleotide. What is the impact of this modification? Could it, for example, decrease the rate of translation as a response to starvation?
ii) Many different variants of the novel Class I RNAs are expressed. What is the function of these abundant RNAs? Is the conserved stem structure a binding site for proteins? If so, does the variable loop region interact with other RNAs by base-pairing, thereby conferring specificity to the protein?
iii) What is the role of the observed spliceosomal RNA poly(A) tails? Do they function in quality control, as in fungi, or do they represent intermediates during normal snRNA biogenesis? The *D. discoideum* polyadenylated snRNAs are cytoplasmic, in contrast to the polyadenylated ncRNAs in fungi. Does this reflect a general difference between amoeba and fungi? Is the enzyme responsible for the polyadenylation nuclear or cytoplasmic?
iv) So far, microRNAs have only been found in multicellular animals and plants. However, *D. discoideum* has both unicellular and multicellular life cycles and the burning question; Are there microRNAs in *D. discoideum*?; remains unanswered.
v) Several long RNAs with antisense complementarity to protein-coding mRNAs were discovered and antisense transcription seems to be widespread in *D. discoideum*, as in other organisms. Do these antisense RNAs regulate the expression of the corresponding mRNAs? If so, is the mechanism(s) mediated by the RNAi machinery or by other factors?

Given the enormous potential of *D. discoideum* as a model organism, I am convinced that several of these questions will be answered in the near future.

5. References

Abou Elela, S. & Ares, M., Jr. 1998. Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *Embo J 17*, 3738-46.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. & Tollervey, D. 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *Embo J 18*, 5399-410.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. & Jewell, D. 2003. MicroRNAs and other tiny endogenous RNAs in C. elegans. *Curr Biol 13*, 807-18.

Anderson, J.T. 2005. RNA turnover: unexpected consequences of being tailed. *Curr Biol 15*, R635-8.

Aravin, A. & Tuschl, T. 2005. Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett 579*, 5830-40.

Aravin, A.A. 2003. The small RNA profile during Drosophila melanogaster development. *Dev. Cell 5*, 337.

Asa, S.L., Ramyar, L., Murphy, P.R., Li, A.W. & Ezzat, S. 2001. The endogenous fibroblast growth factor-2 antisense gene product regulates pituitary cell growth and hormone production. *Mol Endocrinol 15*, 589-99.

Avery, O.T., MacLeod, C.M. & McCarty, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med. 79*, 137-158.

Badis, G., Fromont-Racine, M. & Jacquier, A. 2003. A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *Rna 9*, 771-779.

Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. & Philippe, H. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *PNAS 99*, 1414-1419.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell 116*, 281-97.

Baumberger, N. & Baulcombe, D.C. 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A 102*, 11928-33.

Benkovic, S.J., Valentine, A.M. & Salinas, F. 2001. Replisome-mediated DNA replication. *Annu Rev Biochem 70*, 181-208.

Berezikov, E., Thuemmler, F., van Laake, L.W., Kondova, I., Bontrop, R., Cuppen, E. & Plasterk, R.H. 2006. Diversity of microRNAs in human and chimpanzee brain. *Nat Genet 38*, 1375-7.

Blackburn, E.H. 2006. Telomerase RNA. In *The RNA world*. Edited by R.F. Gesteland, T.R. Cech & J.F. Atkins. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY. 419-436. pp.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R. & Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell 123*, 1279-91.

Borsook, H., Deasy, C.L., Haagensmit, A.J., Keighley, G. & Lowy, P.H. 1952. Incorporation in vitro of labeled amino acids into proteins of rabbit reticulocytes. *J Biol Chem 196*, 669-94.

Brachet, J. & Chantrenne, H. 1951. Protein synthesis in nucleated and non-nucleated halves of Acetabularia mediterranea studied with carbon-14 dioxide. *Nature 168*, 950.

Brenner, S., Jacob, F., & Meselson, M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature 190*, 576-581.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J. & Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol 18*, 630-4.

Brown, J.W., Echeverria, M. & Qu, L.H. 2003. Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci 8*, 42-9.

Bumcrot, D., Manoharan, M., Koteliansky, V. & Sah, D.W. 2006. RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat Chem Biol 2*, 711-9.

