CHARACTERIZATION OF EXOGENOUS AND ENDOGENOUS PLANT SMALL RNAs: DISCOVERY OF A UBIQUITOUS TERMINAL MODIFICATION.

by

Holger **Alexander** Ebhardt
Dipl.-Ing., University of Applied Sciences Jena, Germany, 1999

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Spring 2007

APPROVAL

Name: **Holger Alexander Ebhardt**

Degree: **DOCTOR OF PHILOSOPHY**

Title of Thesis: **CHARACTERIZATION OF EXOGENOUS AND ENDOGENOUS PLANT SMALL RNAs: DISCOVERY OF A UBIQUITOUS TERMINAL MODIFICATION.**

Examining Committee:

Chair: **Dr. Nancy Hawkins**
Assistant Professor, Department of Molecular Biology and Biochemistry

---

Dr. Peter J. Unrau
Senior Supervisor
Assistant Professor, Department of Molecular Biology and Biochemistry

---

Dr. Fiona S.L. Brinkman
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

---

Dr. Dipankar Sen
Supervisor
Professor, Department of Molecular Biology and Biochemistry

Dr. Jim Mattsson
Internal Examiner
Assistant Professor, Department of Biological Sciences

Dr. Xuemei Chen
External Examiner
Associate Professor, Botany & Plant Sciences
University of California Riverside

Date Defended/Approved: March 28, 2007

ABSTRACT

RNA silencing is gene regulatory process mediated by 21 to 25 nt RNA which originate from longer single- or double-stranded RNA transcripts. After synthesis, the 21 to 25 nt RNAs are incorporated into argonaute containing RNA-protein complexes acting either pre- or post-transcriptionally – silencing heterochromatin or silencing RNA transcripts, respectively.

Plants can be infected with viral satellites, which are small infectious agents comprised of 200-400 nt single-stranded RNA genomes. These satellite genomes lack obvious protein coding regions and adopt characteristic secondary structures. Viral satellites depend on helper viruses for replication as the latter provide the RNA-dependent RNA polymerase. One viral satellite system is the Cucumber Mosaic Virus (CMV) and its Y-Satellite RNA (Y-Sat). This helper virus (CMV) and satellite (Y-Sat) pair can infect a range of plants including tobacco, causes particular tobacco strains to yellow. In a recent study, Wang et al. (PNAS 101:3275-3280) demonstrated that symptoms of the CMV Y-Sat were no longer evident despite an accumulation of Y-Sat RNA. This symptom suppression was due to an endogenously expressed strong RNA silencing suppressor called P1/HC-Pro.

I cloned and sequenced 698 small RNA from infected and non-infected plants. Analyzing the cloning data I observed a sharp discrepancy between the size distribution of cloned small RNAs in the histogram versus the autoradiogram of 5' radiolabelled RNA in P1/HC-Pro(+) CMV/Y-Sat(+) plants. Using enzymatic assays and a new band-shift assay I proved that viral small RNA in P1/HC-Pro expressing plants were not modified on the 2'-hydroxyl of their 3' terminus, but commonly cloned endogenous micro RNAs were modified. The analysis of these results was greatly facilitated by my software called Ebbie, an automated sequencing data analysis pipeline using a dynamic web page, which assists the process of annotating cloned small RNAs. These findings imply a separation between the multiple small RNA biogenesis pathways for RNA silencing in plants *via* sub-cellular localization and/or recruiting distinct sets of proteins by each pathway. My research also showed that all small RNAs in plants have a 3' terminal modification, a significant difference between animal and plant RNA silencing pathways.

Keywords: RNA silencing; virus infection; small RNAs; 2'-O-Methyl; RNA silencing suppressor HC-Pro; software *Ebbie*.

DEDICATION

Schlaf, Kindchen, schlaf.

Dein Vater hüt' die Schaf.

Die Mutter schüttels Bäumelein,

da fällt herab ein Träumelein.

Schlaf, Kindchen, schlaf.

Schlaf, Kindchen, schlaf!

Am Himmel zieh'n die Schaf.

Die Sterne sind die Lämmerlein,

der Mond, der ist das Schäferlein.

Schlaf, Kindchen, schlaf!

...

Volksmund – German lullaby.

ACKNOWLEDGEMENTS

Zuerst möchte ich mich bei meiner Familie für deren Unterstützung meiner Studien bedanken. This includes my wife Ashley J Ebhardt and daughter Lilyanna Ebhardt. Thank you for being so patient when I had to go to the lab on weekends or to the coffee shop for writing my thesis.

This thesis project and Ph.D. graduation would not have been possible without the guidance of my senior supervisor Dr. Peter J Unrau. I am also thankful to my supervisory committee comprised of Drs. Fiona SL Brinkman, Mark Paetzel and Dipankar Sen. I also like to take this opportunity to thank my collaborators who are co-authors or appear in the acknowledgement of my publications.

Chapter 2 was essentially published previously: Ebhardt HA, Wiese KC, Unrau PJ: *Ebbie*: automated analysis and storage of small RNA cloning data using a dynamic web server. BMC Bioinformatics: 2006 Apr 3; 7:185. Chapter 3 was partly published: Ebhardt HA, Thi EP, Wang MB, Unrau PJ: Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. Proc Natl Acad Sci U S A. 2005 Sep 20;102(38):13398-403.

Last, but not least, I like to thank everyone I met here at SFU who shared their enthusiasm for science with me, or a simple smile. If your defense is coming up, I am convinced you will be able to access all the knowledge you accumulated during your graduate career and impress your examining committee.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Degenerative nucleotides acronyms

N=A,G,C,T	**R**=A,G	**Y**=C,T	**M**=A,C	**W**=A,T	**S**=G,C
K=G,T	**D**=G,A,T	**H**=A,T,C	**B**=G,T,C	**V**=G,A,C	

Unit prefixes

mM	milli Molar (10^{-3})
µM	micro Molar (10^{-6})
nM	nano Molar (10^{-9})
pM	pico Molar (10^{-12})
fM	fempto Molar (10^{-15})
aM	atto Molar (10^{-18})

Abbreviations used in thesis in alphabetical order

19hr2	19 bp RNA helixes with a 3' 2 nt overhang on each side
2'-O-Me	2'-O-methyl
3-CP	3' cloning primer
5-CP	5' cloning primer
bp(s)	base pair(s)
C.elegans	Caenorhabditis elegans
CMV	cucumber mosaic virus
Da	Dalton
DCL	Dicer-like enzyme
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
dpi	days post induction
GenBank	Nucleotide database at the NCBI
GFP	green fluorescence protein

h	hour(s)
het	heterogeneous
hom	homogeneous
HC-Pro	Helper Component Proteinase
IR-PTGS	inverted-repeat post-transcriptional gene silencing
min	minute(s)
miRNA	micro RNA
mRNA	messenger RNA
nat-siRNA	natural anti-sense silencing RNA
NCBI	National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/
nt(s)	nucleotide(s)
PCR	polymerase chain reaction
PVX	potato virus X
PVY	potato virus Y
ra-siRNA	repeat associated silencing RNAs
RDR	host RNA dependent RNA polymerase
RdRp	viral RNA dependent RNA polymerase
RISC	RNA induced silencing complex
RITS	RNA-induced initiation of transcriptional gene silencing
RNA	Ribonucleic acid
rRNA	ribosomal RNA
SAM	S-Adenosyl Methionine
sec	second(s)
siRNA	silencing RNA
smRNA	small RNA
t_0	time 0 (start time)
ta-siRNA	trans-acting silencing RNA
t_{end}	time at final end point
TEV	tobacco etch virus
TGS	transcriptional gene silencing
Tris	Tris (hydroxymethyl) aminomethane
Y-Sat	cucumber mosaic virus Y-Satellite

1 INTRODUCTION

The Nobel Prize in Physiology or Medicine in 2006 was shared by Andrew Z. Fire (Stanford University, California, USA) and Craig C. Mello (University of Massachusetts Medical School in Worcester, USA). The honour was given to these two principal investigators for demonstrating in the nematode *Caenorhabditis elegans* that double stranded RNA directs cleavage of messenger RNAs in a homologous manner. This process was termed RNA interference in animals. Earlier experiments in plants demonstrated that gene silencing occurs at the transcriptional and/or post-transcriptional level, two processes caused by homologous- or viral induced gene silencing. I'd like to dedicate my thesis to the plant biologists who discovered RNA silencing. The next paragraphs will provide an historic overview of some gene silencing experiments in plants as well as in the namatode *Caenorhabditis elegans* discussing some key experiments of more than a decade of research on RNA silencing.

1.1 Homologous gene silencing

The term homologous gene silencing was coined when DNA sequences were introduced into genomes and the newly introduced sequence and homologous sequences were found not expressed or 'silenced'. In the early 1990's there were many examples when protein coding genes, repetitive DNA elements or viral sequences were artificially introduced into plants and the newly introduced sequences were silenced (see Introduction starting at 1.1.1). Following these observations, the question arose if this gene silencing was the result of silencing the genomic loci from which they originated – termed transcriptional gene silencing – or if the genomic locus was transcribed by an RNA polymerase and the message was subsequently degraded in either the nucleus or the cytoplasm – termed post-transcriptional gene silencing.

1.1.1 Post-transcriptional gene silencing

Experiments studying the pigmentation of petunia flowers were some of the early experiments where researchers encountered post-transcriptional gene silencing. Petunia flower pigmentation is comprised of anthocyanins and flavonols, which are synthesized by the flavonoid pathway (for review see (Grotewold, 2006)). By expressing chalcone synthase, the first enzyme

required in the flavonoid pathway, it was expected to enhance floral coloration.

Surprisingly, two research groups noticed that the introduction of additional copies of the chalcone synthase in Petunia, the initial enzyme required in the flavonoid pathway, did not result in the expected enhancement of floral coloration (Napoli, et al., 1990; van der Krol, et al., 1990). On the contrary, additional copies of the chalcone synthase gene resulted in no change of flower colouring or even a 25 % reduction in the colouring of transformed plants. Further molecular analysis revealed that chalcone synthase messenger RNA levels were reduced. Napoli and colleagues termed this phenomenon "co-suppression" and speculated on its mechanism. Van der Krol and colleagues presented the idea of RNA strands interfering with the transcription process itself, e.g. by forming a triple helix made up of double-stranded DNA and single-stranded RNA. As a possible mechanism both groups entertained the idea of DNA methylation being the cause for suppressing messenger RNA (mRNA).

Following the initial research on the co-suppression of chalcone synthase in Petunia, van Blokland and colleagues isolated cell nuclei from Petunia plants and provided the active RNA polymerases with [α-^{32}P] UTP resulting in radiolabelled transcripts. Following purification of these radiolabelled transcripts, they are used on micro array chips to confirm the transcript's sequence. Thus, if a radiolabelled chalcone synthase transcript is confirmed using the micro array, it is silenced post-transcriptionally in the cell. This assay termed nuclear run-on experiment is able to distinguish between transcriptional and post-transcriptional gene silencing (Eick, et al., 1994). Using nuclear run-on experiments, the research group did not find a difference in the expression levels of chalcone synthase mRNA and other non-suppressed mRNAs. Thus, they concluded that the reduction of chalcone synthase was not caused by transcriptional inactivation but rather the suppression had to occur post-transcriptionally due to increased mRNA turnover presumably in the cytoplasm (Vanblokland, et al., 1994).

1.1.2 Transcriptional gene silencing

The soil bacterium *Agrobacterium tumefaciens* can cause various tumors within a plant. The bacterium infects plants with a plasmid sequence termed transferred DNA (T-DNA), which is incorporated into the plant's genome and encodes virulent proteins causing disease symptoms. Various plasmids have been isolated and classified based on their pathogenesis: Ti (tumor-inducing) or Ri (root-inducing) plasmids. These pathogenic plasmids were extensively studied to elucidate the mechanism by which the pathogenic DNA sequence is inserted into the host's genome. Using this observation, geneticists used this bacterial system to introduce their own

genes into plants (Schell, Van Montagu, 1977). A typical Ti-plasmid is comprised of the transgene of interest and a selectable marker which are both flanked by a 25 bp T-DNA repeat sequence 5'-TGACAGGATATATTGGCGGGTAAAC-3' (Wang, et al., 1984). The T-DNA portion of the bacterial plasmid is incorporated into the host's genomic DNA. Using a selectable marker such as kanamycin-resistant gene the successfully transformed plant will survive in media containing kanamycin.

Matzke and colleagues used two homologous T-DNAs to introduce two antibiotic resistant genes into tobacco plants. Initially kanamycin resistant gene termed T-DNA-I was transformed into tobacco plants. Subsequently the same plants were transformed with a T-DNA homologous to T-DNA-I containing hygromycin resistance (T-DNA-II). Some of these sequentially transformed tobacco plants exhibited a loss of function for kanamycin resistance, the primary transformation. Using methylation sensitive restriction enzyme (*Sst*II) assays, Matzke and colleagues proved that *Sst*II successfully cleaved the promoter region of T-DNA-I from the primary transformed plants. In double transformed plants, *Sst*II was unable to cleave at the same position, indicating a complete methylation of the T-DNA-I promoter region occurred. Thus, this research clearly demonstrates that homologous DNA, in this case T-DNA-II was able to silence T-DNA-I at the transcriptional level (Matzke, et al., 1989).

1.1.3 Viral-induced gene silencing

Wassenegger and colleagues noted that many examples in the literature show hypo- or hypermethylation of genomic DNA during cell development (for review, see (Cedar, Razin, 1990; Razin, Cedar, 1991)). But, little progress was made in elucidating the molecular mechanisms of *de novo* DNA methylation. Wassenegger and colleagues studied the *potato spindle tuber* viroid (PSTVd) in tobacco plants. This RNA viroid amplifies *via* an asymmetric rolling-cycle by which the (+)-strand serves as template for generating multimeric linear copies of the (–)-strand which serves as template for generating more (+)-strand viroid RNA. This process can be inhibited by α-amanitin which binds to the funnel-shaped cavity of RNA polymerase II (Bushnell, et al., 2002), suggesting that this process depends on RNA polymerase II for transcription (Muhlbach, Sanger, 1979).

Wassenegger and colleagues inserted multiple copies of the PSTVd DNA into T-DNA plasmids and transformed them into the tobacco genome using *Agrobacterium*. The methylation state of genomic DNA was monitored using Southern hybridization and methylation sensitive restriction enzymes (*Msp*I, *Hpa*II, *Hae*III). Once the viroid RNA started replicating

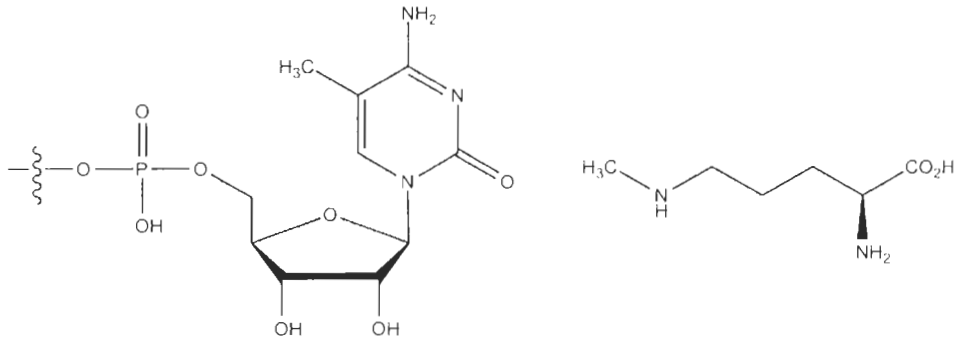
autonomously, the genomic PSTVd DNA became methylated. To prove that homologous gene silencing was possible using these artificially introduced sequences, a secondary transformation using a truncated PSTVd DNA was performed. This truncated PSTVd DNA can not initiate self-replication but was soon methylated. The methylation was restricted to the PSTVd DNA sequence and no other methylation state changes could be observed in four genes tested or the T-DNA flanking the PSTVd sequence. As the PSTVd DNA was only methylated once autonomous replication of the viroid RNA starts, the authors concluded that the *de novo* DNA methylation of the (partial) PSTVd DNA locus required RNA as an intermediate and that post-transcriptional processes are important for silencing (Wassenegger, et al., 1994).

1.1.4 DNA methylation and histone methylation

Up to 30 % of cytosines can be methylated in plant genomes. Cytosine is methylated on the 5-position of the pyrimidine ring and is not distributed randomly throughout the genome. There are at least three DNA methylation patterns: CpG, CpNpG and CpHpH (where N is A, C, G or T and H is A, T or C) (Cao, Jacobsen, 2002). Besides DNA methylation, the linear genomic DNA in eukaryotes is packaged into higher structures, where the DNA wraps around proteins called histone. The lysine residues of histones are often methylated contributing to the transcriptional silencing of genomic regions (Zilberman, et al., 2003). Both DNA methylation and histone methylation contribute to the silencing of parts of the genome often referred to as heterochromatin. Open and accessible regions of the genome are referred to as euchromatin (Allison, 2007).

Figure 1.1: DNA and histone modifications.

Depicted here are two common methylation events: 5-methylcytosine (directly methylating a DNA base) and methylated lysine (methylating a lysine residue within histones).



5-methylcytosine **metylated lysine**

There are two types of methylation, which require distinct methytransferases. Maintenance methylation retains the methylation pattern during replication, in other words the newly synthesized DNA strand will show the same methylation pattern as the parent strand. *De novo* methylation is the process by which a very limited DNA locus is methylated. Maintenance methylation is catalyzed by the DNA methyltransferase Dnmt1, whereas Dnmt3a and Dnmt3b catalyze *de novo* methylation (Brown, 2002).

Methylation sensitive DNA restriction enzymes are often used to verify the methylation state of genomic DNA. This approach has the disadvantage of only comparing overall restriction enzyme cleavage patterns and relies on a negative result, e.g. a band appears in one experiment and not in another, and therefore the DNA might be methylated. An alternative methodology providing a positive signal is the bisulfite treatment of isolated DNA by which cytosine bases are converted into uracil unless the cytosine is methylated on the 5 position of the pyrimidine ring (5-methylcytosine). The treated DNA is then PCR amplified and uracil bases are replaced by thymine. Sequencing individual PCR sequences gives a reliable reflection of the methylation pattern of the amplified DNA sequence (Frommer, et al., 1992).

1.2 Mechanism of gene silencing

Plant biologists and researchers in *C. elegans* established that double-stranded RNA was required for gene silencing by means of transcriptional and post-transcriptional gene silencing. The size of the foreign DNA greatly varied from experiment to experiment but did not affect the gene silencing results (for references, see Introduction 1.1). It was also known that double stranded RNA played a major role in the process of transcriptional (Wassenegger, et al., 1994) or post-transcriptional gene silencing (Fire, et al., 1998). But, how this double stranded RNA of various lengths could affect gene expression on various levels was still a mystery.

There were already some hints that small RNAs could play a major role in RNA silencing. In *C. elegans* a lethal mutation in *lin-4* was linked to normal temporal control of LIN-14 messenger RNA. *Lin-4* temporally down-regulates LIN-14 in the first larval stage (L1). Lee and colleagues were able to locate the *lin-4* regulatory element in four *Caenorhabditis* species and ruled out a protein coding region. Interestingly, all four *Caenorhabditis* clones functionally rescued the *lin-4* null allele of *C. elegans*. Within the functional region was a 22 nt RNA element complementary to the 3' UTR of LIN-14 suggesting an anti-sense mechanism by which *lin-4* regulates expression levels of LIN-14 (Lee, et al., 1993).

1.2.1 Link between small RNAs and post-transcriptional gene silencing

In 1999, Hamilton and Baulcombe performed multiple experiments linking the gene silencing effects to small RNAs ~25 nts in length. First in analogy to Napoli *et al.* 1990 and van der Krol *et al.* 1990, tomato plants were transformed with a tomato 1-aminocyclopropane-1-carboxylate oxidase (ACO) cDNA. Two of five tomato plants exhibited transgene-induced post-transcriptional gene silencing effect and showed in Northern hybridization experiments high levels of ~25 nt antisense small RNA when probing with sense ACO probes.

Three similar experiments all showed these ~25 nt small RNAs to be responsible for post-transcriptional gene silencing, even when the gene silencing was of viral cause. The authors postulated that the ~25 nt RNA serve "as molecular markers" for gene silencing (Hamilton, Baulcombe, 1999).