Carthew, R.W. 2006. Gene regulation by microRNAs. *Curr Opin Genet Dev 16*, 203-8.

Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J. & Huttenhofer, A. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A 97*, 14311-6.

Cerutti, H. & Casas-Mollano, J.A. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet 50*, 81-99.

Chanfreau, G., Rotondo, G., Legrain, P. & Jacquier, A. 1998. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *Embo J 17*, 3726-37.

Chen, L., Lullo, D.J., Ma, E., Celniker, S.E., Rio, D.C. & Doudna, J.A. 2005. Identification and analysis of U5 snRNA variants in Drosophila. *Rna 11*, 1473-7.

Chisholm, R.L., Gaudet, P., Just, E.M., Pilcher, K.E., Fey, P., Merchant, S.N. & Kibbe, W.A. 2006. dictyBase, the model organism database for Dictyostelium discoideum. *Nucl. Acids Res. 34*, D423-427.

Claude, A. 1943. The constitution of protoplasm. *Science 97*, 451-456.

Crick, F.H. 1968. The origin of the genetic code. *J Mol Biol 38*, 367-79.

Decatur, W.A. & Fournier, M.J. 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem 278*, 695-8.

Deng, W., Zhu, X., Skogerbo, G., Zhao, Y., Fu, Z., Wang, Y., He, H., Cai, L., Sun, H., Liu, C., Li, B., Bai, B., Wang, J., Jia, D., Sun, S., He, H., Cui, Y., Wang, Y., Bu, D. & Chen, R. 2006. Organization of the Caenorhabditis elegans small non-coding transcriptome: Genomic features, biogenesis, and expression. *Genome Res. 16*, 20-29.

Ebhardt, H.A., Thi, E.P., Wang, M.B. & Unrau, P.J. 2005. Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. *Proc Natl Acad Sci U S A 102*, 13398-403.

Eddy, S.R. 2002. Computational genomics of noncoding RNA genes. *Cell 109*, 137-40.

Eichinger, L., Pachebat, J.A., Glockner, G., Rajandream, M.-A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B.A., Rivero, F., Bankier, A.T., Lehmann, R., Hamlin, N., Davies, R., Gaudet, P., Fey, P., Pilcher, K., Chen, G., Saunders, D., Sodergren, E., Davis, P., Kerhornou, A., Nie, X., Hall, N., Anjard, C., Hemphill, L., Bason, N., Farbrother, P., Desany, B., Just, E., Morio, T., Rost, R., Churcher, C., Cooper, J., Haydock, S., van Driessche, N., Cronin, A., Goodhead, I., Muzny, D., Mourier, T., Pain, A., Lu, M., Harper, D., Lindsay, R., Hauser, H., James, K., Quiles, M., Madan Babu, M., Saito, T., Buchrieser, C., Wardroper, A., Felder, M., Thangavelu, M., Johnson, D., Knights, A., Loulseged, H., Mungall, K., Oliver, K., Price, C., Quail, M.A., Urushihara, H., Hernandez, J., Rabbinowitsch, E., Steffen, D., Sanders, M., Ma, J., Kohara, Y., Sharp, S., Simmonds, M., Spiegler, S., Tivey, A., Sugano, S., White, B., Walker, D., Woodward, J., Winckler, T., Tanaka, Y., Shaulsky, G., Schleicher, M., Weinstock, G., Rosenthal, A., Cox, E.C., Chisholm, R.L., Gibbs, R., Loomis, W.F., Platzer, M., Kay, R.R., Williams, J., Dear, P.H., Noegel, A.A., Barrell, B. & Kuspa, A. 2005. The genome of the social amoeba Dictyostelium discoideum. *435*, 43-57.

Eichinger, L. & Rivero, F. 2006. *Dictyostelium discoideum protocols*. Humana Press. Totowa, NJ. 580 pp.

Erdmann, V.A., Szymanski, M., Hochberg, A., Groot, N. & Barciszewski, J. 2000. Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res 28*, 197-200.
Esquela-Kerscher, A. & Slack, F.J. 2006. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer 6*, 259-69.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. & Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature 391*, 806.
Ganot, P., Bortolin, M.L. & Kiss, T. 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell 89*, 799-809.
Gilbert, W. 1986. Origin of life: The RNA world. *Nature 319*, 618 - 618.
Glockner, G., Szafranski, K., Winckler, T., Dingermann, T., Quail, M.A., Cox, E., Eichinger, L., Noegel, A.A. & Rosenthal, A. 2001. The Complex Repeats of Dictyostelium discoideum. *Genome Res. 11*, 585-594.
Gottesman, S. 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet 21*, 399-404.
Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W. & Watson, J.D. 1961. Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. *Nature 190*, 581-5.
Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell 35*, 849-57.
Halic, M. & Beckmann, R. 2005. The signal recognition particle and its interactions during protein targeting. *Current Opinion in Structural Biology 15*, 116.
Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. 2002. Two classes of short interfering RNA in RNA silencing. *Embo J 21*, 4671-9.
Hamilton, A.J. & Baulcombe, D.C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science 286*, 950-2.
Hammond, S.M. 2005. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett 579*, 5822-9.
Hammond, S.M., Bernstein, E., Beach, D. & Hannon, G.J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature 404*, 293-6.
Hanley, B.A. & Schuler, M.A. 1991. Developmental expression of plant snRNAs. *Nucleic Acids Res 19*, 6319-25.
Heard, D.J., Filipowicz, W., Marques, J.P., Palme, K. & Gualberto, J.M. 1995. An upstream U-snRNA gene-like promoter is required for transcription of the Arabidopsis thaliana 7SL RNA gene. *Nucleic Acids Res 23*, 1970-6.
Heard, E. 2005. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Current Opinion in Genetics & Development 15*, 482.
Hildebrandt, M. & Nellen, W. 1992. Differential antisense transcription from the Dictyostelium EB4 gene locus: implications on antisense-mediated regulation of mRNA stability. *Cell 69*, 197-204.
Hinas, A. & Soderbom, F. 2006. Treasure hunt in an amoeba: non-coding RNAs in Dictyostelium discoideum. *Curr Genet*
Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. & Zamecnik, P.C. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem 231*, 241-57.
Hongay, C.F., Grisafi, P.L., Galitski, T. & Fink, G.R. 2006. Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell 127*, 735-45.
Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. & O'Shea, E.K. 2003. Global analysis of protein localization in budding yeast. *Nature 425*, 686-91.
Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P. & Brosius, J. 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *Embo J 20*, 2943-53.

Huttenhofer, A. & Schattner, P. 2006. The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat Rev Genet 7*, 475-82.
Huttenhofer, A., Schattner, P. & Polacek, N. 2005. Non-coding RNAs: hope or hype? *Trends in Genetics 21*, 289-297.
Huttenhofer, A. & Vogel, J. 2006. Experimental approaches to identify non-coding RNAs. *Nucleic Acids Research 34*, 635.
Johnson, J.M., Edwards, S., Shoemaker, D. & Schadt, E.E. 2005. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet 21*, 93-102.
Kaneda, S., Gotoh, O., Seno, T. & Takeishi, K. 1983. Nucleotide sequence of Dictyostelium small nuclear RNA Dd8 not homologous to any other sequenced small nuclear RNA. *J. Biol. Chem. 258*, 10606-10613.
Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr., Zhu, J.K., Staskawicz, B.J. & Jin, H. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A 103*, 18002-7.
Kessin, R.H. 2001. *Dictyostelium - Evolution, cell biology, and the development of multicellularity.* Cambridge Univ. Press. Cambridge, UK. xiv + 294 pp.
Kimelman, D. & Kirschner, M.W. 1989. An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. *Cell 59*, 687.
Kishore, S. & Stamm, S. 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science 311*, 230-2.
Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M. & Kiss, T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell 85*, 1077-88.
Kiss, T., Marshallsay, C. & Filipowicz, W. 1992. 7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleolus. *Embo J 11*, 3737-46.
Klein, R.J., Misulovin, Z. & Eddy, S.R. 2002. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci U S A 99*, 7542-7.
Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. & Cech, T.R. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell 31*, 147-57.
Kuhlmann, M., Borisova, B.E., Kaller, M., Larsson, P., Stach, D., Na, J., Eichinger, L., Lyko, F., Ambros, V., Soderbom, F., Hammann, C. & Nellen, W. 2005. Silencing of retrotransposons in Dictyostelium by DNA methylation and RNAi. *Nucleic Acids Res 33*, 6405-17.
Lee, R.C., Feinbaum, R.L. & Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell 75*, 843-54.
Lee, R.C., Hammell, C.M. & Ambros, V. 2006. Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. *Rna 12*, 589-97.
Lee, S.R. & Collins, K. 2006. Two classes of endogenous small RNAs in Tetrahymena thermophila. *Genes Dev 20*, 28-33.
Leng, P., Klatte, D.H., Schumann, G., Boeke, J.D. & Steck, T.L. 1998. Skipper, an LTR retrotransposon of Dictyostelium. *Nucleic Acids Res 26*, 2008-15.
Lerner, M.R. & Steitz, J.A. 1979. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A 76*, 5495-9.
Lerner, M.R. & Steitz, J.A. 1981. Snurps and scyrps. *Cell 25*, 298-300.
Lewis, B.P., Burge, C.B. & Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell 120*, 15-20.
Li, H.W. & Ding, S.W. 2005. Antiviral silencing in animals. *FEBS Lett 579*, 5965-73.
Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. 2005a. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol 15*, 1501-7.