In 2000, Reinhart and colleagues also identified a 21 nt RNA designated *let-7* that regulates developmental timing in the nematode *C. elegans*. *let-7* anneals to the 3' untranslated region of various messenger RNAs (lin-14, lin-28, lin-41 and daf-12), thus causing transcriptional inhibition of these messenger RNA (Reinhart, et al., 2000). In 2001, it became clear that there was a multitude of 21 to 25 nt RNAs both in plants and animals, that could be utilized in a general mechanism by which small RNAs assume a range of regulatory roles (Lagos-Quintana, et al., 2001; Lau, et al., 2001 Lee, Ambros, 2001).

Following the discovery of ~25 nt RNA (Hamilton, Baulcombe, 1999), the same research group identified a second class of small RNA in plants that were 21 to 22 nt in length (Hamilton, et al., 2002). The researchers transformed green fluorescence protein (GFP) into *Nicotiana benthamiana*. Transgenic plants typically display a localized green fluorescence at the site of injection which peaks after two to three days and then dissipates. This silencing effect is associated with two distinct classes of small RNA ranging in length between 21 to 22 nts and 25 nts. Expressing various viral RNA silencing suppressors, the amount of GFP messenger RNA was inversely related to the abundance of small RNA. For example, the viral RNA silencing suppressor P19 from of *tomato bushy stunt virus* caused a compete elimination of GFP related small RNAs 5 and 11 days post induction resulting in the translation of GFP and subsequently a fluorescent signal.

Establishing two sizes of small RNA involved in gene silencing, Hamilton and colleagues addressed the question of small RNA originating from transposable elements in *Arabidopsis thaliana* and *Nicotiana benthamiana*. Using Northern hybridization, all three tested transposable

elements which showed 24 to 26 nt RNA, but no 21 to 22 nt RNA. To establish a link between the 24 to 26 nt long RNA and DNA methylation, the research group used methylation sensitive restriction enzymes (*Hae*III and *Sau*3A1) to digest the *Arabidopsis* AtSN1 transposable element in wild-type and silencing-defective mutants. PCR amplification using AtSN1 specific primers resulted in PCR products in wild-type and most silencing-defective mutants, except the silencing-defective mutant *sde3*. This result suggests that at least two distinct silencing pathways exist within *Arabidopsis*.

1.3 Small RNA biogenesis in animals

Half a decade of research into small RNAs in plants and animals revealed two main classes of regulatory small RNAs: silencing and micro RNAs (siRNA and miRNA). The former is generated from double stranded RNA transcripts while the latter is generated from single stranded RNA transcripts that fold back on themselves forming an imperfect stem-loop structure, often referred to as hairpins (Jones-Rhoades, et al., 2006).

1.3.1 RNA silencing in animals

RNA silencing is a multi-facetted regulatory mechanism in plants. In animals, the major post-transcriptional RNA silencing pathway is called RNA interference (RNAi). The first RNAi experiments in animals were done by Fire and colleagues. Their model organism *Caenorhabditis elegans* encodes the *unc-22* gene which encodes a myofilament protein. Low levels of this protein result in severe twitching of the namatode. Thus, if the *unc-22* gene could be silenced, the worm would twitch. The research group injected single stranded sense or antisense RNA corresponding to *unc-22* mRNA into the nematode with no observable consequence. Only when double stranded RNA was injected did the worm twitch severely, demonstrating that the *unc-22* was silenced (Fire, et al., 1998).

Further research into the mechanism of RNAi revealed, that this pathway is triggered when double stranded RNA encounters the RNase III enzyme Dicer. Dicer is found in the cytoplasm where it cuts long dsRNA into 21 – 25 nts long double stranded silencing RNAs (siRNA). In analogy to miRNA, one strand of the siRNA-duplex is incorporated into RISC. Unlike typical miRNAs in animals targeting the 3'UTR causing inhibition of the mRNA, siRNA guide RISC to mRNAs for an endonucleolytic cleavage event, leading to the degradation of mRNA thus silencing gene expression posttranscriptionally.

Besides primary siRNA derived from Dicer cleavage it has been postulated that there is a

class of secondary siRNA, whose synthesis require RNA-dependent RNA polymerases. Analyzing small RNA populations in *C. elegans*, Pak and Fire established that the vast majority of siRNAs are actually secondary siRNAs (Pak, Fire, 2007).

1.3.2 Micro RNA biogenesis in *C.elegans*

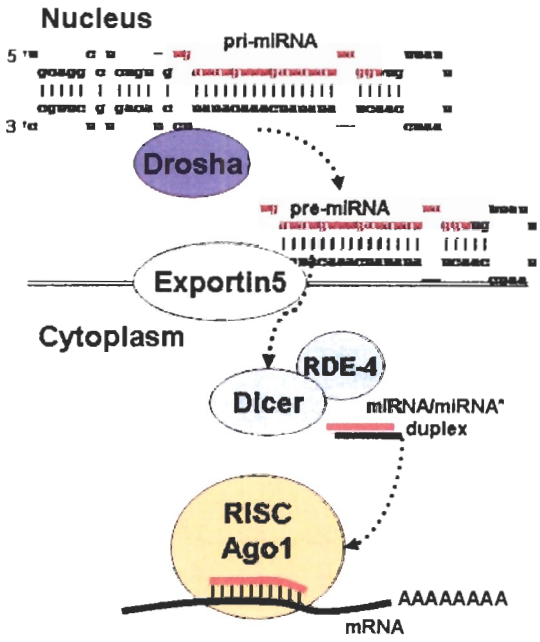
Micro RNAs (miRNAs) are transcribed from genomic DNA by RNA polymerase II (Lee, et al., 2004). The transcript folds into a characteristic imperfectly paired stem loop structure termed primary miRNA (pri-miRNA) (Lim, et al., 2003). This long pri-miRNA is recognized by ribonuclease Drosha and truncated to a shorter hairpin, which in turn is exported from the nucleus to the cytoplasm requiring Exportin 5 as protein factor (Bohnsack, et al., 2004; Lund, et al., 2004; Yi, et al., 2003). In the cytoplasm, the hairpin is further cleaved by Dicer and RDE-4 (R2D2 in other animals), creating the miRNA/miRNA* 19 bp duplex RNA with a 2 nt overhang on each end (Liu, et al., 2003; Tabara, et al., 2002). The miRNA of this duplex is incorporated into the RNA-induced silencing complex (RISC) while the miRNA* degrades rapidly (Schwarz, et al., 2003). A core component of RISC is an argonaute protein, which cleaves the targeted mRNA (Liu, et al., 2004; Song, et al., 2004). Besides the core component, many other proteins, including Dicer, are thought to be included in RISC. Current size estimates of a native RISC ranges from several hundred kDa to 80S (Filipowicz, 2005).

The mature micro RNA has a 5' phosphate and a 3' hydroxyl (Lim, et al., 2003). The 21 nts long miRNA contains a 7 nt seed sequence which contributes the bulk of miRNA – mRNA binding energy. Mutational changes within the seed sequence can alter the target mRNA completely (Schwarz, et al., 2006). This seed sequence is important in genome evolution of animals as the 3'UTR and the corresponding miRNA co-evolve. Using comparative genome analysis in mammals, Farh and colleagues uncovered evidence for selective avoidance, whereby genes avoid certain 7mers in their 3'UTR avoiding the translational inhibition by miRNA guided RISC. These 7mers are often sufficient for miRNA-directed translational inhibition of mRNAs (Farh, et al., 2005). How the binding of micro RNAs to the 3'UTR causes translational inhibition is still debated (Chu, Rana, 2006; Liu, et al., 2005; Mendell, 2005).

Many research groups established the importance of miRNAs in the development of organs and organisms by governing the temporal and spatial expression of mRNAs. This process guides cell differentiation, but it is not clear if miRNA are the cause or reaction of this process (Bartel, 2004; Jones-Rhoades, et al., 2006). Figure 1.2 shows the miRNA pathway in animals and how miRNAs affect translation of mRNAs.

Figure 1.2: Micro RNA pathways in *C.elegans*.

Micro RNAs (miRNAs) are transcribed from genomic DNA. The single stranded RNA folds into a characteristic hairpin termed primary miRNA (pri-miRNA). In animals, Drosha reduces the length of the pri-miRNA obtaining the precursor miRNA (pre-miRNA); the pre-miRNA is exported into the cytoplasm by Exportin5 where the pre-miRNA is cleaved by Dicer obtaining the miRNA/miRNA* duplex. The miRNA is incorporated into an argonaute containing RNA-induced silencing complex (RISC). Through the complementary between miRNA and the target mRNA, RISC is guided to the target mRNA and causes translational inhibition (Filipowicz, 2005).



Micro RNA mediated translational inhibition of mRNA.

1.4 Small RNA biogenesis in plants

RNA silencing in plants is described in the literature using multiple names, depending on the double stranded RNA's origin. The main idea is the existence or generation of endogenous or exogenous double stranded RNA which is cleaved into 21 to 25 nt long RNA duplexes by Dicer-like proteins. Of the short duplex one strand is incorporated, or 'loaded', into an argonaute protein containing RNA-protein complex. The latter complex might recruit other protein factors and is the biologically active complex acting either pre- or post-transcriptionally – silencing heterochromatin or silencing RNA transcripts, respectively.

Primary transcripts are typically generated by RNA polymerase II, in the case of mRNAs, or RNA polymerase IV, in the case of transposable and repetitive elements (Onodera, et al., 2005). These primary single stranded transcripts are converted into double stranded RNAs by RNA-dependent RNA polymerases.

1.4.1 RNA-dependent RNA polymerases in plants

Plant genomes encode RNA dependent RNA polymerases termed RDR. RNA dependent RNA polymerases of viral origin are termed RdRp (Wassenegger, Krczal, 2006). Sequence alignments of RDRs and RdRps show substantial differences which could be attributed to two independent evolutionary appearances of these polymerases or the high viral mutation rate encrypts homology. The latter explanation is supported by studies focusing on the catalytic domain of RDRs, RdRps and DNA-dependent RNA polymerases. As all groups share a common core motif, these enzymes are thought to be of ancient origin (Iyer, et al., 2006).

Arabidopsis encodes six RDRs ranging from approximately 100 to 130 kDa in size. RDR1 was discovered in Chinese cabbage upon viral infection (Astier-Manifacier, Cornuet, 1971) and its polymerase activity studied *in vitro* (Schiebel, et al., 1993). RDR1's *in vivo* role remains ambiguous as does RDR3a, RDR3b and RDR3c.

RDR2 and RDR6 were studied in more detail and are linked to some RNA silencing pathways in plants. RDR2 is involved in *de novo* methylation of genomic DNA (Xie, et al., 2004; Zilberman, et al., 2003) and RDR6 is linked to post-transcriptional silencing (Gasciolli, et al., 2005b). Further details on these two RDRs are given in the sections below.

An interesting concept of transitive silencing involves RDR6. Here, a messenger RNA is cleaved by a micro RNA. The resulting cleaved messenger RNA is amplified in a primer independent manner and cleaved by DCL4, resulting in secondary trans-acting silencing RNAs (ta-siRNAs). The ta-siRNAs are incorporated into RISC thereby amplifying the effect of the primary miRNA. The newly synthesized ta-siRNAs are also not complementary to the original miRNA, rather they are generated every 21 nts 5' and 3' of the original cleavage site (Petersen, Albrechtsen, 2005). The section 'Small RNAs from DCL4 cleavage' discusses the concept of transitive silencing in more detail.

1.4.2 Dicer-like proteins in plants

Besides RDRs, another important ribonuclease termed Dicer in animals is involved in small RNA biogenesis. Dicer and dicer-like enzymes belong to the RNase III family of double stranded RNA specific endoribonucleases, an ancient class of enzymes found in all domains of life. In higher eukaryotes, Dicer and dicer-like proteins are multi-domain proteins containing an RNase III domain, a double stranded RNA binding domain and range in size from 135 to 200 kDa (Shi, et al., 2006). In animals, Dicer liberates the pre-miRNA from its terminal loop whereas in

plants Dicer-like enzymes (DCL) take over the function of both Drosha and Dicer. In animals, endogenous small RNAs are used for house-keeping functions and development of tissues, whereas plants used the ability of DCL's to cleave double stranded RNA. As will be discussed below, most plant viruses are single stranded RNA viruses, which propagate *via* a double stranded RNA stage, and plants evolved to cut them efficiently. If plants are always able to cut viral RNA efficiently, viruses would be unable to propagate in plant cells, e.g. the coat protein messenger RNA was constantly degraded thus no virus particles were made, thus no spread of the virus. Viruses fought back and acquired proteins that would undermine the plants RNA silencing machinery. These proteins are called RNA silencing suppressors and they act in various ways, as will be discussed below. One can imagine this being a never ending battle between the DCL's in plants trying to stop viral infection and viruses stopping the plant's DCL's (Itaya, et al., 2007).

Arabidopsis was annotated to contain four dicers (Finnegan, et al., 2003; Schauer, et al., 2002). Margis and colleagues designed DCL2 and DCL3 specific primers and PCR amplified reverse transcribed messenger RNA. The research group found two alternatively spliced transcripts of DCL2 which differ in the RNA helicase region potentially altering the specificity of DCL2. Investigating other available monocot plant genomes, a paralog of DCL3 termed DCL3b was observed. DCL3b in *Oryza sativa* has a 57 % amino acid identity to DCL3a and its double stranded RNA binding domain diverged significantly from all other Dicer-like proteins. *Oryza sativa* encodes six dicer genes (*dcl1*, *dcl2a*, *dcl2b*, *dcl3a*, *dcl3b* and *dcl4*) (Margis, et al., 2006). Currently, there is no genetic data available deciphering the role of alternatively spliced dicer-like ribonucleases (in the case of DCL2) or the two DCL3 paralogs. Thus, for the remainder of the introduction I will focus on the main four dicers.

1.4.2.1 Small RNAs from DCL1 cleavage

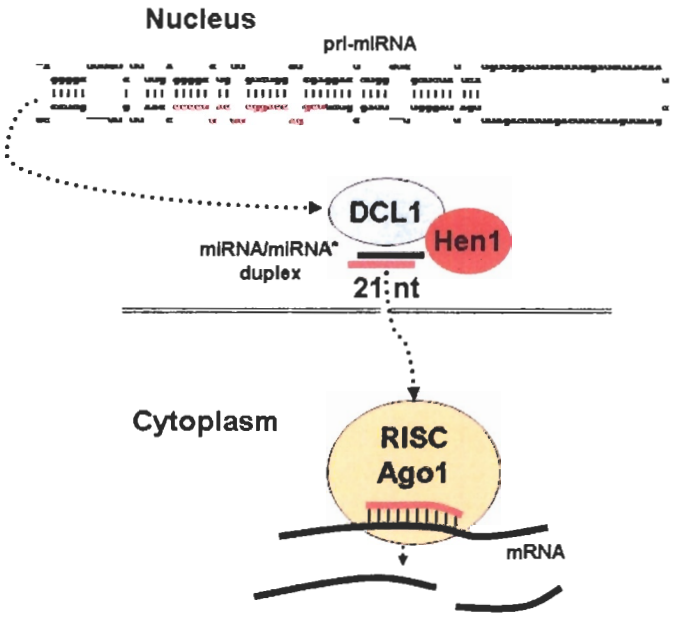
In plants, primary miRNAs are also transcribed by RNA polymerase II and fold into characteristic hairpin structures (Kurihara, Watanabe, 2004; Lee, et al., 2004). The hairpins only require one ribonuclease to obtain the miRNA/miRNA* duplex which is termed Dicer-like (DCL). Most plants encode at least four distinct dicer-like proteins: DCL1, DCL2, DCL3 and DCL4. Each one was found to be predominantly responsible to cleave certain types and sizes of small RNA. DCL1 is located in the nucleus and essential for generating the miRNA/miRNA* duplex (Hiraguri, et al., 2005; Park, et al., 2002).

Associated with DCL1 and miRNA biogenesis is HEN1 (HUA ENHANCER1). *hen1* mutants showed multifaceted developmental defects in *Arabidopsis* (Boutet, et al., 2003; Park, et

al., 2002). HEN1 has a S-Adenosyl Methionine (SAM) binding pocket, methylates miRNA/miRNA* duplexes *in vitro* (Yu, et al., 2005) and does not require guide RNAs to methylate its targets (Omer, et al., 2002). In *Arabidopsis* this 3' terminal methylation prevents 3' uridylation (Li, et al., 2005). Figure 1.3 shows the micro RNA pathway in plants and how miRNAs affect stability of mRNAs.

Figure 1.3: Micro RNA pathways in plants.

Micro RNAs (miRNAs) are transcribed from genomic DNA. The single stranded RNA folds into a characteristic hairpin termed primary miRNA. In plants, only one protein cleaves the pri-miRNA, which is referred to as Dicer-like protein (DCL1), resulting in the miRNA/miRNA* duplex that is immediately methylated by HEN1. Through an uncharacterized mechanism, the miRNA/miRNA* duplex is exported into the cytoplasm. In both animals and plants one strand of the duplex is incorporated into Ago1 containing RISC, which causes translational inhibition of target mRNAs in animals and cleavage of target mRNAs in plants.



Micro RNA mediated cleavage of mRNA.

Micro RNAs both in animals and plants are incorporated into RISC. In three quarters of all cases, the strand with the lower 5' terminus melting temperature is loaded into the RISC (Khvorova, et al., 2003; Schwarz, et al., 2003). A key component of RISC is ARGONAUTE1 (AGO1), a protein part of an ancient protein class that contains both RNA binding and RNA cleavage domains (Boutet, et al., 2003). Argonaute proteins are comprised of a PIWI and PAZ domain and are found in all domains of life (Archaea, Bacteria and Eukaryotes) (Parker, Barford, 2006). In plants, argonaute proteins are 115 to 140 kDa in size (Swiss-Prot). Unlike miRNA RISC in animals that causes translational inhibition, the miRNA RISC complex in plants causes mRNA degradation. The target mRNA is recognized by a near perfect 21 nt long complementary

match between the miRNA and the mRNA (Llave, et al., 2002).

1.4.2.2 Small RNAs from DCL3 cleavage

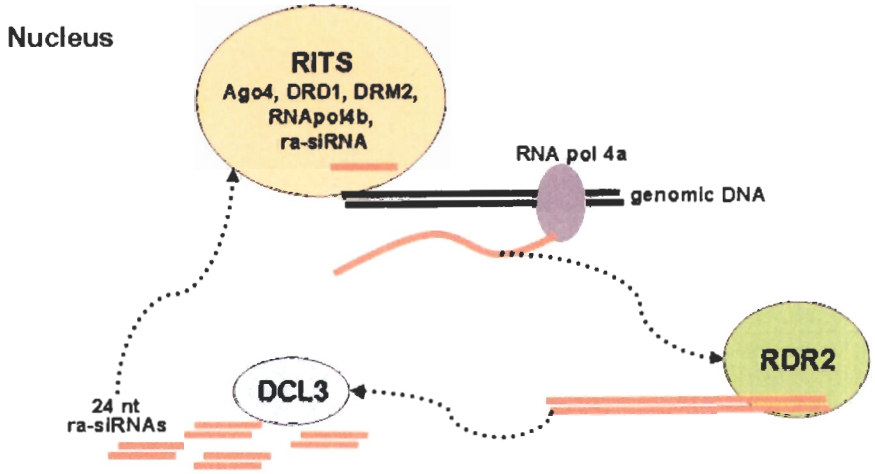
RNA polymerase IVa transcribes repetitive genomic regions that are converted by RDR2 into double stranded RNA (Pontes, et al., 2006). These double stranded transcripts from repeat associated genomic loci are cleaved by DCL3 into short 22 nt long silencing RNA duplexes with a 2 nt overhang on each side. As these silencing RNAs are repeat associated, they are termed in the literature as repeat-associated silencing RNA, short repeat associated silencing RNAs (ra-siRNA). The type of DNA methylation guided by 24 nt RNAs is termed RNA-dependent or RNA-directed DNA methylation (RdDM). When RdDM targets a promoter region it is called transcriptional gene silencing (TGS) (Mette, et al., 2001).

RDR2 and DCL3 were first linked to DNA methylation patterns when Xie and colleagues studied AtSN1 (the *Arabidopsis* thaliana short interspersed element 1). *rdr2* and *dcl3* were identified in lowering the asymmetric CpHpH methylation by about 75 % and 50 %, respectively. The levels of histone methylation at H3K9 decreased by ~ 50 % in both *rdr2* and *dcl3* and increased by ~ 50 % for H3K4 methylation. Because of the low DNA and histone methylation patterns, transcription levels of AtSN1 were 3 to 9 times higher in *dcl3* and *rdr2* mutants, respectively, when compared to wild-type *Arabidopsis* (Xie, et al., 2004).