Li, S.G., Zhou, H., Luo, Y.P., Zhang, P. & Qu, L.H. 2005b. Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from Schizosaccharomyces pombe. *J Biol Chem 280*, 16446-55.
Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. & Bartel, D.P. 2003. The microRNAs of Caenorhabditis elegans. *Genes Dev 17*, 991-1008.
Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. & Hannon, G.J. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science 305*, 1437-41.
Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J. & Meyers, B.C. 2006. MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res 16*, 1276-88.
Lund, E., Kahan, B. & Dahlberg, J.E. 1985. Differential control of U1 small nuclear RNA expression during mouse development. *Science 229*, 1271-4.
Maden, B.E. 1990. The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol 39*, 241-303.
Maeda, M., Sakamoto, H., Iranfar, N., Fuller, D., Maruo, T., Ogihara, S., Morio, T., Urushihara, H., Tanaka, Y. & Loomis, W.F. 2003. Changing patterns of gene expression in dictyostelium prestalk cell subtypes recognized by in situ hybridization with genes from microarray analyses. *Eukaryot Cell 2*, 627-37.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature 437*, 376-80.
Marker, C., Zemann, A., Terhorst, T., Kiefmann, M., Kastenmayer, J.P., Green, P., Bachellerie, J.P., Brosius, J. & Huttenhofer, A. 2002. Experimental RNomics. Identification of 140 Candidates for Small Non- Messenger RNAs in the Plant Arabidopsis thaliana. *Curr Biol 12*, 2002-13.
Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P. & Nellen, W. 2002. RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol Biol Cell 13*, 445-53.
Matsui, M., Mizuseki, K., Nakatani, J., Nakanishi, S. & Sasai, Y. 2000. Xenopus kielin: A dorsalizing factor containing multiple chordin-type repeats secreted from the embryonic midline. *Proc Natl Acad Sci U S A 97*, 5291-6.
Matzke, M.A. & Birchler, J.A. 2005. RNAi-mediated pathways in the nucleus. *Nat Rev Genet 6*, 24-35.
May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L. & Martienssen, R.A. 2005. Differential regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. *PLoS Genet 1*, e79.
McCarroll, R., Olsen, G., Stahl, Y., Woese, C. & Sogin, M. 1983. Nucleotide Sequence of the Dictyostelium discoideum Small-Subunit Ribosomal Ribonucleic Acid Inferred from the Gene Sequence: Evolutionary Implications+. *Biochemistry 22*, 5858-5868.
Mehler, M.F. & Mattick, J.S. 2006. Non-coding RNAs in the nervous system. *J Physiol 575*, 333-41.
Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. & Tuschl, T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell 15*, 185-97.

Mochizuki, K. & Gorovsky, M.A. 2004. Small RNAs in genome rearrangement in Tetrahymena. *Curr Opin Genet Dev 14*, 181-7.
Moore, P.B. & Steitz, T.A. 2006. The roles of RNA in the synthesis of protein. In *The RNA world*. Edited by R.F. Gesteland, T.R. Cech & J.F. Atkins. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY. 257-285. pp.
Mount, S.M., Gotea, V., Lin, C.F., Hernandez, K. & Makalowski, W. 2007. Spliceosomal small nuclear RNA genes in 11 insect genomes. *Rna 13*, 5-14.
Munroe, S.H. & Zhu, J. 2006. Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective. *Cell Mol Life Sci 63*, 2102-18.
Nazar, R.N. 2004. Ribosomal RNA processing and ribosome biogenesis in eukaryotes. *IUBMB Life 56*, 457-65.
Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev 18*, 1655-66.
Orgel, L.E. 1968. Evolution of the genetic apparatus. *J Mol Biol 38*, 381-93.
Orum, H., Nielsen, H. & Engberg, J. 1992. Structural organization of the genes encoding the small nuclear RNAs U1 to U6 of Tetrahymena thermophila is very similar to that of plant small nuclear RNA genes. *J Mol Biol 227*, 114-21.
Pak, J. & Fire, A. 2007. Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science 315*, 241-4.
Pasquinelli, A., Reinhart, B., Slack, F., Maller, B. & Ruvkun, G. 2000. Conservation across animal phylogeny of the sequence and temporal regulation of the 21 nucleotide C. elegans let-7 heterochronic regulatory RNA. *Nature 408*, 86.
Patel, A.A. & Steitz, J.A. 2003. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol 4*, 960-70.
Pintsch, T., Zischka, H. & Schuster, S.C. 2002. Hisactophilin is involved in osmoprotection in Dictyostelium. *BMC Biochem 3*, 10.
Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A. & Panning, B. 2002. XIST RNA AND THE MECHANISM OF X CHROMOSOME INACTIVATION. *Annual Review of Genetics 36*, 233-278.
Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A.D., Dehay, C., Igel, H., Ares, M., Jr., Vanderhaeghen, P. & Haussler, D. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature 443*, 167-72.
Popova, B., Kuhlmann, M., Hinas, A., Soderbom, F. & Nellen, W. 2006. HelF, a putative RNA helicase acts as a nuclear suppressor of RNAi but not antisense mediated gene silencing. *Nucl. Acids Res. 34*, 773-784.
Poy, M.N. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature 432*, 226.
Reinders, Y., Schulz, I., Graf, R. & Sickmann, A. 2006. Identification of novel centrosomal proteins in Dictyostelium discoideum by comparative proteomic approaches. *J Proteome Res 5*, 589-98.
Reinhart, B.J. & Bartel, D.P. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science 297*, 1831.
Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. & Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature 403*, 901-6.
Rich, A. & Watson, J.D. 1954. Some Relations Between Dna And Rna. *Proc Natl Acad Sci U S A 40*, 759-64.
Rivas, E. & Eddy, S.R. 2001. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics 2*, 8.
Romby, P., Vandenesch, F. & Wagner, E.G.H. 2006. The role of RNAs in the regulation of virulence-gene expression. *Current Opinion in Microbiology 9*, 229.
Rosen, E., Sivertsen, A. & Firtel, R.A. 1983. An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in Dictyostelium. *Cell 35*, 243-51.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H. & Bartel, D.P. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell 127*, 1193-207.

Russell, A.G., Charette, J.M., Spencer, D.F. & Gray, M.W. 2006. An early evolutionary origin for the minor spliceosome. *Nature 443*, 863-6.

Sarkar, N. 1997. Polyadenylation of mRNA in prokaryotes. *Annu Rev Biochem 66*, 173-97.

Schattner, P. 2002. Searching for RNA genes using base-composition statistics. *Nucleic Acids Res 30*, 2076-82.

Schramm, L. & Hernandez, N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev 16*, 2593-620.

Shrimali, R.K., Lobanov, A.V., Xu, X.-M., Rao, M., Carlson, B.A., Mahadeo, D.C., Parent, C.A., Gladyshev, V.N. & Hatfield, D.L. 2005. Selenocysteine tRNA identification in the model organisms Dictyostelium discoideum and Tetrahymena thermophila. *Biochemical and Biophysical Research Communications 329*, 147.