The exact mechanism by which the RNA-induced initiation of transcriptional gene silencing (RITS) acts on genomic loci is under active investigation. Pontes and colleagues demonstrated that RDR2, DCL3, AGO4 and NRPD1b, the largest subunit of RNA polymerase IVb, co-localize with ra-siRNAs within the nucleolus (Pontes, et al., 2006). They further postulate that the proteins found in the nucleolus form the processing centre in which the single stranded RNA polymerase IVa transcripts are converted into double stranded RNA by RDR2. DCL3 is responsible for cleaving the now double stranded RNA into 22 bp long RNA helixes with a 2 nt overhang on each side. One strand of the duplex is loaded into the AGO4-containing RITS (Zilberman, et al., 2004). The loaded RITS relocates to the nucleoplasm where it associates with RNA polymerase IVb to guide the chromatin remodeler DRD1 as well as *de novo* cytosine methyltransferase DRM2 to the genomic repeat regions. The exact mechanism or role each protein plays in this process is still unknown. I propose, as RNA polymerase VIb is still associated with the loaded RITS the polymerase melts the double stranded genomic DNA and once there is a match between the ra-siRNA in RITS and the genomic DNA the *de novo* cytosine methyltransferase DRM2 is recruited by RITS to methylate cytosines as illustrated in Figure 1.4.

Figure 1.4: Chromatin silencing guided by small RNAs.

RNA polymerase IVa (RNA pol 4a) transcribes RNA from repeat associated genomic regions. An endogenous RNA-dependent RNA polymerase RDR2 converts the single stranded transcripts into double stranded RNA. The double stranded RNA is cleaved by Dicer-like 3 (DCL3) into 24 nt double stranded RNA. A single strand 24 nts in length (ra-siRNA) is incorporated into RITS. By an enigmatic mechanism, RITS guides transcriptional silencing of repeat associated genomic regions.



1.4.2.3 Small RNAs from DCL4 cleavage

DCL4 has multiple roles in plant cells: typically DCL4 generates 21 nt RNA associated with trans-acting silencing RNAs and silencing RNAs for RNA interference. Upon infection, DCL4 also engages in cleaving invading RNAs (Blevins, et al., 2006b).

RNA interference is the most effective and commonly used technique to silence plant genes post-transcriptionally. Typically, an inverted repeat corresponding to the target sequence is introduced into the plant and the double stranded RNA is cleaved by a dicer-like protein. One strand of the silencing RNA duplex leads RISC to the target mRNA, which is cleaved and subsequently degraded. In the literature this process is historically referred to as inverted-repeat post-transcriptional gene silencing (IR-PTGS) or, more recently, RNA interference in plants (RNAi) (Beclin, et al., 2002).

Which of the dicer-like proteins in plants cleaves these inverted repeats is still not clear as no mutant defective in the inverted-repeat-pathway has been isolated. Two lines of evidence support the idea of DCL4 responsible for the RNAi pathway: first the size of RNA DCL4 cleaves is 21 nts – RNAi associated small RNAs are 21 nts. Second, DCL4 was required for accumulation of siRNA derived from a moderately expressed inverted repeat (Dunoyer, et al., 2006).

Another RNA silencing pathway in which DCL4 plays an important role is in the

biogenesis of trans-acting silencing RNAs (ta-siRNA) which are important for general house-keeping duties in plant development, e.g. for juvenile-to-adult phase transitions (Vaucheret, 2005). Ta-siRNA are generated from miRNA cleaved transcripts that are converted by RDR6 into double stranded RNA substrates for DCL4 (Gasciolli, et al., 2005a). DCL4 mutants lack ta-siRNAs and accumulate ta-siRNA target mRNAs (Xie, et al., 2005).

It is interesting to note that Dunoyer and colleagues established that plant RNAi and the generation of ta-siRNA are two distinct processes requiring DCL4, as a double mutant of DCL4 and RDR6 showed normal levels of RNAi related siRNAs but a lack of ta-siRNAs (Dunoyer, et al., 2006).

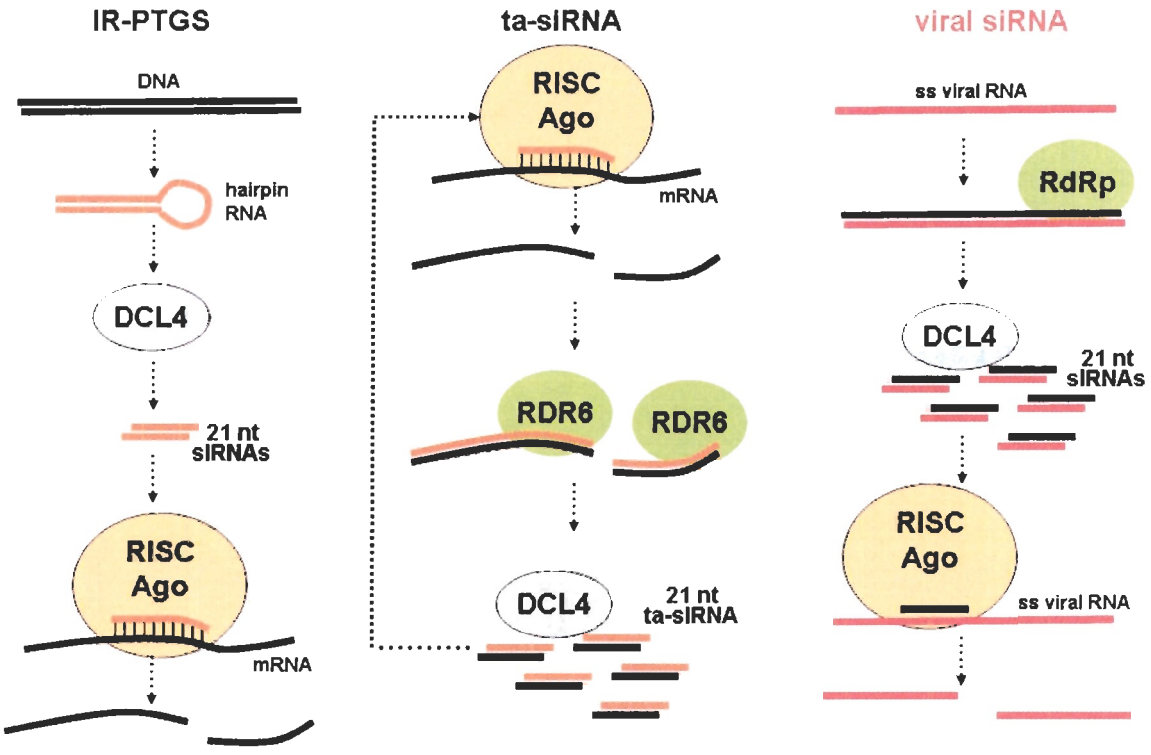
During viral infection, DCL4 targets viral dsRNA producing siRNAs that in turn target the viral RNA. Blevins and colleagues studied the cytoplasmic tobamovirus ORMV and found only viral siRNAs generated by DCL4. The same research group did not find any viral siRNA accumulation dependence on host RNA-dependent RNA polymerases (RDR1, -2, -3a, -3b, -3c and RDR6). However there was a link between virus induced gene silencing, RDR6 and DCL4 (Blevins, et al., 2006a).

Blevins *et al.*'s research would suggest that DCL4 resides in the cytoplasm. In contrast Hiraguri and colleagues located DCL4 in the nucleus using GFP-tagged DCL4 (Hiraguri, et al., 2005). The latter experiment was performed in non-infected plants. It would be interesting to observe the localization of GFP-tagged DCL4 in tobamovirus challenged plants.

Figure 1.5 shows a summary of RNA silencing pathways involving DCL4.

Figure 1.5: Multiple RNA silencing pathways involving DCL4.

Inverted-repeat post-transcriptional gene silencing (IR-PTGS), also known as plant RNAi, is the process by which introduced DNA is transcribed into RNA which folds into a hairpin. DCL4 recognizes the double stranded RNA and generates 21 nt double stranded siRNAs. A single strand of the siRNA duplex is incorporated into RISC, which is guided to the siRNA's target mRNA. Trans-acting siRNAs amplify the effects of miRNA-guided mRNA cleavage. The cleaved mRNA serves as template for RNA-dependent RNA polymerase (RDR6) amplification. The now double stranded RNA is recognized by DCL4, cleaved into 21 nt siRNAs which target the original mRNA. During a viral infection, a viral RNA-dependent RNA polymerase (RdRp) converts the single stranded viral genome into double stranded RNA. This double stranded RNA is digested by DCL4 into siRNAs that target the intruding viral RNA.



1.4.2.4 Small RNAs from DCL2 cleavage

Like DCL4, DCL2 also plays a dual role in plant cells. In uninfected plants, DCL2 is involved in natural anti-sense silencing RNAs. Upon infection, DCL2 also engages in cleaving invading RNAs (Deleris, et al., 2006).

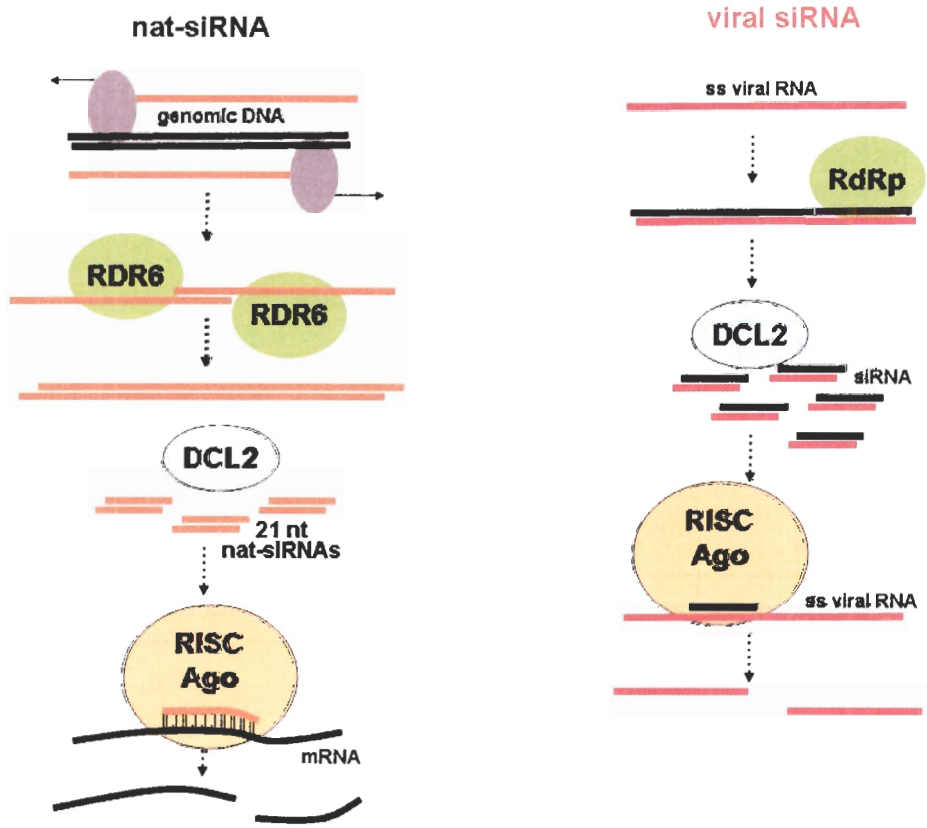
Natural anti-sense silencing RNAs (nat-siRNA) are derived from genomic regions where two neighboring genes are encoded on opposite DNA strands (termed *cis*-antisense gene pairs). Once these genes are transcribed, starting from opposite sites, the overlapping transcript gives rise to nat-siRNA. The biogenesis of nat-siRNAs involves a number of enzymes, including DCL2 and RDR6. Nat-siRNAs are 24 nts in length and guide cleavage of one of the parent transcripts,

depending on which nat-siRNA was incorporated into RISC. Interestingly, the effect of nat-siRNAs is amplified as secondary phased 21 nt siRNAs are generated involving DCL1 (Borsani, et al., 2005). As *cis*-antisene gene pairs are common in plant and other genomes, nat-siRNA might exemplify wide spread gene regulation. Nat-siRNAs also contribute to plant stress adaptaion (Vaucheret, 2006).

Upon viral infection *Arabidopsis'* DCL2 generates ~22 nt RNA, which are incorporated in to RISC and conduct the cleavage of primary viral RNA. Borsani and colleagues determined that DCL4 and DCL2 have redundant activities but act hierarchicaly during viral infection (Borsani, et al., 2005). How the two different splice forms of DCL2 influence these pathways is currently unknown. Figure 1.6 illustrates RNA silencing pathways involving DCL2.

Figure 1.6: RNA silencing pathways involving DCL2.

Natural anti-sense silencing RNAs (nat-siRNA) are derived from genomic regions where two neighbouring genes are encoded on opposite DNA strands (termed *cis*-antisense gene pairs). The overlapping transcripts are converted into double stranded RNA by RDR6 and cleaved into 21 nt long nat-siRNAs by DCL2. Another role for DCL2 is during viral infection, where, like DCL4, DCL2 will cleave viral double stranded RNA. In the viral siRNA generation, DCL4 and DCL2 overlap, but show a hierarchical order: DCL4, then DCL2.



1.5 Comparison of plant and animal immune systems

Plants and animals share some aspects of non-specific immune defence against invading parasites, bacteria or viruses. Both use external tissue to protect the inner cells and organs. In the case of plants, the outer bark consists of alternating layers of periderms and rhytidome. In animals, the outer layer is comprised of epithelial tissues. In case of wounding, plants respond rapidly by depolarization of cell membranes, release of host or pathogen cell wall fragments. Long term immune responses include complex biosynthetic reactions, formation of boundary tissues and activation of the RNA silencing machinery. As discussed earlier, a key component of the plant's immune system are multiple layers of RNA silencing machinery and partially overlapping DCL activities. Thus far, six DCL's have been annotated in rice (Margis, et al., 2006) and it will be interesting to explore the evolution of DCL's in other plant genomes once they become available.

Animals on the other hand do not use the RNA silencing machinery as their main immune system. Animals will respond to invading parasites, bacteria and viruses *via* non-specific immune defense mechanisms including an inflammatory response, accumulation of white blood cells and increase of body temperature to prevent the growth of bacteria. Besides the non-specific immune response, there is a specific immune response. There, B-cells produce specific antibodies against the invading pathogen, called antigens. B-cells are also matured in the thymus where they mature to T-cells, which also fight invading germs (Janeway, et al., 2004).

1.6 Plant viruses

Viruses are infectious agents that rely on the host cells to supply components for their own replication. There are a multitude of viruses infecting different plant species and their genetic information is encoded mostly, but not solely, by single-stranded RNA. Other forms of encoding hereditary information are by means of double-stranded RNA (e.g. wound tumor virus), single-stranded DNA (e.g. gemini virus) or double-stranded DNA (e.g. cauliflower mosaic virus).

The hereditary nucleic acid is typically encapsidated by a protective protein coat or shell. These proteins are often referred to as capsid or coat proteins. The infectious mature virus particle is often referred to as virion and can adopt various shapes, e.g. spherical or rod-shaped.

Transmission of plant viruses varies, e.g. entering plants *via* wounds or insects act as an intermediate invertebrate host of a virus. If the plant virus is comprised of multiple components, e.g. a tripartite genome, all three components have to enter the host cell simultaneously. Once a

host cell has been infected, the virus takes advantage of the host cell's energy metabolism to multiply within the cell. The infection is usually recognized by a mosaic-like leaf pattern (chlorosis) or stunted growth (necrosis) (Walkey, 1991).

Most viruses encode viral RNA-dependent RNA polymerases (RdRps) to ensure amplification of their genome with a high error rate (10^{-4}) to rapidly evolve viral genomes undermining the host's anti-viral defense system (Crotty, et al., 2001). RdRps have also been linked to acquiring new genes from viruses or the host and recombination of viral genomes through strand switching during replication (Hill, et al., 1997).

1.6.1 The cucumber mosaic virus

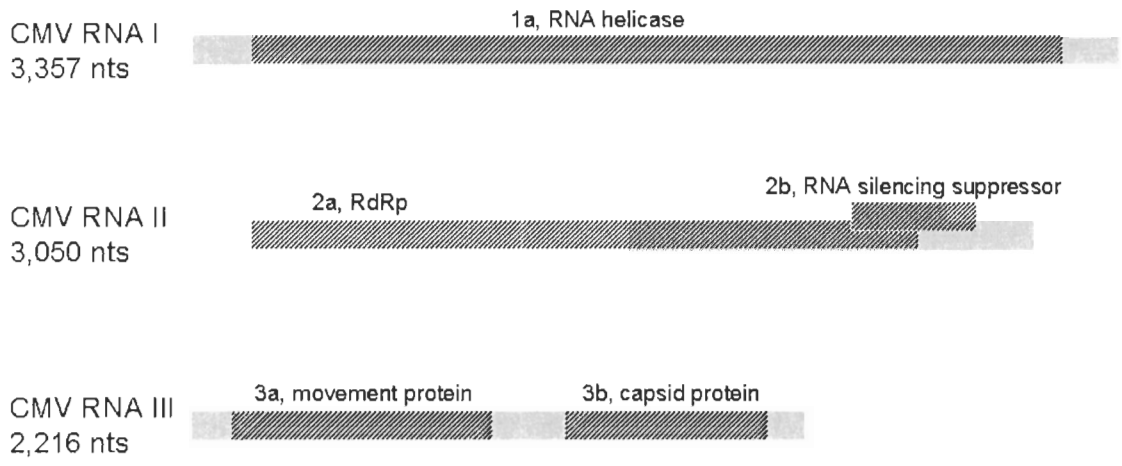
One such plant virus is the cucumber mosaic *cucomovirus* (ICTV 10.0.4.0.001) naturally infecting cucumber, tomato and spinach. The *cucomovirus* is transmitted by an insect or by mechanical inoculation and can infect a wide variety of hosts, inducing tobacco (VIDE Database, http://image.fs.uidaho.edu/vide/).

In the early 1970's a lethal necrosis in tomato plants afflicted the French Alsace (Marrou, 1973; Marrou, et al., 1973). Waterworth and colleagues discovered that the cucumber mosaic virus (CMV) and its satellite RNA caused these symptoms in tomato (Waterworth, et al., 1978; Sleat, et al., 1994). In tobacco, depending on the strain and satellite, CMV causes yellow or white chlorosis due to loss of chlorophyll or chlorophyll and xanthrophyll, respectively (Sleat, Palukaitis, 1990; Zhang, et al., 1994).

The helper virus CMV is comprised of a tripartite positive single-stranded RNA genome (Lot, et al., 1974). CMV RNA I is 3,357 nts in length and encodes a putative RNA helicase (Refseq: NC_002034). CMV RNA II is 3,050 nts in length and encodes two proteins: an RNA dependent RNA polymerase and the 2b protein, a putative RNA silencing suppressor (Refseq: NC_002035). CMV RNA III is 2,216 nts long and encodes the 3a protein, a putative movement protein, and the caspid protein required for packaging all infectious RNA molecules (Refseq: NC_001440). The CMV does not require a DNA stage for amplification. Figure 1.7 shows the CMV tripartite genome and the protein(s) each RNA encodes.

Figure 1.7: Tripartite CMV protein coding genome.

The tripartite CMV single stranded RNA genome encodes a RNA helicase (1a), a viral RNA dependent RNA polymerase (2a), a putative RNA silencing suppressor (2b), a protein essential for viral movement (3a) and a capsid protein for packaging the viral agents (3b).



1.6.2 Satellites of cucumber mosaic virus

When analyzing plants infected with CMV, researchers often found a fourth 334 to 391 nts long parasitic RNA molecule termed Satellite RNA that requires CMV as helper virus for its own replication (Gould, et al., 1978). CMV's satellite RNA parasitic nature was demonstrated in field studies in Spain. Aranda and colleagues sampled over eight years six different sites and found various isolates of the helper virus CMV but only one form of the satellite RNA (Aranda, et al., 1997).

A search for 'cucumber mosaic virus satellite RNA' in NCBI's GeneBank reveals 126 sequenced CMV satellite RNAs (search performed Feb. 15, 2007). Closer examination of the CMV satellites revealed that all isolates share a common 5' and 3' terminus, with 10 or 8 nt sequence identity, respectively. Hidaka and colleagues established that certain satellites could be translated into short poly-peptides *in vitro* (Hidaka, et al., 1990). However, none of these products were ever detected *in vivo*. Furthermore, when comparing different satellite isolates, these peptides were not conserved and a third position mutation bias, reminiscent of protein coding regions, was not observed (Masuta, Takanami, 1989).