Sierra-Montes, J.M., Pereira-Simon, S., Smail, S.S. & Herrera, R.J. 2005. The silk moth Bombyx mori U1 and U2 snRNA variants are differentially expressed. *Gene 352*, 127-36.

Sijen, T., Steiner, F.A., Thijssen, K.L. & Plasterk, R.H. 2007. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science 315*, 244-7.

Soderbom, F. 2006. Small nucleolar RNAs: identification, structure, and function. In *Small RNAs: analysis and regulatory functions*. Edited by W. Nellen & C. Hammann. Springer-Verlag Berlin Heidelberg. 31-55. pp.

Sucgang, R., Chen, G., Liu, W., Lindsay, R., Lu, J., Muzny, D., Shaulsky, G., Loomis, W., Gibbs, R. & Kuspa, A. 2003. Sequence and structure of the extrachromosomal palindrome encoding the ribosomal RNA genes in Dictyostelium. *Nucleic Acids Res 31*, 2361-8.

Szafranski, K., Lehmann, R., Parra, G., Guigo, R. & Glockner, G. 2005. Gene organization features in A/T-rich organisms. *J Mol Evol 60*, 90-8.

Takeishi, K. & Kaneda, S. 1979. Low molecular weight nuclear RNA species in Dictyostelium discoideum. *Nucleic Acids Symp Ser*, s125-7.

Takeishi, K. & Kaneda, S. 1981. Isolation and characterization of small nuclear RNAs from Dictyostelium discoideum. *J. Biochem. 90*, 299-308.

Talmor-Neiman, M., Stav, R., Klipcan, L., Buxdorf, K., Baulcombe, D.C. & Arazi, T. 2006. Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. *Plant J 48*, 511-21.

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J. & Huttenhofer, A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc Natl Acad Sci U S A 99*, 7536-41.

Tycowski, K.T., Kolev, N.G., Conrad, N.K., Fok, V. & Steitz, J.A. 2006. The ever-growing world of small nuclear ribonucleoproteins. In *The RNA world*. Edited by R.F. Gesteland, T.R. Cech & J.F. Atkins. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY. 327-368. pp.

Upadhyay, R., Bawankar, P., Malhotra, D. & Patankar, S. 2005. A screen for conserved sequences with biased base composition identifies noncoding RNAs in the A-T rich genome of Plasmodium falciparum. *Mol Biochem Parasitol 144*, 149-58.

Waibel, F. & Filipowicz, W. 1990. RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. *Nature 346*, 199.

Valadkhan, S. 2005. snRNAs as the catalysts of pre-mRNA splicing. *Current Opinion in Chemical Biology 9*, 603.

Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev 20*, 515-24.

van Hoof, A., Lennertz, P. & Parker, R. 2000. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol 20*, 441-52.
Vankan, P., Edoh, D. & Filipowicz, W. 1988. Structure and expression of the U5 snRNA gene of Arabidopsis thaliana. Conserved upstream sequence elements in plant U-RNA genes. *Nucleic Acids Res 16*, 10425-40.
Washietl, S., Hofacker, I.L. & Stadler, P.F. 2005. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A 102*, 2454-9.
Wassenegger, M. & Krczal, G. 2006. Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci 11*, 142-51.
Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N. & Imai, H. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev 20*, 1732-43.
Watson, J.D. & Crick, F.H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature 171*, 737-8.
Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev 20*, 759-71.
Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P. & Crete, P. 2004. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell 16*, 69-79.
Werner, A. & Berdal, A. 2005. Natural antisense transcripts: sound or silence? *Physiol Genomics 23*, 125-31.
Wickens, M., Anderson, P. & Jackson, R.J. 1997. Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev 7*, 220-32.
Will, C.L. & Luhrmann, R. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol 13*, 290-301.
Will, C.L. & Luhrmann, R. 2006. Spliceosome structure and function. In *The RNA world*. Edited by R.F. Gesteland, T.R. Cech & J.F. Atkins. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY. 369-400. pp.
Williams, R.S., Boeckeler, K., Graf, R., Muller-Taubenberger, A., Li, Z., Isberg, R.R., Wessels, D., Soll, D.R., Alexander, H. & Alexander, S. 2006. Towards a molecular understanding of human diseases using Dictyostelium discoideum. *Trends Mol Med 12*, 415-24.
Win, T.Z., Draper, S., Read, R.L., Pearce, J., Norbury, C.J. & Wang, S.W. 2006. Requirement of fission yeast Cid14 in polyadenylation of rRNAs. *Mol Cell Biol 26*, 1710-21.
Wise, J.A. & Weiner, A.M. 1980. Dictyostelium small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. *Cell 22*, 109-118.
Wise, J.A. & Weiner, A.M. 1981. The small nuclear RNAs of the cellular slime mold Dictyostelium discoideum. *J. Biol. Chem. 256*, 956-963.
Woese, C. 1967. *The genetic code*. Harper & Row. New York. 200 pp.
Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A. & Wagner, E.G. 2003. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res 31*, 6435-43.
Voinnet, O. 2005. Induction and suppression of RNA silencing: insights from viral infections. *Nat Rev Genet 6*, 206-20.
Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. & Martienssen, R.A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science 297*, 1833-7.
Yoshida, H., Kumimoto, H. & Okamoto, K. 1994. dutA RNA functions as an untranslatable RNA in the development of Dictyostelium discoideum. *Nucleic Acids Res 22*, 41-6.
Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. & Chen, X. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science 307*, 932-5.

Yuan, G., Klambt, C., Bachellerie, J.P., Brosius, J. & Huttenhofer, A. 2003. RNomics in Drosophila melanogaster: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res 31*, 2495-507.

Yukawa, Y., Felis, M., Englert, M., Stojanov, M., Matousek, J., Beier, H. & Sugiura, M. 2005. Plant 7SL RNA genes belong to type 4 of RNA polymerase III- dependent genes that are composed of mixed promoters. *Plant J 43*, 97-106.

Zamore, P.D. & Haley, B. 2005. Ribo-gnome: the big world of small RNAs. *Science 309*, 1519-24.

Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell 101*, 25-33.

Zuker, C., Cappello, J., Lodish, H.F., George, P. & Chung, S. 1984. Dictyostelium transposable element DIRS-1 has 350-base-pair inverted terminal repeats that contain a heat shock promoter. *Proc Natl Acad Sci U S A 81*, 2660-4.

6. Acknowledgements

First and foremost, I want to thank my eminent supervisor **Fredrik** for accepting a complete stranger as your first PhD student five years ago. Taking part in building up your research group has been extremely educational and inspiring. I am really impressed by your crazily enthusiastic, yet controlled way of conducting research as well as your diplomatic skills (I could probably use a little more of those myself). Thanks also for being a real friend in need, great Slime BBQs, Roskilde trips, Reggae festival company and the coolest hairdo in the department (Sorry Al!).

Gerhart, my second supervisor, thank you for great discussions, and for writing countless letters of recommendation without (too much) complaint. They seem to have worked well!

Prof. John Näslund's foundation (Norrlands nation), Knut and Alice Wallenberg's foundation (SLU), and the Swedish Royal Academy of Sciences are gratefully acknowledged for financing my attendance to conferences and courses.

Past and present members of the Uppsala Dicty group: **Anders** for great sense of humour and introducing me to the Ketchup radio show (I think we were both well above the target group age though); **Lotta** for drawing some of Fredrik's attention away from my experiments during this past year, for taking the beer club to a completely different (definitely better!) level and for being a very cool person. Project students: **Björn G** for great help with the GC-rich RNA project and **Louise A** for skilful work on the siRNA project.

Thanks also to collaborators during the years: **Pontus L** for finding more RNAs than anyone could wish for and being so patient; **Johan R** for understanding the mind of an experimentalist ("this is what you need" is a great way to name an Excel sheet!); **Victor Ambros** at Dartmouth Medical School for constructing the small RNA cDNA library, never-ending optimism, patience and letters of recommendation. **Wolfgang Nellen** for letters of recommendation, interesting discussions and taking good care of me in Germany. All other members of the Kassel group for making me feel at home and trying to teach me some German during my two months in your lab.