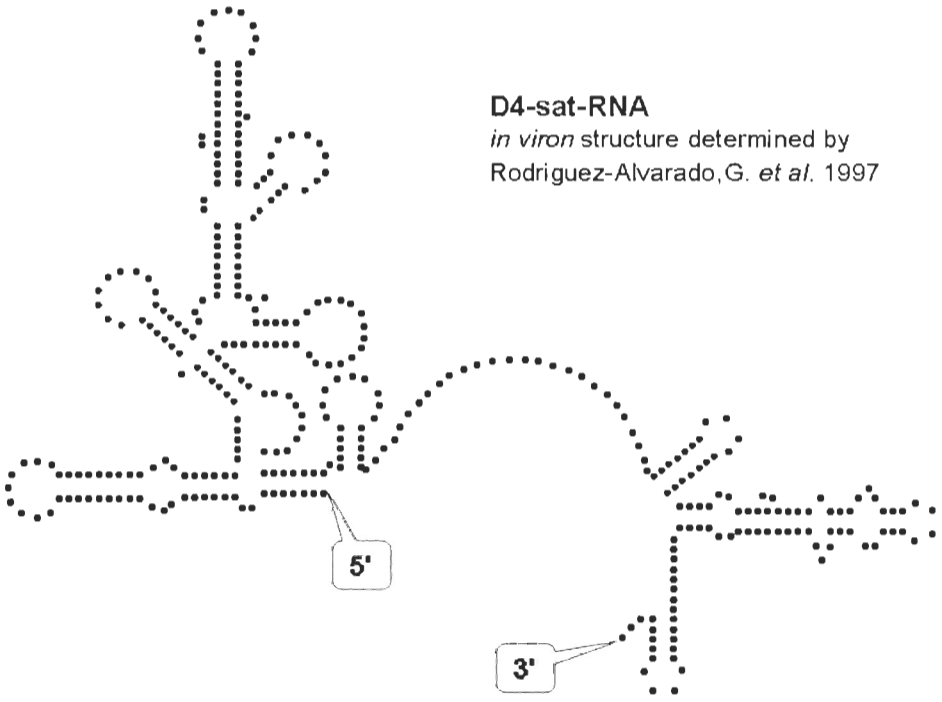
If the single stranded satellite RNA does not encode proteins then the secondary and tertiary structure of this molecule must be important for its biological function. Gordon and

Symons studied a 336 Satellite isolate they termed Q-sat-RNA. The Q-sat-RNA was probed with nuclease S1 and RNase T1 and found it to be highly structured with 52 % of all residues base paired and seven major stems forming (Gordon, Symons, 1983). Later, closely related B2- and B3-sat-RNA were found not to fit the proposed model that well, and different cleavage patterns were observed in regions of identity (Hidaka, et al., 1988).

In 1997 Rodrigues-Alvarado and Roossinck did a comprehensive study on another isolate named D4-sat-RNA by combining *in silico*, *in vitro*, *in vivo* and *in virions* modification studies of adenines and cytosines using dimethyl sulfate (DMS). In essence, their *in virion* model shows a highly structured 5' terminus, an alternatively folded or unfolded central region and a highly structured 3' terminus (Rodriguez-Alvarado, Roossinck, 1997). Figure 1.8 shows the schematic representation of the D4-sat-RNA secondary structure as determined by Rodriguez-Alvarado and Roossinck.

Figure 1.8: Schemata of D4-sat-RNA.

The (•) represent the nucleotides of the determined *in viron* secondary RNA structure of cucumber mosaic virus D4-sat-RNA (Rodriguez-Alvarado, Roossinck, 1997). Clearly visible is the highly structured 5' part of the D4-sat-RNA, the unstructured centre part and the structured 3' terminus. The 5' and 3' termini are labelled.



Masuta and colleagues further characterized CMV's Y-sat-RNA. Expressing the Y-sat-RNA alone in transgenic tobacco plants did not induce yellow chlorosis (Masuta, Takanami, 1989). Further research revealed two distinct regions in the satellite of CMV causing necrosis in tomato and yellowing in tobacco (Devic, et al., 1989; Devic, et al., 1990). In the case of Y-Sat / CMV infecting tobacco, Jaegle and colleagues noticed that not all CMV satellite strains cause the same intensity of yellowing in tobacco. While studying mutant forms of the satellites, Jaegle and colleagues determined two nucleotides at position 185/186 in the Y-Sat sequence to cause yellowing in tobacco plants. If these two nucleotide positions were altered, the yellowing in tobacco was eliminated, but not the necrosis in tomato (Jaegle, et al., 1990).

Masuta and colleagues observed that most wild type tobacco species are susceptible to yellowing by Y-Sat and CMV infection but found two species that did not show the typical yellowing: *Nicotiana clevelandii* and *N. debneyi*, which show a green mosaic phenotype instead (Masuta, et al., 1993). *N. bigelovii* shows the typical yellowing phenotype upon infection with Y-Sat / CMV. Masuta and colleagues crossed a non-yellowing (*N. clevelandii*) and a yellowing species (*N. bigelovii*) which resulted in a F1 generation. Crossing the F1 plants resulted in a F2 generation, which upon infection of Y-Sat / CMV resulted in yellow mosaic plants, intermediate plants and green mosaic plants with a segregation of 1:2:1. Crossing the F2 plants with the green mosaic phenotypic parents (*N. clevelandii*) resulted in intermediate and green phenotypic plants in a ratio of 0:1:1 (yellow : intermediate : green). Crossing the F2 with the yellow mosaic plants (*N. bigelovii*) resulted in plants showing a ratio of 1:1:0 (yellow : intermediate : green) after infection with Y-Sat / CMV. From these observations, Masuta and colleagues concluded that Y-Sat / CMV target a single incompletely dominant gene in tobacco.

1.7 Viral infection: plant defense and counter defense

Viral infections are common to all organisms. Various strategies have been employed to combat infection: from digestion of foreign DNA in the case of intruding phage to immune systems with T and B memory cells that help memorize previous intruders. Plant developed a different kind of anti-viral defense mechanism. As previously discussed, dicer degrades double stranded RNA (from either single or double stranded transcripts). Most plant viruses are comprised of RNA genomes. There are some DNA viruses which have to ultimately go through an RNA stage in order for the encoded message to be translated. Thus, plants evolved to degrade viral RNA into short 21 to 24 nt long RNA in order to direct sequence specific degradation of viral messenger RNA.

1.7.1 RNA silencing suppressors

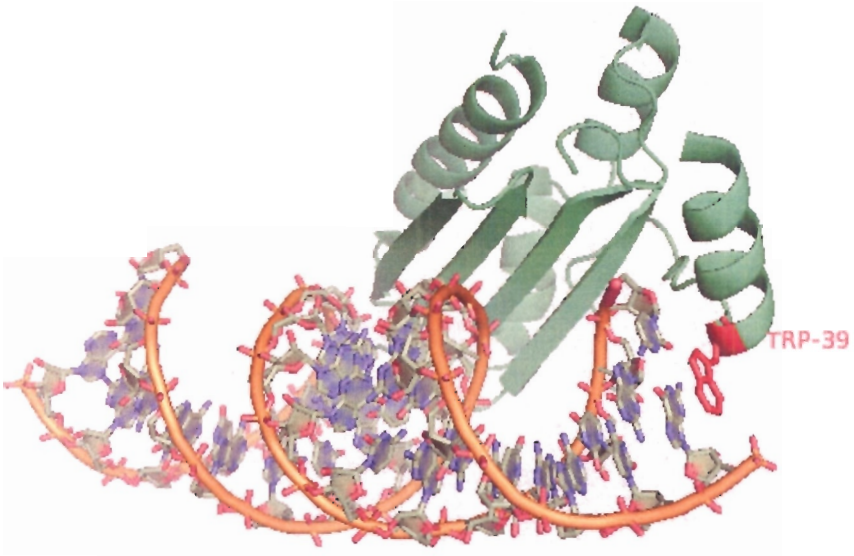
One very successful strategy of overcoming the plant's anti-viral defence system is to block any of the major players involved. Tombusviruses encode a RNA silencing suppressor called p19, which binds double stranded RNA helixes 19 base pairs in length with great affinity (K_d = 170 pM) and specificity. As mentioned before, Dicer-like cleavage of double stranded precursor RNA results in 19 base pair helixes with 2 nt 3' overhang on each side. p19 is specific for RNA, not DNA, and is size specific for 19 base pair helixes. It tolerates helixes of 17 to 22 base pairs but does not bind to longer or shorter constructs *in vitro*. Figure 1.9 shows the crystal structure of *Carnation ring spot virus* p19 bound to silencing RNA duplex at 1.85 Å (Ye, et al., 2003a). It should be noted that the assumed biological molecule is a homo-dimer – Figure 1.9 only shows a single p19 unit. The length of double helix p19 can accommodate is dictated by tryptophan, shown as side chain with an electron density cloud. The indole ring of tryptophan serves as a bracket to measure the length of the RNA double helix. This and other stacking interactions as well as hydrogen bond formation between the amino acid side chains and the phosphate backbone build this strong RNA-protein interaction.

The RNA binding domain of p19 is almost identical to that of ribosomal protein L1 binding to ribosomal stem-loop structure (Vargason, et al., 2003). Tombusviruses might have acquired this domain during RdRp dependent strand switching during viral replication.

The RNA silencing suppressor p19 clearly acts by binding to 19 bp RNA helixes with a 3' 2 nt overhang on each side (19hr2). In my opinion, this can only be the first step in the RNA suppressor cycle of p19. If p19's sole role is soaking up cleavage products of Dicer-like proteins, the plant cell would burst open with p19-siRNA complexes, a loose-loose situation for plant cell and virus. To ensure turnover, a feasible hypothesis could be that p19 binds the siRNA double helix with 2 nt overhang and presents this very overhang to cellular nucleases. p19 releases the cleaved blunt end 19 bp RNA helix, which is now no longer recognized by RISC as siRNA duplex, and binds to the next cleavage product of Dicer-like proteins. The evidence for this hypothesis is currently spotty, but blunt end 19 bp RNA helixes in *Drosophila* embryo lysate do not cause RNAi (Elbashir, et al., 2001). An experiment in support of the hypothesis could be to measure the accumulation of internally radiolabelled 19 nt long small RNA that were originally in a 21 nt duplex in cell free extract from infected plants expressing p19.

Figure 1.9: RNA silencing suppressor p19.

The suppressor binds double stranded RNA 21 nt in length. Shown here is the solved X-ray structure with the protein depicted as a green ribbon cartoon. The assumed biological molecule is a homo-dimer. p19 and the RNA double helix interact *via* multiple non-specific amino-acid – phosphate backbone interactions. Tryptophan 39, labelled in red, dictates the length of double stranded RNA helix p19 can accommodate. (Reference: Figure generated by Ebhardt using PyMOL; pdb id: 1R9F (Ye, et al., 2003a)).



1.7.2 Helper component proteinase

Potyviruses encode one long polypeptide that is cleaved by viral proteinases to yield the different functional enzymes, one of which is the Helper Component Proteinase (HC-Pro). It was initially isolated from tobacco vein mottling virus (TVMV) and potato virus Y (PVY) infected tobacco plants. The molecular weight of a unit of the isolated Helper component (HC) was approximately 55 kDa. Berger and colleagues identified HC as an essential enzyme for viral host-to-host transmission by aphids (Berger, et al., 1989). Further biochemical studies revealed that HC-Pro is capable of cleaving its own C-terminus for release from the precursor polypeptide (Carrington, et al., 1989a; Carrington, et al., 1989b). Cronin and colleagues discovered in a mutant screen, that HC-Pro was also involved in viral long-distant movement in the host plant's phloem (Cronin, et al., 1995). Atreya and Pirone showed that HC-Pro was essential for aphid transmission and regulation of virulence in potyvirus infected plants (Atreya, Pirone, 1993).

HC-Pro is the second enzyme in the long potyvirus peptide. As seen in Figure 1.10, the P1/HC-Pro/P3-fragment contains at least two proteases: P1 and HC-Pro. It was observed on many

occasions that P1 enhances the RNA silencing effects of HC-Pro (Anandalakshmi, et al., 1998; Kasschau, Carrington, 1998; Pruss, et al., 1997). In light of this finding, the materials and methods section of some HC-Pro studies are ambiguous in terms of using P1/HC-Pro or only HC-Pro.

Figure 1.10: P1 – HC-Pro – P3 potyvirus polypeptide.

The N-terminus of the potyvirus polypeptide is comprised of at least two proteases: P1 and HC-Pro, which self-cleave to release themselves as functional units. The four dots at the C-terminus represent the continuation of the polypeptide.

Most of the initial experiments were done using plants infected with potyviruses. If HC-Pro is taken out of this context, there are conflicting results regarding disease symptoms. For example when P1/HC-Pro was expressed in transgenic tobacco plants, P1/HC-Pro amplified the disease symptoms of CMV and Potato Virus X (Pruss, et al., 1997). In contrast, Wang and colleagues showed that CMV/Y-Sat symptoms were eliminated in transgenic tobacco plants expressing P1/HC-Pro (Wang, et al., 2004).

Brigneti and colleagues expressed a green fluorescent protein (GFP) transgene in tobacco plants. Typically, the GFP transgene is silenced post-transcriptionally. Upon infection of these plants with Potato Virus Y (PVY) and the Cucumber Mosaic Virus (CMV), fluorescence levels went up. This phenomenon was traced to the ability of PVY's HC-Pro and CMV's 2b to inhibit the post-transcriptional gene silencing of the host plant (Brigneti, et al., 1998). Two other labs identified HC-Pro as post-transcriptional gene silencing suppressor (Kasschau, Carrington, 1998) as well as the transgene – and virus-induced gene silencing (Anandalakshmi, et al., 1998).

The mechanism by which P1/HC-Pro acts as RNA silencing suppressor is debated in the literature. The initial experiments purifying HC-Pro was done using an oligo-dT column, a tribute to HC-Pro's ability to bind nucleic acids (Pirone, Thornbury, 1984). Maia and Bernardi show *in vitro* that HC-Pro binds to ssRNA, DNA and dsRNA (Maia, Bernardi, 1996). Lakatos and colleagues demonstrate that siRNA duplexes are bound by recombinant HC-Pro (Lakatos, et al., 2006). The conclusion of this thesis will discuss some of these publications in more detail.

1.7.3 Research prior to this study

Wang and colleagues hypothesized that RNA silencing is involved in viroid and viral satellite pathogenicity (Wang, et al., 2004). To support this hypothesis with experimental data, Wang and colleagues expressed the viral RNA silencing suppressor P1/HC-Pro as transgene in tobacco plants. Transgenic and non-transgenic plants were then challenged with CMV/Y-Sat where Y-Sat infection causes yellowing in tobacco. Following infection, non-transgenic plants developed yellowing symptoms after 17 days post induction (dpi) and severe systemic chlorosis after 25 dpi. None of the 21 plants expressing P1/HC-Pro showed any of these symptoms, supporting the hypothesis that RNA silencing is involved in viral satellite pathogenicity. Northern hybridization experiments using a Y-Satellite probe showed high levels of Y-Sat specific siRNAs in infected plants. In P1/HC-Pro(+) plants, the Y-Sat specific siRNA levels were even higher than in P1/HC-Pro(–) plants.

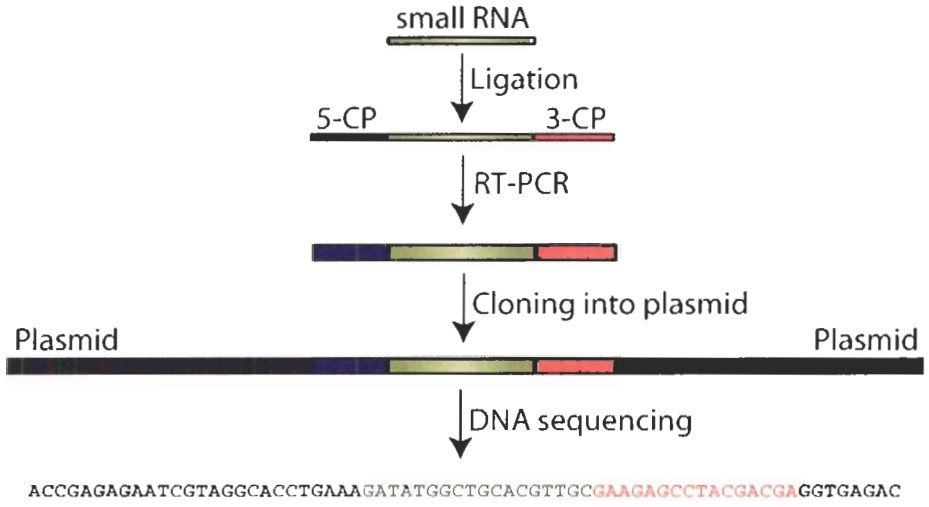
These results confirm that P1/HC-Pro does not interfere with the cleavage of siRNAs from double stranded transcripts and supported the hypothesis that RNA silencing is involved in viral satellite pathogenicity. The question how the RNA silencing suppressor P1/HC-Pro functions was not addressed. It also presented a paradigm: in the Northern hybridization experiments of infected plants, levels of Y-Sat derived siRNAs were higher in P1/HC-Pro(+) CMV/Y-Sat(+) than in P1/HC-Pro(–) CMV/Y-Sat(+) plants. But, the P1/HC-Pro(+) plants did not show any Y-Sat related disease symptoms. Therefore, my research addressed the question of how P1/HC-Pro acts as a RNA silencing suppressor.

1.8 Small RNA cloning and sequencing

DNA sequencing is used ubiquitously: from deciphering genomes (Ng, et al., 2000) to determining the primary sequence of small RNA. The first small RNAs were cloned from *C. elegans*, Drosophila and humans in 2001 and ever since small RNA cloning proved to be a vital aspect of small RNA research (Lagos-Quintana, et al., 2001; Lau, et al., 2001; Lee, Ambros, 2001). The process of small RNA cloning appends known primers to each end of the small RNA using T4 RNA ligase. The resulting constructs are then reverse transcribed and PCR amplified (Figure 1.11). Cloning and sequencing of these PCR products determines the full sequence of the cloned small RNA and is a conventional approach that gives this information.

Figure 1.11: Small RNA Cloning Scheme.

Small RNAs are isolated by size from a sample of total RNA. Then, 3'- and 5'-cloning primers (5-CP, 3-CP) are ligated onto the small RNA using T4 RNA ligase and converted into cDNA by reverse transcriptase. The cDNA is ligated into a plasmid, transformed into bacteria and plasmids from single colonies isolated for DNA sequencing. The resulting DNA sequence contains the original small RNA sequence flanked by two cloning primers (5-CP and 3-CP).



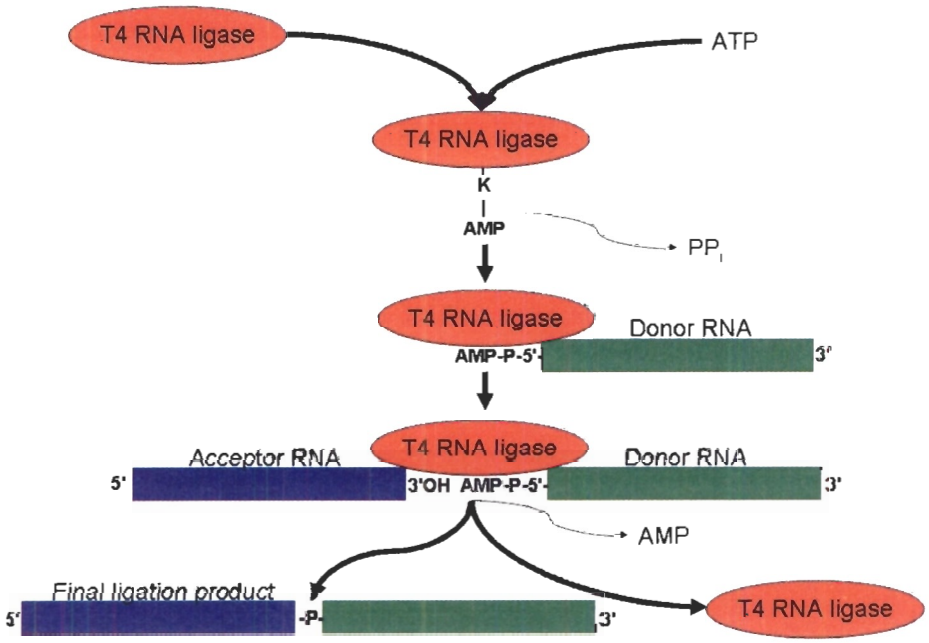
1.8.1 T4 RNA ligase

The crucial first and second step of the RNA cloning protocol involves T4 RNA ligase, which joins the 5' phosphate of an acceptor RNA to the 3' hydroxyl of the donor RNA (see Figure 1.12). The multi-step reaction mechanism includes the initial adenylylation of a lysine residue in the enzymatic core of the T4 RNA ligase. Then the enzyme binds a donor RNA and the adenosine monophosphate is transferred from the lysine to the 5' phosphate of the RNA *via* a 5'-5' linkage. The T4 RNA ligase now binds the acceptor RNA and facilitates the phosphor-disester bond formation between the acceptor RNA's 3' hydroxyl and the 5' adenylylated donor RNA. The two RNA molecules are covalently bound and the T4 RNA ligase is available for another reaction. (Bryant, Benkovic, 1982; Cherepanov, de Vries, 2002; Ho, et al., 2004; Wang, Silverman, 2006).