I would also like to thank my two fellow PhD musketeers: **Martin W** (PhD – how hard can it be?) for great company on the climbing wall (although we really suck) and beer club co-regency; and super room-mate **Urszula** for being my social and organizational role model (I still have a long way to go…). It has been great to be able to share half-time evaluation stress, paper-writing stress and just plain stress with the two of you all these years, I will miss that. Thanks also for trying to teach me some crystallography, even though I never remember which kinase you are currently working on…

For superb technical assistance: **Christer**, there seems to be nothing you cannot fix (or do not keep somewhere in your office). **Erling**, thanks for great computer support and patience with computer illiterates. I promise I will never again take computer advice from Fredrik.

Past and present xray people for making BMC a nice place to be: **Rosie** and **Malin** for room-sharing during my first PhD months, **Andreas K**, **Hasse**, **Stefan**, **Norsk-Nisse**, **Margareta**, **Ulla**, **Elleonor**, **Anton**, **Anatoly**, **Tom**, **Inger**, **Linda**, **Saeid**, **Evalena**, **Jens**, **Martin**, **Martin** , **Sinisa**, **Martin**… and the rest of the **Åqvist** group, Angel **Torsten**, **Nisse**

(oldest Uppsala friend!), **Annette**, **Jenny**, **Al**, **Emma**, **Anna S-L**, **Adrian, Jerry**, **Jonas**, **Mats etc.**

Microbiology/ICM at Uppsala University: Thanks for selflessly sharing your radioactivity room! In particular **Magnus L** for sharing my terrible sense of humour; probably distant relative **Mathias B** for Gammalsvenskby and hair color discussions; **Klas U** for hopeless conversations; **Ema K** for keeping stylish even after more than four years in the lab; **Erik** for being nice and smart at the same time; **Ulrika** for always helping out; **Staffan S**, **Emma R**, **Cia, Helene, Leif, Shiying, Santanu, Nora, Nina, Sonchita and everyone else.**

All the **Protein Technology students** for letting me practice my teaching skills, at least I enjoyed it!

My hairdresser **Kerstin** for trying hard to keep me from looking too much like a scientist.

Craig Hunter for relieving a great deal of stress by accepting me as a postdoc well before I even started writing this thesis and for already making me feel at home in your group.

My friends: **Anna S** for dammsugar- and dubbelfika, and more recently, with the help of **Gonçalo,** for my first Portuguese wedding. **Hanna Li** and **Hanna La** for great trips to London, Helsinki, Berlin, and Halland, and for the excellent 23 December tradition. **Emma V** for our yearly get-togethers. **Andreas C** for being a great friend and technology consultant.

Orchestra buddies from **Kungliga Akademiska Kapellet** for providing well-needed distraction from RNA and slime molds in the form of music, parties, and "tours" to Oxford, Lund and Linköping, in particular cello players **Tina**, **Anna K**, **Maria S**, **Stina** and **Björn T**.

Relatives: uncle **Leffe**, I still find it hard to grasp that you are not around anymore. Aunt **Tutta** and favourite cousins from the countryside **Jens** and **Annika**, I look forward to visiting you in Lund/Malmö soon. Aunt and uncle **Gertrud & Göran** for good company at Gammalsvenskby reunions. All of you who are like relatives: the **Björne & Britt**, **Moa**, **Mikaela**, **Tomas**, **Julia**, **Lisa**, **Gunilla**, **Magnus & Eva**, **Eva & Håkan**, **Viveka**, **Marianne & Anders**, thanks for being so enthusiastic about my dissertation party, I am sure we will have fun!

Enormous thanks go to my fantastic family: **Mamma** and **Pappa** for endless support and for always trying hard to understand what I do. **Farmor**, the most positive person I know, for always bragging about me and hardly missing a single of my concerts in almost 20 years. **Torsten** for encouraging my athletic (well…) ambitions and for knowing all the sports facts. My brother **Kristian** with all the adult points, for always helping out, **Fia** for being a cool and caring person and **Doris** for at least sometimes leaving my shoes alone. My sister **Maria** and brother-in-law **Jannis** for nice dinners and discussions, and to the greatest vertebrate nieces ever **Petrina** and **Nelly**.