RNA ligase is specific to ligating two RNA molecules. Does the enzyme tolerate deoxynucleotides? Nandakumar and colleagues addressed the specificity of RNA ligase (Nandakumar, Shuman, 2004). The researchers conclude from their experiments that RNA ligase is able to join a 3'-hydroxyl RNA to a 5'-phosphate RNA or DNA. Investigating changes to the 2'-OH on the 3' terminus, the researchers observed a modest decrement of ligation efficiency when the researchers introduced a 2'-O-CH_3. Earlier experiments already established that T4 RNA ligase can ligate a 5' adenylylated DNA sequence to an RNA 3' hydroxyl (Sugino, et al., 1977).

Figure 1.12: T4 RNA ligase reaction mechanism.

T4 RNA ligase is adenylylated at a lysine residue. The enzyme then binds a donor RNA and transfers the AMP onto the donor RNA. By binding the acceptor RNA, the close proximity of the 3' hydroxyl and the adenylylated 5' terminus of the donor RNA will lead to phosphor diester bond. The two RNA molecules are covalently bound and the T4 RNA ligase is available for another ligation.

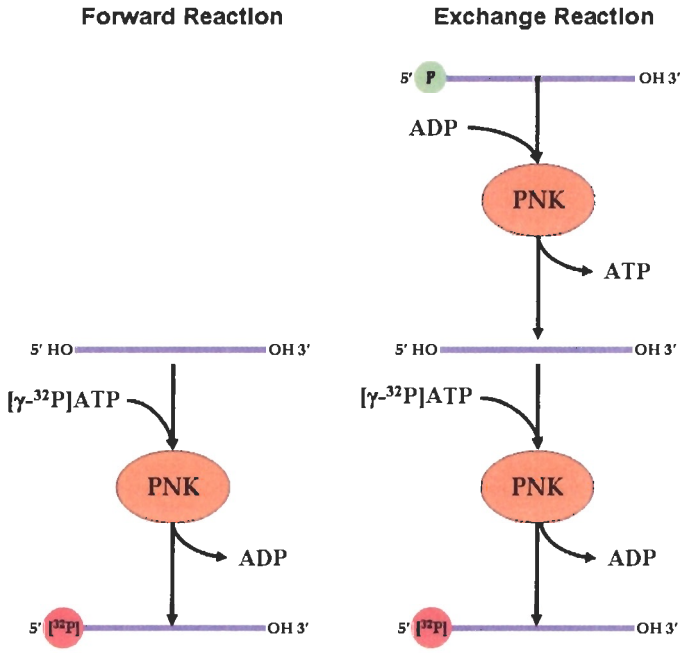


1.8.2 Polynucleotide kinase

To visualize the ligation reaction on an autoradiogram, T4 polynucleotide kinase is routinely used to transfer a radioactive gamma phosphate from ATP onto the 5' hydroxyl of a RNA or DNA strand (forward reaction). If the 5' terminus is phosphorylated, polynucleotide kinase will first remove the 5' phosphate leaving a 5' hydroxyl and subsequently perform the forward reaction (this process is also known as exchange reaction). Figure 1.13 visualizes both mechanisms of T4 polynucleotide kinase (Eastberg, et al., 2004; Wang, et al., 2002).

Figure 1.13: Polynucleotide kinase.

Polynucleotide kinase carries out a forward reaction, which transfers the gamma phosphate of an ATP molecule onto a 5' hydroxyl DNA or RNA strand. If the 5' terminus is phosphorylated, polynucleotide kinase first removes the phosphate and then carries out the forward reaction, a process known as exchange reaction.



1.8.3 Analysis of sequencing data

To learn more about the effects of P1/HC-Pro on small RNAs, the total population of small RNAs from tobacco plants were cloned and sequenced. A four point comparison strategy was used to identify individual and compound affects. Cloning and sequencing is currently the only available methodology to determine the sequence of a small RNA. Mainstream techniques such as microarrays (Lim, et al., 2005) and Northern analysis (Valoczi, et al., 2004) require prior knowledge of the primary small RNA sequence.

Another novel approach of sequencing DNA is the massive parallel signature sequencing developed by Solexa. The DNA sequence of interest is appended on each side with PCR amplification primers using DNA ligase. The DNA anneals on a glass chip, which contains a dense lawn of primers. Then the bound DNA is PCR amplified in a so-called bridge amplification step. The DNA molecules are then denatured and are available for another round of PCR amplification. This process is iterated until clusters of DNA molecules are amplified on the glass chip. For the actual DNA sequencing step, primers, fluoresently labelled nucleotides, which are

reversely blocked on their 3' terminus, and DNA polymerase is added. After each cycle the excess reagents are removed, the fluoresence signal is determined and the 3' terminus unblocked for another addition of reversely blocked flouresent nucleotides (Solexa sequencing technology, Illumina Inc., San Diego, CA.). The process is iterated and will lead to a short 20 to 30 nt long sequence, a portion of which is the annealing primer. Thus, massively parallel signature sequencing can give estimates of total small RNA numbers, but it does not provide complete primary sequence information (Lu, et al., 2005).

A recent development in DNA sequencing is the advent of commercially available 454 sequencing generating large numbers (500,000) of DNA sequences up to 300 nts in length, which essentially follows the traditional cloning protocol without inserting the cloning fragment into a bacterial vector (Margulies, et al., 2005).

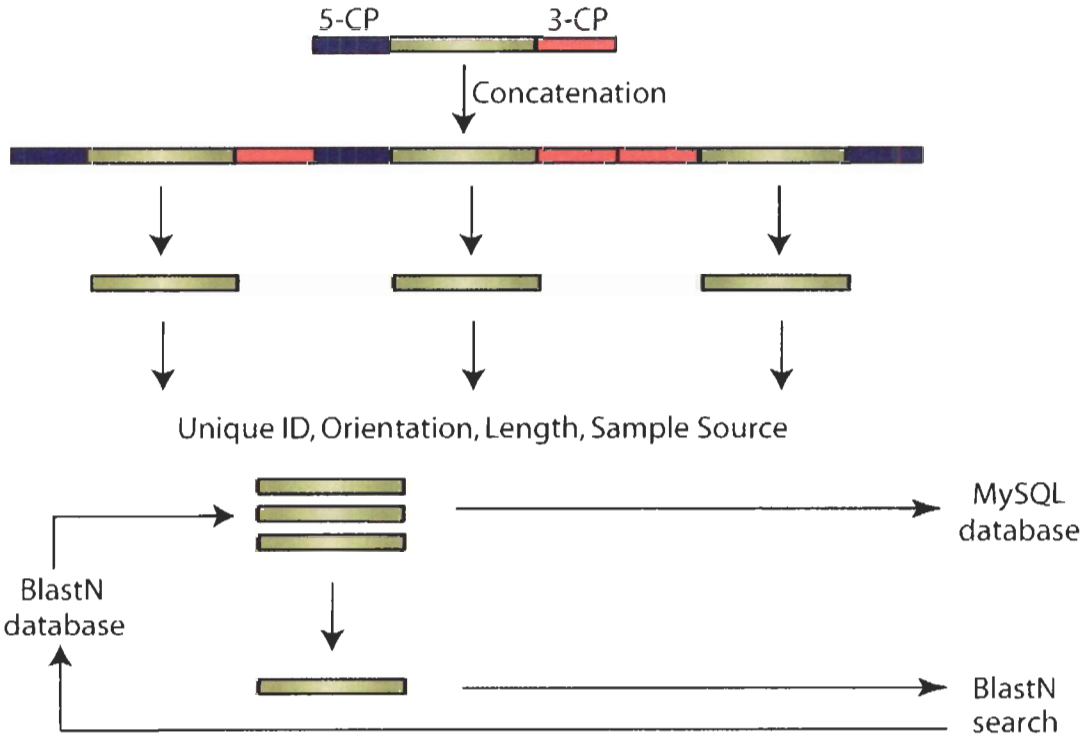
As a result of the conventional small RNA cloning strategy it is obvious that the small RNA sequence is found between the 5-CP and 3-CP. Therefore, a sequencing text file containing single or concatenated small RNA sequences would be scanned for matching 5-CP and 3-CP pairs. To our knowledge, no software performing this task was available publicly. Other DNA sequencing programs are concerned with automated base calling, e.g. *phred* (Ewing, Green, 1998; Ewing, et al., 1998). The closest DNA sequence analysis tools are vector-trimming programs, which remove external vector sequences from the DNA sequence (Chou, Holmes, 2001; Li, Chou, 2004). Removing the vector from small RNA sequencing data leaves the small RNA sequence surrounded by their 5-CP and 3-CP. Other shortfalls of vector removal programs are that they do not determine the directionality of the insert and fail at excising multiple inserts. Also, once the vector is removed there is typically no further analysis of the remaining sequence, e.g. reliable storage and BlastN search. Finding no easily accessible research tool to enter and analyze small RNA sequences I developed *Ebbie*. This tool has since been expanded and utilized for a number of high-throughput sequencing projects by the Unrau lab.

2 *EBBIE*: ANALYSIS AND STORAGE OF SMALL RNA CLONING DATA USING A DYNAMIC WEB SERVER.

To facilitate the analysis and comparison of newly acquired sequencing data I developed *Ebbie*: an automated analysis and storage pipeline using a dynamic web server. Figure 2.1 shows a concept overview of *Ebbie*. In essence, a single or multiple small RNA sequences are excised from a DNA sequencing file using 5' and 3' constant sequence regions. Then, each sequence obtains a unique ID and is deposited into the MySQL database in the correct 5' - 3' direction. The length of the cloned small RNA is automatically counted and the sample source determined from the first letter of the sequencing file.

Figure 2.1: ***Ebbie*** **– Concept Overview.**

A sequencing text file containing single or concatenated small RNA sequences is searched for matching 5-CP and 3-CP pairs. The sequence in between these flanking sequences is excised, a unique ID, orientation, length of insert and sample source automatically determined. All these facts are recorded in a MySQL database and each newly isolated small RNA sequences is analyzed using BlastN to locally installed databases. The new annotation is deposited into the MySQL and BlastN databases and the process repeated until no further unannotated sequences remain.



2.1.1 External software

Blast v2.2.9 is a heuristic local alignment tool essential for comparing query sequences to large databases (Altschul, et al., 1997). When installed locally, it proves to be a powerful and versatile tool for comparing new sequences to personalized local databases. For our published study, a blast-database containing 1,919 small RNA sequences (43,724 nucleotides) from the *Arabidopsis* small RNA cloning project (Gustafson, et al., 2005) was installed locally. Querying this database using BlastN, overlaps of at least 8 consecutive base pairs were detectable using default parameters [-e Expectation value (E) default = 10.0; -W Word size, default = 11; -S Query strands to search against database where 3 is both, 1 is top, 2 is bottom default = 3; -F Filter query sequence (DUST with blastn) default = T]. This was sufficient for our cloning project of 698 small RNA clones. For larger databases, optimized Blast parameters might be necessary. If genomic sequence data is available, BLAT (Kent, 2002) or other non-heuristic alignment algorithms such as Smith-Waterman (Waterman, Eggert, 1987) might be considered for annotating small RNAs to a genome.

2.1.2 Components of *Ebbie*

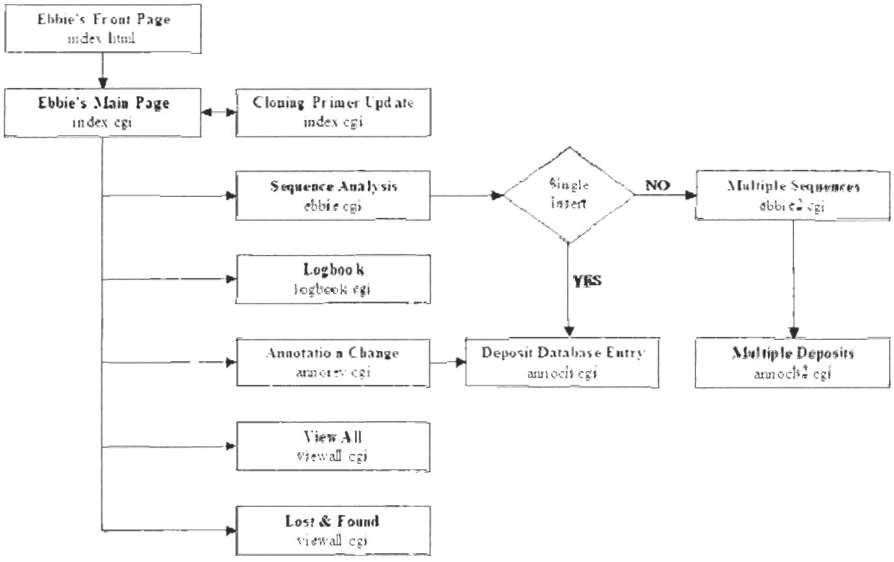
MySQL v4.1.10a-Max was chosen as a database due to its compatibility with Perl. Perl v5.6.0 was chosen as a programming language because of its strength in analyzing and manipulating strings. Perl serves well in creating dynamic web pages (Castro, 2001), interacting with MySQL databases (DuBois, 2001) and communicating with the operating system. Most Linux systems are distributed with these programs. *Ebbie* was originally implemented on a standard PC with Linux Novell SuSE 9.3 operating system (standard PC: AMD Athlon 1.1 GHz processor with 256 KB cache, 512 MB RAM, 60 GB HD) using the root's web server directory. The dynamic web tool was since implemented under user accounts on the http://bioinformatics.org/ebbie Linux server and local server http://unrau1.mbb.sfu.ca/~aebhardt/ (the latter PC runs Linux Novell SuSE 10.0 operating system with Intel D 3 GHz, x86_64 architecture with 2,048 KB cash, 2 GB RAM and 230 GB HD).

2.1.3 Flowchart of *Ebbie*

Ebbie is comprised of eight perl-cgi scripts. They share four common libraries, which contain common subroutines, e.g. querying the MySQL database and drawing a table with the results. Figure 2.2 shows *Ebbie*'s functions and the perl-cgi scripts required for each.

Figure 2.2: Flowchart of *Ebbie*.

Indicated are *Ebbie*'s functions and the cgi-scripts required for each. *Ebbie*'s front page is created by index.html. From the main page, several scripts can be called. DNA sequencing text files are uploaded and analyzed by ebbie.cgi. Other features include database review/manipulation tools and a logbook



Ebbie's front page is created by index.html. The database information is submitted to index.cgi, *Ebbie*'s main page. Index.cgi can update the currently used cloning primers (5'-cloning primer: 5-CP and 3'-cloning primer: 3-CP). From the main page, several scripts can be accessed. The *Logbook* lists all analyzed files (table created by logbook.cgi). DNA sequencing text files are uploaded and analyzed by ebbie.cgi when *Sequence Analysis* is selected. If a single insert is found, annoch.cgi is called to deposit the insert into the MySQL database. If multiple inserts are found, ebbie2.cgi and annoch2.cgi are used to add these entries to the MySQL database. From *Ebbie*'s main page a range of database review/manipulation tools can also be accessed: *View All, Lost & Found* (tables created by viewall.cgi) and *Annotation Change* (which also uses annoch.cgi to update a database entry). For a detailed discussion of these functions see text below.

2.1.4 Description of *Ebbie*

Ebbie's dynamic web pages are platform and browser independent (browsers tested: Mozilla Firefox 1.0.4 on Linux and Windows XP operating system, Safari 2.0 for Macintosh, MS Internet Explorer 6.0 on Windows XP). With *Ebbie* v3.0.9 a login was implemented. The login serves as database name and *Ebbie* subsequently works with this database. As seen in Figure 2.3, *Ebbie*'s main page has four basic functions. First, it retrieves the current 5'- and 3'-cloning primers (5-CP, 3-CP) from the selected MySQL database and displays their sense and antisense

sequence. The user can edit these sequences as required. All sequences entered into these text fields are capitalized. Second, the user can browse the local computer's file manager to select and analyze a DNA sequencing text file. *Ebbie* maintains a log of all processed files that can be reviewed by clicking on the *Display logbook* function. Through *Ebbie*'s main page, three online database management tools are available: *Annotation Change*, *View All* entries of database and *Lost & Found*.

The uploaded file name serves as the primary *id* for the MySQL database entry. If no file is selected or if the *id* already exists in the database, an error message is displayed and the process aborted. If a file is valid (i.e. *id* is new and unique), the DNA sequence data is converted into a string, capitalizing the characters A, C, G and T. All other characters remain unchanged. Perl's *index* function is used to confirm that at least one 5-CP and 3-CP pair exists, if this condition is not met or if perl's *index* function identifies an uneven number of 5-CP and 3-CP pairs, then an appropriate error message is generated in the logbook. The algorithm starts at the 5' end of the DNA sequence and finds the first occurrence of a 5-CP (or antisense 3-CP). Moving in the 3' direction, the next 3-CP (or antisense 5-CP) is located. An insert is deposited into the MySQL database, if a sequence of length > 0 is found between the two primer pairs. If no insert is found in the DNA sequencing text-file, a message is displayed and recorded in the logbook.

2.1.4.1 Database selection

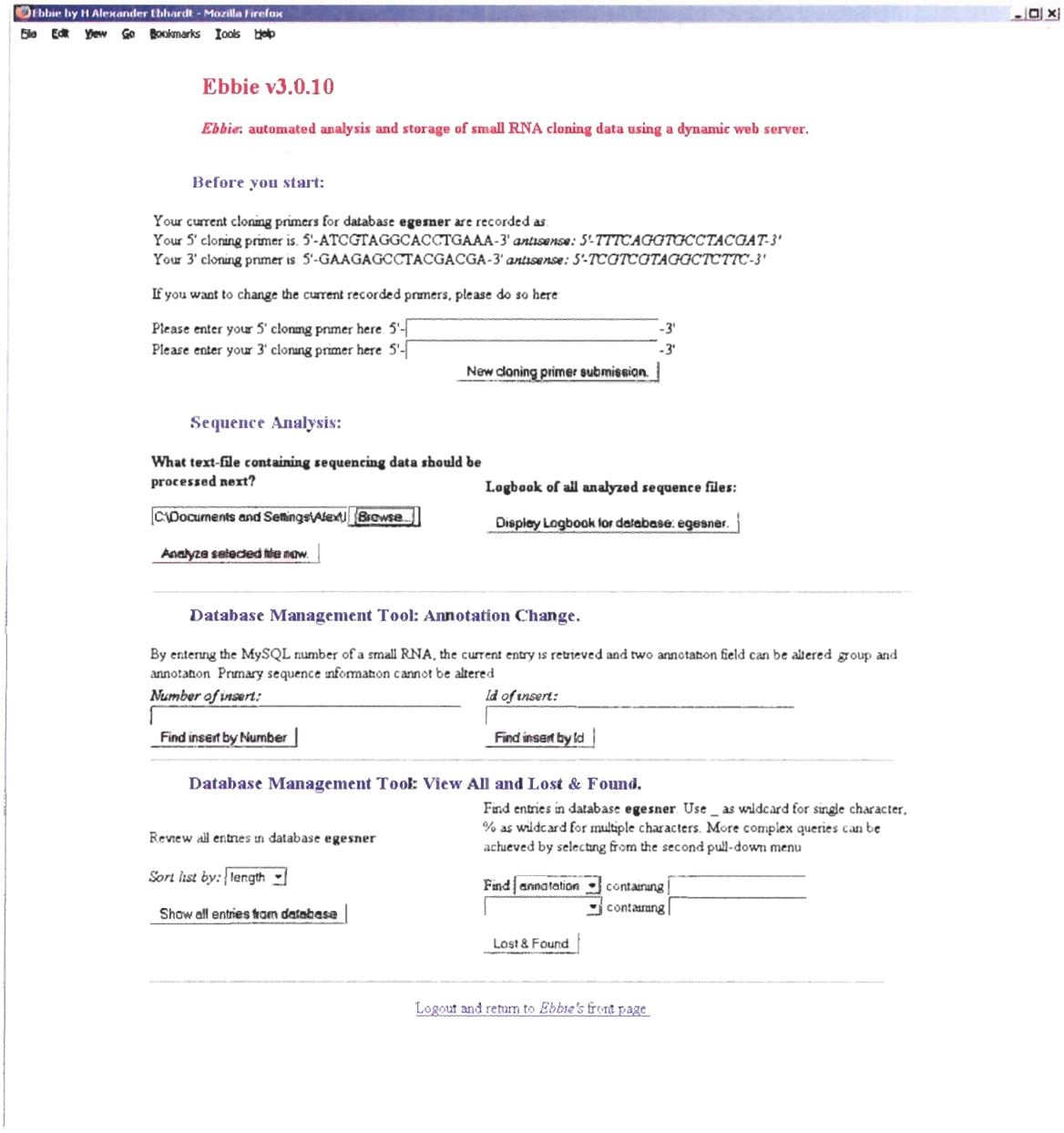
With *Ebbie* v3.0.9, a login was implemented. Each login user name is associated with a database. Thus, once logged on, *Ebbie* will use the same database until the user logs out and logs in with a different user name. This feature is advantageous to keep various small RNA cloning projects separate. Also, the user can choose the DNA sequencing file names wisely and use the same database with different samples from the same cloning project (see sample source description below).

Each MySQL database is accompanied by a BlastN database. The latter database is a dynamic text file to which each new entry is appended in FASTA format.

The *'Database Management Tool: Annotation Change'* allows the user to change only two fields of each insert: '*annotation*' and '*group*', all other fields (*no, id, sequence, length, orientation* and *sample source*) cannot be edited from the web interface in order to preserve the integrity of the database. This restriction was deliberately chosen to maximize the integrity of primary data.

Figure 2.3: Screenshot of *Ebbie's* front page.

The user is able to (1) review the current cloning primer pairs and update them if necessary. (2) Sequencing file analysis by uploading a text file containing the sequencing data and keeps a logbook of events. (3) Database management tool: *Annotation Change*. A database entry is retrieved and annotations can be updated. (4) Database management tools: *View All* and *Lost & Found*.



2.1.4.2 Analysis of inserts

Once an insert is found, *Ebbie*:

- deposits the *id* and sequence insert into the dynamic BlastN database,
- deposits the insert into the MySQL database, in the correct sense specified by the

orientation of 5-CP and 3-CP,

- determines the inserts *length*,
- determines its *id* based on the file name, and
- determines its *sample source*, which is inferred from the first character of the file name.

The latter function relies on *grep* to determine the initial character and then assigns the sample source by referencing an external text file. This sample source assignment can easily be manipulated by editing the external text file (/ebbie/mod/source.nt). Currently, file names starting with 1, 2, ... 9 have an automatic sample source assigned; other file names will result in '*unknown*' sample source.

Following the automated sequence deposition, the sequence insert is subject to BlastN searches against locally installed databases. In our case, the BlastN searches included the BlastN database from the *Arabidopsis* smRNA-cloning project (Gustafson, et al., 2005), the genomes of Y-Satellite plus its helper virus Cucumber Mosaic Virus (NCBI Genbank accession numbers for viral genomes: NC_002034.1, NC_002035.1, NC_001440.1, D10038.1) and a complete BlastN database of all previously found inserts. The latter dynamic BlastN database extends each time a new insert is found, allowing for rapid identification of new groups. Each BlastN analysis is scanned by *grep*, probing for '*No hits found.*' in which case it will only print one line onto the screen, indicating the unsuccessfully searched database. Otherwise, the complete BlastN analysis is displayed on the web page to facilitate user-mediated annotation. Figure 2.4 shows the analysis of a clone, in this case finding a complete match against three previous entries.

The user can now fill out three additional annotation fields:

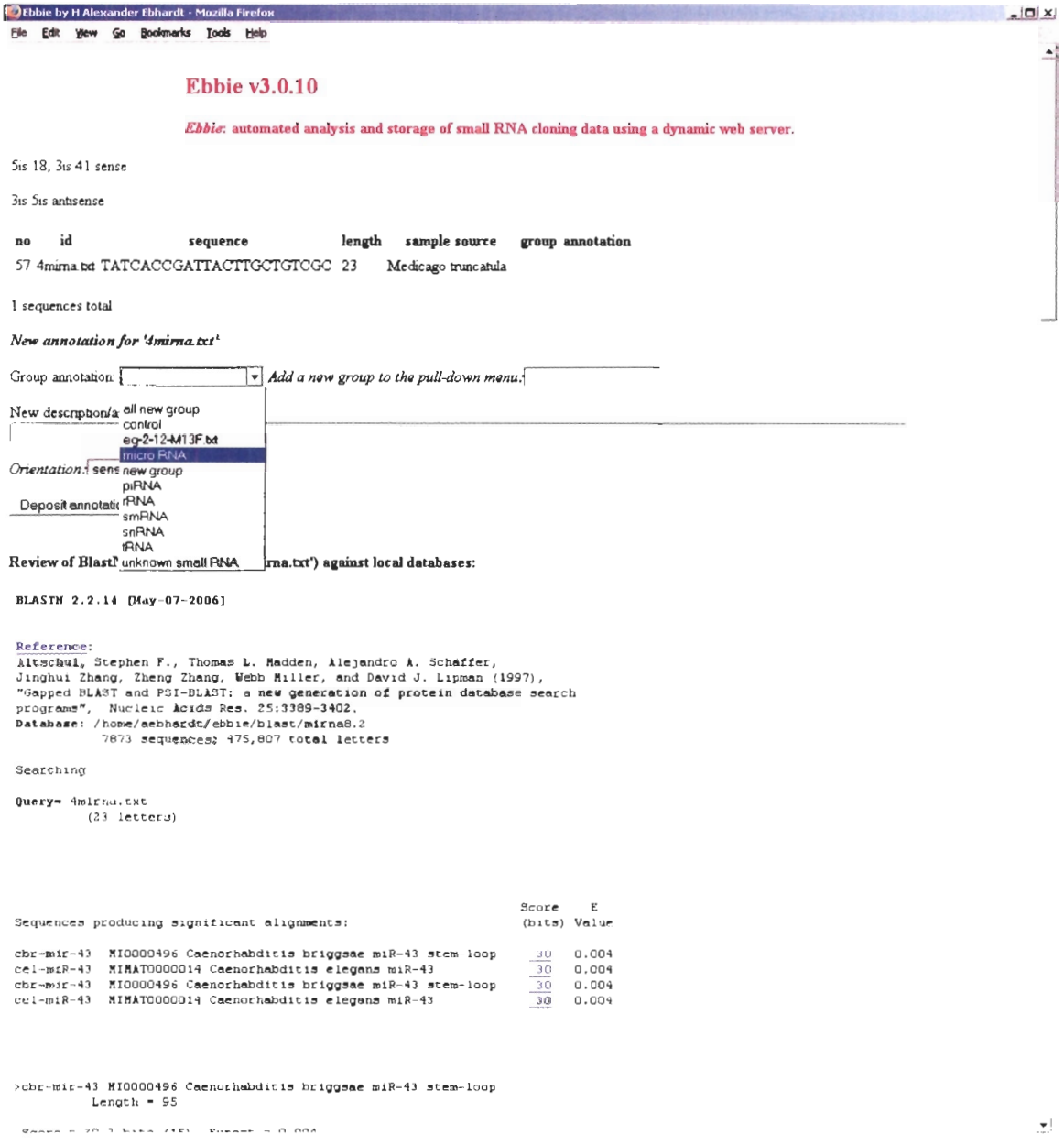
Group pull-down menu: The *group* pull-down menu offers standard RNA types found previously during data entry and analysis. A *new group* can be added through the accompanying text field if a *group* is identified by local BlastN analysis. Once submitted, this *new group* description is simultaneously added to the small RNA annotation in the MySQL database, the BlastN dynamic database and the *group* pull-down menu. The latter menu is sorted alphabetically and is made available for subsequent *group* annotations. This form of annotation proved quite powerful in the analysis of our set of small RNA.

Annotation field: a text field allowing for user generated comments based on the automatic BlastN searches or external BlastN searches (our BlastN searches were limited by the

amount of RAM available on the server).

Figure 2.4: Analysis of a small RNA sequence using *Ebbie*.

The first table shows the MySQL database entry, which is created automatically once a cloned small RNA is found (entering id/filename, sequence of small RNA in correct 5'-3' orientation, length and sample source). The group annotation pull-down menu offers the user a choice from a variety of group annotations. A new group can be named in the adjacent text field. Further annotation details can be added in the annotation text field. The orientation field offers three choices: N/A, antisense and sense.



Orientation pull-down menu: allows the selection of three categories: N/A, sense and antisense to classify the BlastN search results. This feature is also used when working with

multiple inserts – all non-annotated inserts in the MySQL database will have a "NULL" entry in this column. Once the insert is annotated, the "NULL" will be replaced by the above mentioned choices.

Once annotated, the insert's MySQL entry is updated by pressing the submit button. Consecutively, *Ebbie*'s deposit algorithm appends the newly annotated entry to the BlastN text file in FASTA format. The flat file is subsequently formatted into BlastN files for subsequent BlastN analysis. The newly created web page displays the MySQL entry (*id*, *sequence*, *length*, *group* and *annotation*) and allows the user to return to *Ebbie*'s main page or make annotation changes to newly discovered group members.

An example: ribosomal RNA group 01

We classified a group as two sequences overlapped by 12 or more consecutive base pairs. This empirical overlap proved to be stringent in retrospect; a 16 base pair non-gapped overlap would have resulted in 32 groups. The percentage overlap score from BlastN is sometimes misleading, as a BlastN search might not align the whole query sequence to any given subject, thus misleading the user about the percentage identity.

During our small RNA cloning project of virally infected tobacco plants[6], *Ebbie* identified 33 groups among 700 small RNA sequences. The first group *Ebbie* identified in infected/non-infected tobacco plants was a 24 nt small RNA resulting from the 5' start of the small ribosomal RNA. This accumulation is an intriguing fact and does not seem random, considering that 18S ribosomal RNA is approximately 1,800 nts in length. This group was further investigated (see Chapter 4 starting at page 62).

2.1.4.3 Multiple inserts

If the number of inserts in the sequencing file exceeds one, all inserts are automatically inserted into the MySQL database in the correct 5'-3' orientation, together with their *length* and *sample source*. The *id* for each insert is specified uniquely by appending to the end of the filename a unique insert number. The user is notified of the number of primer pairs found and the number of inserts deposited into the MySQL database. To analyze individual sequences, a pull-down menu displays all inserts found in the current sequencing file. Following the selection of an insert, the user can analyze each one individually (as described in the previous section above). As long as unannotated inserts are in the database, the user can select from the pull-down menu remaining unannotated cloned small RNA sequences.

2.1.4.4 Logbook function

The logbook function is accessible from *Ebbie*'s main page. Each time a file is uploaded and analyzed by *Ebbie*, the system time is recorded, together with the filename. Once the file is analyzed, a comment is recorded depending on the outcome of the analysis: 'Sorry, there was no insert found', 'Single insert found.', 'There were *x* primer pairs and *y* inserts deposited into *z*' (where *x* is the number of primer pairs found, *y* the number of inserts deposited and *z* the database used) and 'Number of 5'- and 3' -cloning primers not even!'. The latter comment is displayed in red, as this file may need manual intervention to rescue its content before subjecting it again to the insert excision algorithm.

2.1.4.5 Review database

All entries in the selected database can be reviewed and ordered by *id*, *length*, *group* and *number* fields using the '*View All*' button. For each database, *Ebbie* will remember the last selection of this pull-down menu. This feature is useful while generating a database and allows a quick survey of the database during data entry.

2.1.4.6 Lost & found

The *Lost & Found* function allows the user to use one or more wildcard characters for querying the database. '_' is used for single character and '%' for multiple character wildcard. From a pull-down menu the user selects a category, e.g. *id*, and in the adjacent text field the query is entered, e.g. '3%'. In this example, all entries with the starting character of '3' are displayed.

For more complex queries, a second pull-down menu is available, which includes AND/OR BOOLEAN operators. For example, all small RNA belonging to the class of "Y-Sat RNA" AND length of "21" nucleotides can be selected.

2.1.4.7 Change annotation

To update the *annotation* of individual inserts, a *change annotation* script was implemented. The script searches for either the *id* or *number* of the insert. The *id* is useful once a new *group* is identified in a BlastN search. The *number* is convenient after reviewing the database. Once a *number* or *id* has been submitted, the record of the *id* is retrieved from the database (*no, id, sequence, length, sample source, group* and *annotation*). With *Ebbie* v3.0.10 the BlastN search against locally installed databases is repeated and printed below the annotation fields. This gives the user a sense of certainty that the RNA sequence was annotated correctly in

the first place and any annotation ambiguity can be resolved. The user can then choose a standard *group* description from the *group* pull-down menu or add a *new group*. The *'Annotation field'* will display the current annotation in the text field, allowing the user to add supplementary information to it. Once adjustments are made, the new *annotation* can be submitted and the corresponding fields in the MySQL database are updated. Further, if the *group* annotation is changed, the BlastN flat file will be edited to reflect the current *group* annotation. The user is unable to use a wild card character for the *change annotation* function.

2.1.5 Limitations of *Ebbie*

The algorithm will experience difficulties if low complexity or ambiguous repetitive 5-CP or 3-CPs primer sequences are used, which should be avoided by the correct design of primer pairs. Similar fundamental problems are encountered when cloning RNA using poly(A)-polymerase to extend the 3' end of a sequence which may already contain poly(A) residues. Additional wet lab experiments (e.g. primer extension assays) need to be conducted in order to determine the RNA's true 3' end length. Also, no wild card characters are allowed when identifying 5-CP and 3-CP primers within the DNA sequence file, to ensure the quality of the DNA read. Imperfect primers can be identified by a subsequent manual examination of sequence files that are flagged as having uneven or no primer pairs in the logbook.

3 EXTENSIVE 3' MODIFICATION OF PLANT SMALL RNAS IS MODULATED BY HELPER COMPONENT-PROTEINASE EXPRESSION.

In a previous study, Wang and colleagues infected tobacco plants with Y-Satellite and its helper virus Cucumber Mosaic Virus (Y-Sat/CMV) causing the wild-type tobacco plants to yellow. Infection of tobacco plants expressing the strong RNA silencing suppressor P1/HC-Pro (HC-Pro from here on) did not show the yellowing phenotype (Wang, et al., 2004). This implies that Y-Sat/CMV uses the endogenous RNA silencing pathway to cause disease symptoms. Using Northern hybridization experiments wild-type infected plants showed low levels of small RNA (21/22 nt in size) originating from Y-Sat. In HC-Pro expressing infected tobacco plants, the levels of Y-Sat small RNA were 10 times higher, yet did not cause the yellowing symptoms. To explore this paradox, where small amounts of Y-Sat small RNA cause yellowing in HC-Pro(–) plants and appreciable amounts of Y-Sat small RNA in HC-Pro(+) caused no obvious phenotype, small RNA from above mentioned plants were isolated, cloned and sequenced. Further, my research aimed to address the question how P1/HC-Pro acts as a RNA silencing suppressor.

3.1 Small RNAs from tobacco plants

For a two variable experiment, in this case HC-Pro expression and/or infection with CMV/Y-Sat, only one variable is changed at a time. Exploring all options leads to a four-point-comparison-study, allowing the researcher to attribute for the changed variable. Thus, small RNA were isolated from four different groups of tobacco plants, namely HC-Pro(+)/uninfected, HC-Pro(–)/uninfected, HC-Pro(+) CMV/Y-Sat(+) and HC-Pro(–) CMV/Y-Sat(+) representing the four possible scenarios for comparison.

3.1.1 5' radiolabelling of small RNAs from tobacco plants

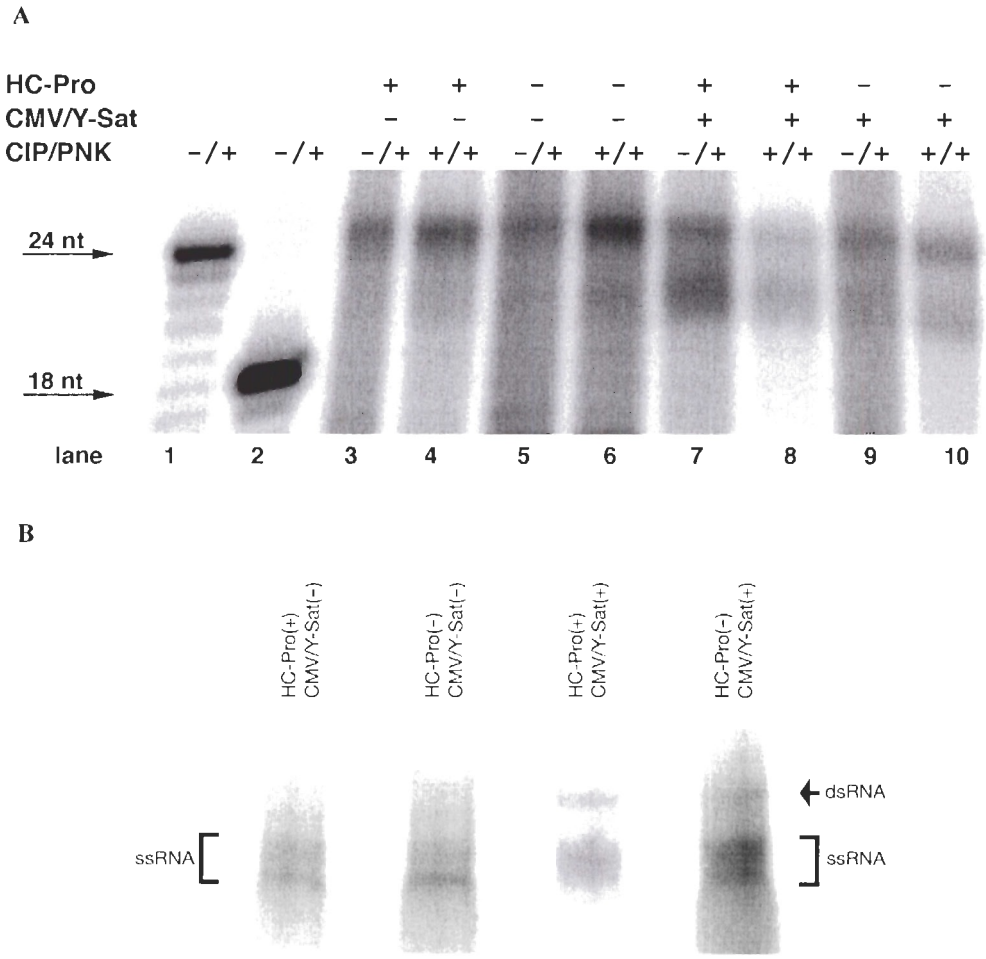
All small RNA samples were 5' radiolabelled with or without prior phosphatase treatment using polynucleotide kinase and [^{32}P]γ-ATP. The radiolabelled small RNA samples were analyzed on a 23.75 % denaturing PAGE, with the resulting autoradiogram shown in Figure 3.1. The intensity of phosphatase treated RNA samples was on average 2.5 times more intense than those RNA that were initially not treated with phosphatase. This observation confirms published

results indicating that mirco- and silencing RNA are 5' phosphorylated. There is a dominant 24 nts band for all plant samples. Y-Sat/CMV infected tobacco plants show an additional band at 21/22 nts.

Analyzing the 5' radiolabelled RNA from all four plants in a native PAGE, it shows that upon infection with CMV/Y-Sat a slower migrating band corresponding to double stranded small RNAs is visible. This indicates that upon viral infection more duplex small RNAs are formed than available RISC are available.

Figure 3.1: 5' Radiolabelling of Small RNA from (un-)infected tobacco plants.

(A) Four tobacco plant small RNA extracts were radiolabelled using polynucleotide kinase (PNK) with an excess of [γ-32P]ATP. The samples were either untreated (–) or pretreated with CIP (+) before labeling, which resulted in nearly identical gel mobility and relative labeling intensities. To obtain approximately equal labeling intensities, the samples pretreated with CIP were diluted by an average of 2.5-fold relative to the samples that were untreated. The RNA size markers in the first two lanes indicate the mobilities of 24- and 18-nt RNA species. The ratio of the background intensity to the intensity of the 24-nt band in each sample was within 2-fold for the uninfected plants. A similar background ratio was observed for the infected plant samples. (B) 5' radiolabelled RNA was run into a native 10 % PAGE. Indicated are the single- and double stranded RNAs.



3.1.2 Cloning and sequencing of small RNAs from tobacco plants

For cloning, the small RNA were gel purified and appended on the 3' terminus with an adenylylated DNA oligo (App16.5) using T4 RNA ligase. Following this ligation, the RNA-DNA hybrid was gel purified and appended on the 5' terminus with an RNA oligo (r17.50) using T4 RNA ligase. Then, the construct was reverse transcribed, PCR amplified and cloned into the TA cloning vector pCR2.0. The transformed bacteria (*E.coli JM109*) were plated agar plates containing Ampicillin and X-Gal. The latter component is used for blue/white screening of colonies. Most pUC vectors encode the first 146 aa of β-galactoside. If a fragment is successfully cloned into the polycloing site, the 146 aa fragment will not be enzymatically active, thus the bacterial colonies appear white (Ullmann, et al., 1967). PCR products from individual colonies were evaluated for containing inserts of appreciative length by agarose gel electrophoresis and send for DNA sequencing to determine the primary sequence of the cloned small RNA.

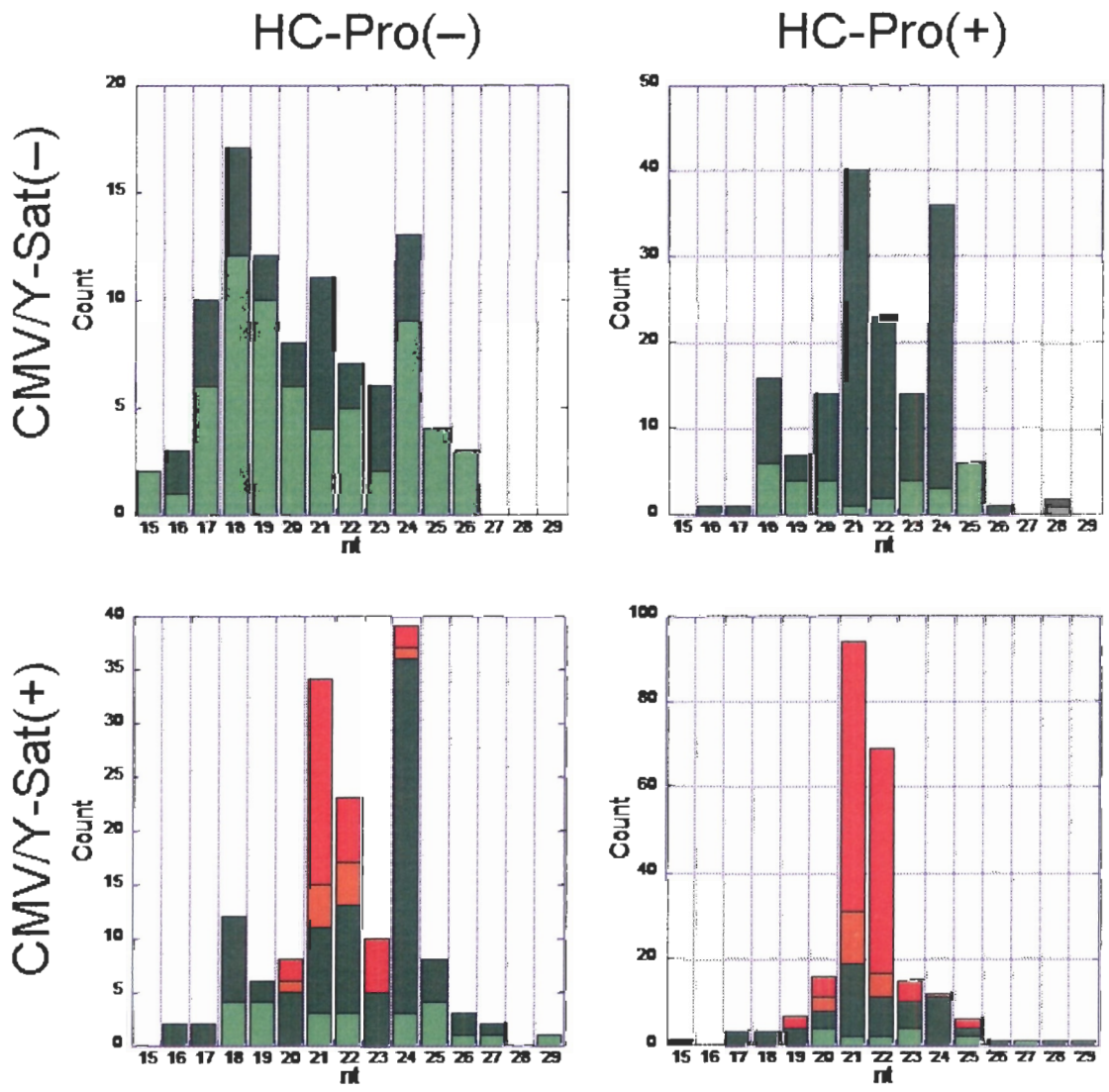
In total 698 small RNA sequences from all four plant types were deposited into the MySQL database using *Ebbie*. These sequences were classified into four major groups based on their annotation: ribosomal or transfer RNA fragments, cucumber mosaic virus, Y-Satellite and plant sequences. The latter group included sequences that were easily identifiable, such as micro RNA or messenger RNA fragments, and small RNA with unknown annotation. Plotting the length distribution of these small RNAs and colour coding the four major groups resulted in the following histograms (see Figure 3.2).

Looking at the size distribution of cloned small RNA from uninfected plants (Figure 3.2 top panels), it is notable that the amount of 24 nt RNA is appreciatively less than what would be expected from the 5' radiolabelling (Figure 3.1 lanes 3 to 6). 21/22 nt small RNA are twice as abundant than 24 nt small RNA in the histogram of cloned small RNA. This is in contrast to five times more 24 nt over 21/22 nt RNA in the 5' radiolabelling. Comparing the uninfected plants, HC-Pro expression causes the number of cloned ribosomal and transfer RNA fragments to decrease significantly.

The histogram of cloned small RNA from HC-Pro(−) CMV/Y-Sat(+) plants (Figure 3.2 bottom left panel) and their 5' radiolabelling (Figure 3.1 lanes 9, 10) agree to the greatest extend. When comparing the 5' radiolabelling autoradiogram of small RNAs from HC-Pro(−) (un-) infected plants (Figure 3.1, lanes 5 and 6 *versus* lanes 9 and 10), one observes that upon viral infection 21/22 nt small RNAs appear. The small RNA cloning confirms that the majority of 21/22 nt small RNA in infected plants are of viral origin, confirming previous Northern hybridization experiments (Wang, et al., 2004).

Figure 3.2: Histogram of cloned small RNAs.

The top two histograms are from uninfected plants, the bottom two are from Y-Sat/CMV infected tobacco plants. On the left are HC-Pro(–) plants, on the right HC-Pro(+) plants. In light green are tRNA and rRNA fragments; dark green plant related small RNA sequences; orange are small RNA originating from CMV and in red are small RNA from Y-Sat.



3.2 Discrepancy between 5' radiolabelling and cloned small RNA histogram

The size distributions in the histogram of cloned small RNA (Figure 3.2) agree well with the 5' radiolabelling (Figure 3.1), except for HC-Pro(+) CMV/Y-Sat(+) plants. The HC-Pro(+) infected plants showed a bimodal distribution of small RNAs in the autoradiogram of the 5' radiolabelling (Figure 3.1, lanes 7 and 8), but only a mono-modal distribution in the histogram of

cloned small RNAs (Figure 3.2, bottom left panel). This discrepancy was further investigated (see below).

3.2.1 Ligation experiments

Investigating this discrepancy further, a time course of the first 3' adaptor ligation (using App16.5 and T4 RNA ligase) with 5' radiolabelled small RNA from Y-Sat/CMV infected tobacco plants HC-Pro(–)/(+) was carried out. Ligation time points were analyzed on a 23.75 % denaturing PAGE and visualized by phosphor imaging. The resulting autoradiogram (Figure 3.3 A) shows clearly the time dependent accumulation of ligation products for a control RNA (lanes 1 to 4) and small RNA populations from HC-Pro(–)/(+) infected plant (lanes 5 to 12).

The time dependent ligation profile for HC-Pro(–) CMV/Y-Sat(+) plants shows a bimodal distribution (Figure 3.3 A, lanes 9 to 12). In contrast, HC-Pro(+) CMV/Y-Sat(+) plant RNAs show a bimodal distribution before ligation (Figure 3.3 A, lane 5) but only a mono-modal distribution of time dependent ligation product (Figure 3.3 A, lanes 6 to 8).

To further visualize the effect of the 3' adaptor ligation, the starting population in t_0 and the ligated population in t_{final} were visualized in a histogram. In Figure 3.3 B t_0 (blue) and t_{final} (red) overlap in RNA ligations from HC-Pro(–)/CMV/Y-Sat(+), but shows a lack of large ligation products corresponding to the ligation of App16.5 to 24 nt RNA in HC-Pro expressing infected plant samples. Thus, the discrepancy seen between the 5' radiolabelling and the histogram of cloned small RNA is due to the first 3' terminus ligation step.

The ligation results in Figure 3.3 A, lanes 5 to 12 clearly demonstrate that the discrepancy between the 5' radiolabelling and the histogram of cloned small RNAs is due to the ligation on the 3' terminus of the RNA, the first step in the cloning protocol. It still leaves the possibility that the 24 nt small RNAs did not ligate well because they were blocked on their 3' terminus. To address the question if the 24 nt small RNA in HC-Pro(+) CMV/Y-Sat(+) plants were blocked on their 3' terminus, gel purified 24 nt RNA from all four plant types were ligated to the 3' adaptor App16.5 using T4 RNA ligase. A time dependent accumulation of ligation product is visible in lanes 13 to 28 of Figure 3.3 B. Thus, all 24 nt small RNA are in principle ligatable.

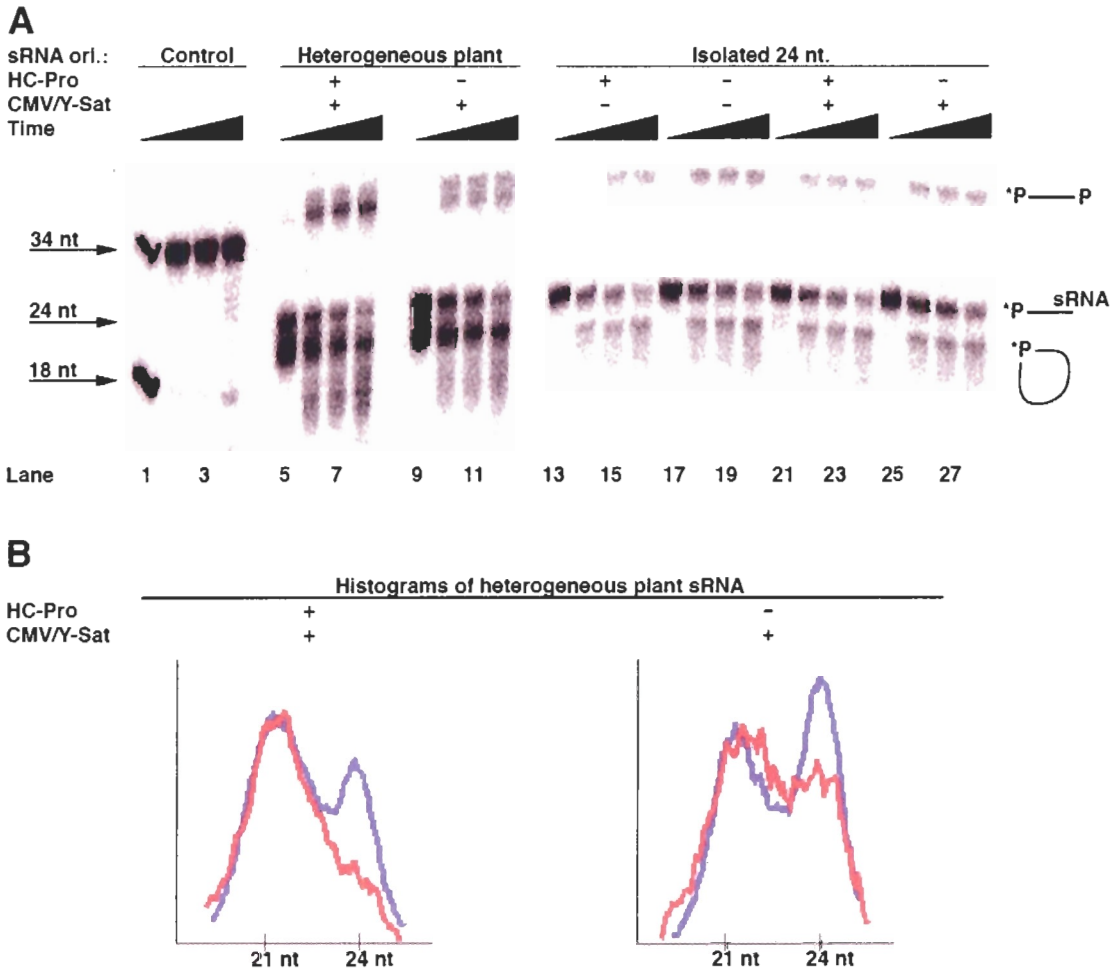
Taking together all facts accumulated thus far, the 24 nt RNA would have been cloned if they were the sole component of the ligation reaction. But, the RNA sample from HC-Pro(+) CMV/Y-Sat(+) plants had two populations of small RNAs. Because the 21/22 nt small RNAs

cloned much more efficiently, they were a better substrate for T4 RNA ligase. One possible hypothesis is, that in HC-Pro(+) CMV/Y-Sat(+) plants the Y-Sat 21/22 nt small RNA are not modified on their 3' terminus, thus they are the preferential substrate for T4 RNA ligase over modified small RNAs.

Besides the expected longer ligation products (RNA and adenylylated DNA are ligated), a second time dependent band migrating faster than the unligated 24 nt RNA. Judging by the mobility of the band, the time dependent product might be circular RNA. To provide evidence for this hypothesis, the ligation material was subject to alkaline phosphatase treatment and only the band below the unligated material maintained its 5' radiolabel. This circularization (approximately 15 %) is due to T4 RNA ligase basic requirements of a 5' phosphoryl and a 3' hydroxyl for ligation. Ideally, the ligation takes place between two RNA molecules, but in 15 % of the cases the molecule is ligated in *cis*. This *cis*-ligation could be avoided by using a T4 RNA ligase mutant that requires a 5' adenylylated nucleic acid molecule and a 3' hydroxyl. The only *cis*-ligation possible for this T4 RNA ligase mutant is the circularization of the 5' adenylylated adaptor molecule, not the desired small RNA molecule. This circularization might also be concentration dependent, as the 5' phosphorylated control RNA did not accumulate significant amounts of circular RNA species.

Figure 3.3: 3' Adaptor Ligation.

(A) The 21- to 24-nt small RNAs from infected HC-Pro(+) or HC-Pro(–) plants (lanes 5-12) or the 24-nt small RNAs from both infected and uninfected HC-Pro(+) and HC-Pro(–) plants (lanes 13-28) were gel-purified and ligated to a 16-nt adenylated DNA oligonucleotide. A synthetic 18-nt RNA control ligation is also shown (lanes 1-4). The ligation was analyzed at four time points for all of the samples: 1 min, 1 h, 2 h, and overnight. The slower-migrating species correspond to ligated RNA-DNA ligation products; the faster-migrating bands that emerge during ligation correspond to circularized RNA (only these bands remain labelled upon treatment with CIP, indicating the circular nature of the RNA). The RNA-DNA ligation product from infected HC-Pro(+) plants that corresponds to the 21- to 22-nt RNAs is significantly enriched (lanes 5-8) relative to the 24-nt RNAs. This effect does not occur in the absence of HC-Pro expression (lanes 9-12). The 24-nt RNAs from either infected or uninfected plants [HC-Pro(+) or HC-Pro(–)] preferentially form self-ligated circular species, with a 3:1 ratio between this circular species and the DNA-RNA ligation products. In contrast, the ligation of the 18-nt control RNA (having the terminal sequence of... AAG-3') is much faster and only gives a negligible amount of circular product (lanes 1-4). (B) Histograms showing the small RNA distribution before ligation (blue) or in the RNA-DNA ligated population (red) from viral-infected HC-Pro(+) plants (Left) or HC-Pro(–) plants (Right).

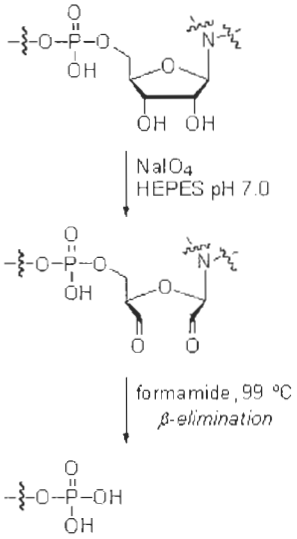


3.2.2 Periodate cleavage experiments

Following the T4 RNA ligase experiments, it was hypothesized that within the population of small RNAs there are subpopulation with differential 3' terminal modification status. To explore the status of a blocked 3' terminus, periodate cleavage experiments were performed. Sodium periodate converts the 3' terminal ribose into an unstable dialdehyde, which can be removed in a β-elimination step effectively separating the 3' terminal nucleoside from the RNA. This reaction only takes place when both 2' and 3' hydroxyls are unmodified. If untreated and periodate treated samples are run into a high percentage denaturing PAGE, the periodate sensitive sample will migrate as an approximately 2 nt shorter fragment. This is due to the loss of the 3' terminal nucleoside, e.g. a 21 nt 5' phosphorylated RNA is comprised of 21 nucleosides and 21 phosphates. Due to the periodate treatment and β-elimination, the treated RNA will now only have 20 nucleosides, but still 21 negatively charged phosphates. The periodate cleavage reaction steps and subsequent β-elimination of an unmodified 3' terminal ribose are shown in Figure 3.4.

Figure 3.4: Periodate mediated cleavage of unmodified RNA.

Shown here is the 3' terminus of an RNA molecule. The terminal sugar is not modified, thus sodium periodate is able to open the cyclic ribose ring in neutral buffer conditions. This cleavage results in an unstable dialdehyde, which is cleaved off in formamide and heat (99 °C), a process termed β-elimination.

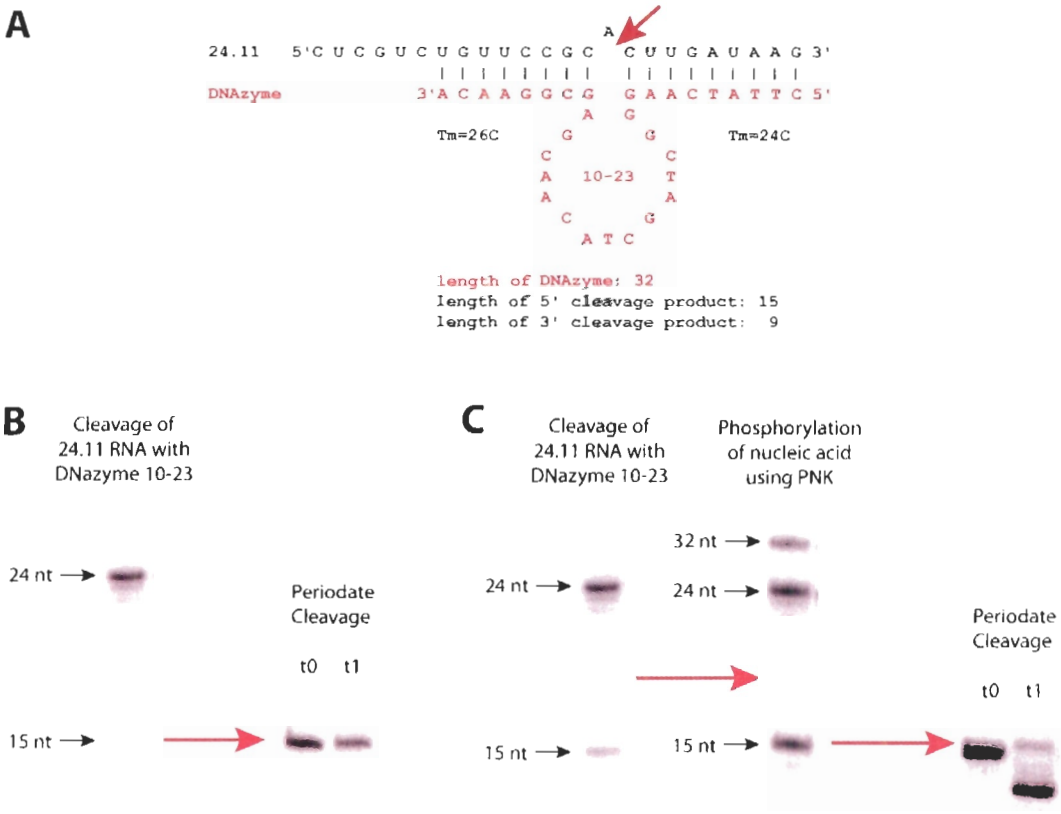


One typical modification of the 3' terminal is a 2',3'-cyclic phosphate originating from endo- or exonucleolytic cleavage of RNA. To prove that T4 polynucleotide kinase could remove a cyclic phosphate, a 5' radiolabelled 15 nts RNA substrate with a 3' terminal is a 2',3'-cyclic phosphate was generated using DNAzyme 10-23 (Santoro, Joyce, 1997). The cleavage product of this reaction was subject to periodate cleavage and proved to be periodate insensitive (Figure 3.5,

panel B, t0 vs. t1). The same cleavage reaction was used, but after the initial cleavage event, the entire reaction was phosphorylated using [^{32}P]γ-ATP and polynucleotide kinase (PNK). After this reaction, the 15 nt RNA was gel purified and proved to be sensitive to periodate treatment (Figure 3.5, panel C, t0 vs. t1). Judging from these observations, T4 polynucleotide kinase successfully removed the 2',3'-cyclic phosphates. The removal of the cyclic phosphate is done in a two-step process by which 2',3'-cyclic phosphate is converted to a 3'-PO_4 and then hydrolyzes the 3'-PO_4 to a 3'-OH (Cameron, Uhlenbeck, 1977).

Figure 3.5: Polynucleotide kinase removes 3' terminal 2',3'-cyclic phosphates.

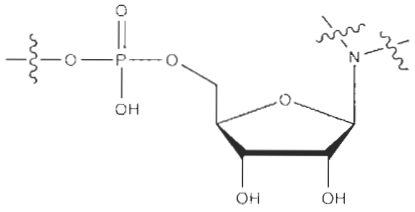
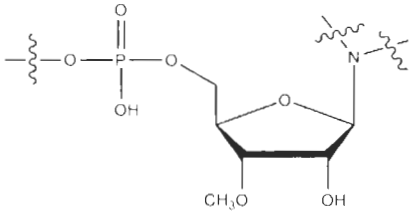
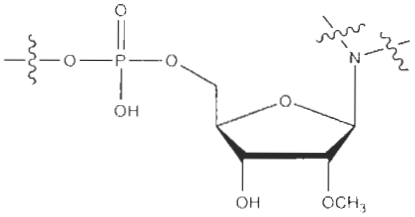
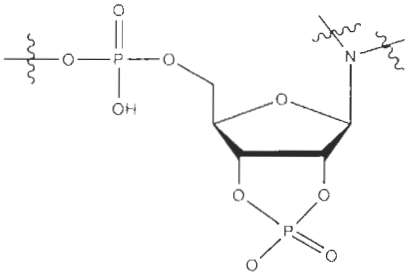
(A) The 24.11 RNA was targeted by 10-23 DNAzyme to be cleaved into two asymmetric products, 9 and 15 nts in length, respectively. A red arrow indicates the cleavage site. The 15 nt long fragment carries the 2',3'-cyclic phosphate on its 3' terminus, whereas the 9 nt fragment carries a 3' OH on its 3' end. (B) Following the cleavage of the 5' radiolabelled 24.11 RNA, only the undigested substrate and the 15 nt fragment carrying the 5' radiolabel are visible in a autoradiogram of the 23.75 % denaturing PAGE. The 15 nt fragment was then gel purified (panel B, t0) and subject to periodate treatment and β-elimination (panel B, t1). As seen in the autoradiogram, no periodate dependent cleavage is observable. (C) Following the cleavage of 24.11 RNA, the reaction was phosphorylated using [^{32}P]γ-ATP and polynucleotide kinase (PNK). Visible are most nucleic acids: 32 nt band is the DNAzyme, 24 nt band is the uncleaved 24.11 RNA, 15 nt band is the 5' cleavage product and the 9 nt band is the 3' cleavage product is not visible in this figure. The 15 nt band was gel purified (panel C, t0) and subjected to periodate treatment (panel C, t1). The periodate dependent cleavage is almost quantitatively.



Another ubiquitous modification of the RNA ribose is the methylation of the 2' hydroxyl. This modification was first detected in ribosomal RNA, especially in functional regions of the ribosome, and is thought to structurally stabilize the ribosomal RNA (Cunningham, et al., 1991). This 2'-O-methylation is facilitated by an RNA-protein complex containing small guide RNAs called C/D box snoRNAs (small nucleolar RNAs) (Omer, et al., 2000). This post-transcriptinal modification is catalyzed by the methyltransferase fibrillarin, which is guided to the target RNA by C/D box snoRNAs (Omer, et al., 2002).

Table 3.1: 3' terminal modifications on ribose of RNA.

Shown in this table are various modifications of the 2' and 3' position of the 3' terminal ribose. Each column represents an enzymatic or chemical assay and its ability to detect the 3' terminal ribose modification. Some experimental results are shown in Figure 3.8.

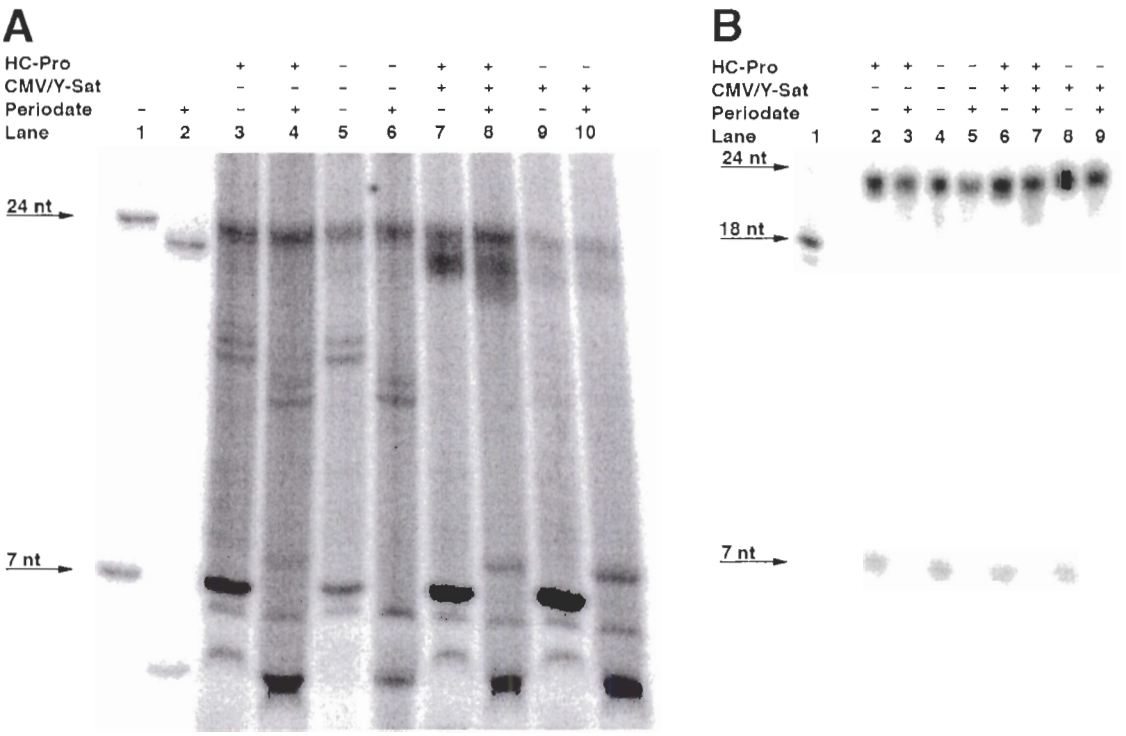
	$NaIO_4$	Ligase	PAP	PNK
	+++	+++	+++	—
	—	—	—	—
	—	++	—	—
	—			+++

Small RNA from all four plant types were first phosphorylated, effectively removing the 2',3'-cyclic phosphate, and then subjected to sodium periodate cleavage. A seven nucleotide unmodified synthetic RNA was added to each RNA sample as a control for successful periodate cleavage. Figure 3.6 A shows the autoradiogram of size separated periodate cleaved RNA samples. In all samples, the 7 nt control RNA was cleaved quantitatively and the 24 nt small RNA are mostly periodate resistant. In infected tobacco plants, the 21/22 nt small RNA was periodate resistant in HC-Pro(–) plants (Figure 3.6 A, lanes 9 and 10). In HC-Pro(+) plants, the 21/22 nt band showed sensitivity to periodate cleavage (Figure 3.6 A, lanes 7 and 8). To further demonstrate that only 21/22 nt are sensitive to periodate treatment, 24 nt small RNA from all four plant samples were gel purified and exposed to periodate as seen in Figure 3.6 B (lanes 3, 5, 7 and 9). This experiment confirmed, that only a small fraction (5 to 12 %) of 24 nt small RNA are periodate sensitive regardless of HC-Pro expression within the cell. The 7 nt control RNA cleaved quantitatively in all reactions.

The periodate cleavage assays on populations of small RNA, albeit by size, support the before mentioned hypothesis that not all subpopulations are equally modified on their 3' terminus, thus leading to the discrepancy observed during 3' adaptor ligation using T4 RNA ligase. In infected plants, the 21/22 nt RNA population is comprised of viral and endogenous small RNA. To analyze which subpopulation is sensitive to periodate cleavage, specific small RNA were isolated from the total population of small RNA. To isolate specific small RNA, 3' biotin-labelled DNA probes were designed antisense to commonly cloned 21/22 nt RNA, namely MIR166, MIR168, as well as sense and antisense probes against Y-Sat small RNA from region 173 to 193. In separate reactions, these antisense probes were annealed to total 5' radiolabelled small RNA. Following the annealing step, the mixture was incubated with streptavidin and all fragments were separated by size in a native PAGE (Figure 3.7 A). The shifted material contained MIR166, MIR168 for all four plant samples and Y-Sat small RNA (position 173 to 193) from infected plants. The small RNA were cut from the gel, eluted and tested for periodate cleavage sensitivity. The resulting autoradiograms of the denaturing 23.75 % PAGE is shown in Figure 3.7 B and C. Y-Sat small RNA were only sensitive to periodate in HC-Pro expressing infected plants (≈ 50 % Figure 3.7 B lanes 3 and 4; Figure 3.7 C lanes 7 and 8). On the contrary, Y-Sat small RNA from HC-Pro(–) CMV/Y-Sat(+) plants were not sensitive to periodate, nor were endogenous micro RNA MIR166 and MIR168 in either HC-Pro(–) or HC-Pro(+) plants (Figure 3.7 B lanes 7 to 14; Figure 3.7 C lanes 3 to 6).

Figure 3.6: Periodate cleavage of small RNA populations.

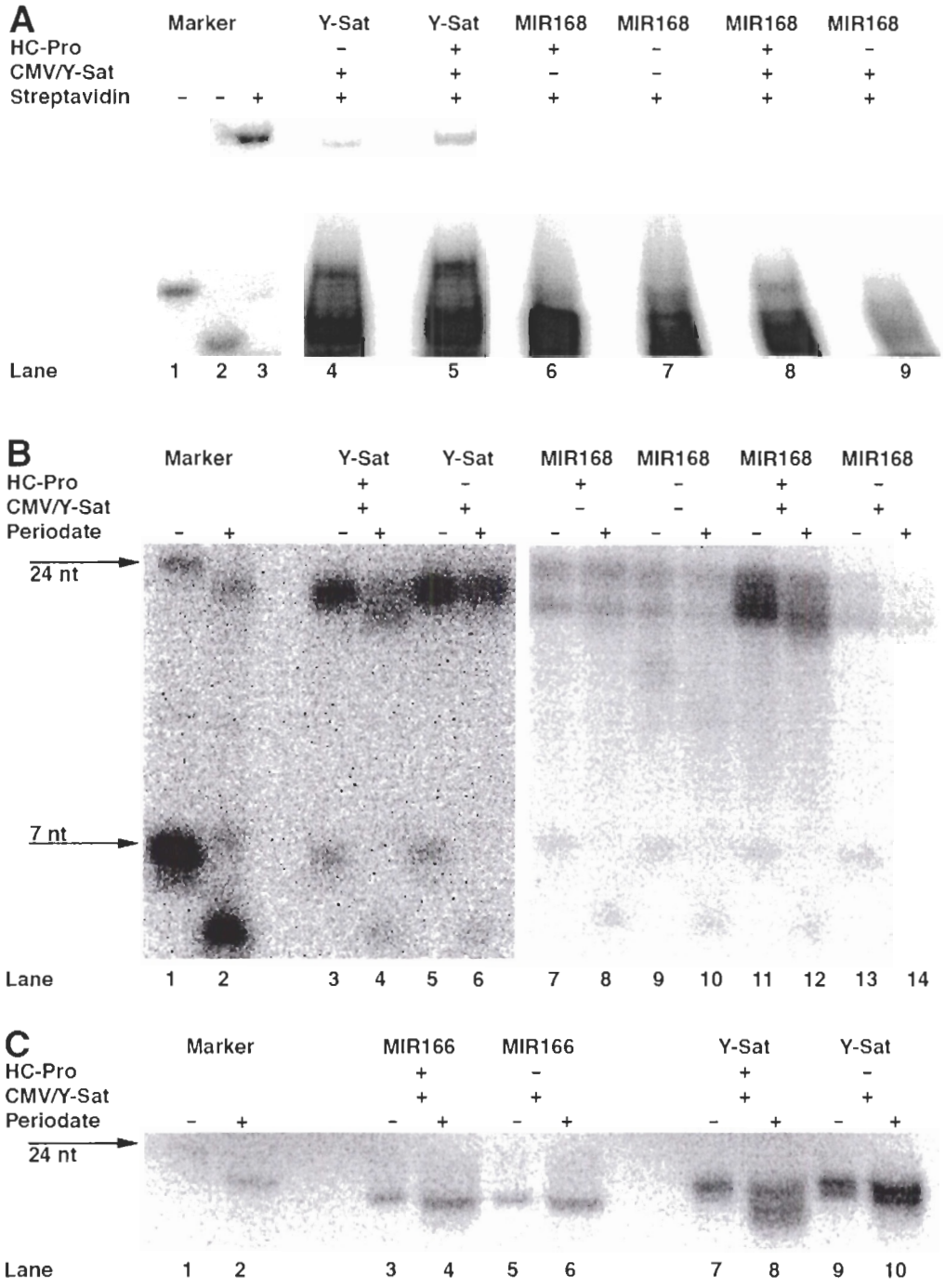
(A) Small RNAs were radiolabelled with PNK (lanes 3, 5, 7, and 9) and then subjected to periodate treatment (lanes 4, 6, 8, and 10). Only RNA from virally infected HC-Pro(+) plants showed significant sensitivity to periodate (lane 8). A synthetic 7-nt internal control RNA was added to all RNA fractions before chemical treatment to confirm the efficient periodate-mediated elimination of the terminal nucleoside from unmodified RNA. A 24-nt RNA marker reacts to completion when exposed to periodate, shifting downward by 2 nt in mobility (lanes 1 and 2). Note that the 15-nt double bands in the two uninfected samples are also fully sensitive to periodate (lanes 3-6).
(B) The 24-nt-long RNAs from both the infected and uninfected plant samples were gel-purified and subjected to periodate treatment separately. Quantification of cleavage products indicates that cleavage in uninfected plants is weakly dependent on HC-Pro expression [7% and 5% in lanes 3 and 5, respectively, after background subtraction using the (–) periodate lanes]. In infected plants, HC-Pro effects a 2-fold increase in periodate cleavage (12% and 6% in lanes 7 and 9, respectively).



These experiments confirm that within the total population of small RNAs from HC-Pro(+) CMV/Y-Sat(+) tobacco plants, a subpopulation is not modified on its 3' terminus. No other small RNA sample in the four-point comparison study showed a large discrepancy between the cloning data's histogram and their 5' radiolabelling autoradiogram. From these observations two conclusions can be drawn. First, all small RNAs are equally modified on their 3' terminus. Second, HC-Pro modulates the 3' terminal modification of viral small RNAs, a subpopulation of the total small RNA population.

Figure 3.7: Periodate cleavage of small RNA species.

(A) Purification of small RNAs with streptavidin-induced native gel shifts. A single-stranded 5'-radiolabelled oligonucleotide (lane 2) being annealed to an antisense 3'-biotin-tagged oligonucleotide only shifted marginally in the absence of streptavidin (lane 1). When streptavidin is added, a nearly quantitative and large upward gel shift occurs (lane 3) in a native 10 % polyacrylamide gel. Specific small RNAs (Y-Sat and MIR168) were therefore isolated by annealing radiolabelled small RNAs from plants to corresponding biotinylated antisense oligonucleotides, followed by separation using native polyacrylamide gel in the presence of streptavidin (lanes 4-9). Annealing of the labelled small RNAs with a random sequence did not result in a detectable gel shift, nor did annealing of small RNAs from uninfected plants with the Y-Sat probe (data not shown). The signal intensities of the RNA:DNA duplexes relative to those of the total small RNA population are 0.62, 1.17, 0.25, 0.24, 0.21, and 0.30%, respectively, for lanes 4-9. The same technique was used to isolate MIR166 (data not shown). (B) Periodate treatment of MIR168 and Y-Sat small RNAs. Lanes 1 and 2 are the 24- and 7-nt control RNA before and after periodate treatment. The 7-nt control RNA was added to all samples to serve as an internal control for periodate cleavage. Lanes 3-6 show Y-Sat small RNA from infected HC-Pro(+) or HC-Pro(–) plants treated with periodate (+) or untreated (–). Lanes 7-14 are MIR168 with or without periodate treatment. Periodate cleavage is negligible in all miRNA samples (infected HC-Pro(+) plants show minor cleavage). (C) Periodate treatment of MIR166 and Y-Sat small RNAs. MIR166 from infected HC-Pro(+) or HC-Pro(–) plants was substantially resistant to periodate (lanes 4 and 6). Only the Y-Sat small RNAs from HC-Pro(+) plants were sensitive to periodate (lane 8).


A
Marker
Y-Sat
Y-Sat
MIR168
MIR168
MIR168
MIR168
HC-Pro
CMV/Y-Sat
Streptavidin
Lane
1 2 3 4 5 6 7 8 9
B
Marker
Y-Sat
Y-Sat
MIR168
MIR168
MIR168
MIR168
HC-Pro
CMV/Y-Sat
Periodate
24 nt
7 nt
Lane
1 2 3 4 5 6 7 8 9 10 11 12 13 14
C
Marker
MIR166
MIR166
Y-Sat
Y-Sat
HC-Pro
CMV/Y-Sat
Periodate
24 nt
Lane
1 2 3 4 5 6 7 8 9 10

3.2.3 Modification of 2' or 3' hydroxyl?

Having established that most small RNA in plants are modified on their 3' terminus, two questions remained: (1) what is the modification and (2) is the modification on the 2' or 3' hydroxyl? The first question was answered on February 11, 2005 when Yu and colleagues published that micro RNA are methylated on their 3' terminus by HEN1 (Yu, et al., 2005). The research group observed a S-adenosyl methionine (SAM) binding motif within HEN1 and were able to show *in vitro* the transfer of a [^{14}C] radiolabelled SAM onto a micro RNA. Together with our data, it builds a strong case that all small RNA are methylated on their 3' terminus.

This leaves the second question if the modification is on the 2' or 3' hydroxyl. Our cloning data showed that the methylated small RNA are able to be ligated. This suggests that the 3' hydroxyl is not modified, as both T4 RNA ligase and Poly-(A) polymerase strongly prefer a 3' hydroxyl to form a phosphodiester bond between adjacent nucleotides. To explore how efficient a methylated 2' hydroxyl substrate is for ligation with T4 RNA ligase and extension using poyl(A)-polymerase, two seven nucleotide RNA molecules were synthesized. These two synthetic RNA molecules were identical in sequence and differed only in their 3' terminus modification, 2'-hydroxyl (2'-OH) or 2'-O-methyl (2'-O-Me). In an extension experiment, poly(A) polymerase was not able to extend the 2'-O-Me substrate, even though the enzyme was active as seen by the extension of an internal unmodified 126 nt control RNA (Figure 3.8 A lanes 3 and 4). Poly(A) polymerase extended rapidly the 2'-OH synthetic RNA after one hour (Figure 3.8 A lanes 1 and 2). T4 RNA ligase was able to append the adenylylated DNA (App16.5) to the 2'-O-Me RNA, but with lower efficiency (39 % ligation yield after one hour versus 86 % using 2'-OH substrate). Thus, an extended ligation would facilitate ligation of 2'-O-Me modified small RNA.

3.3 Concluding remarks

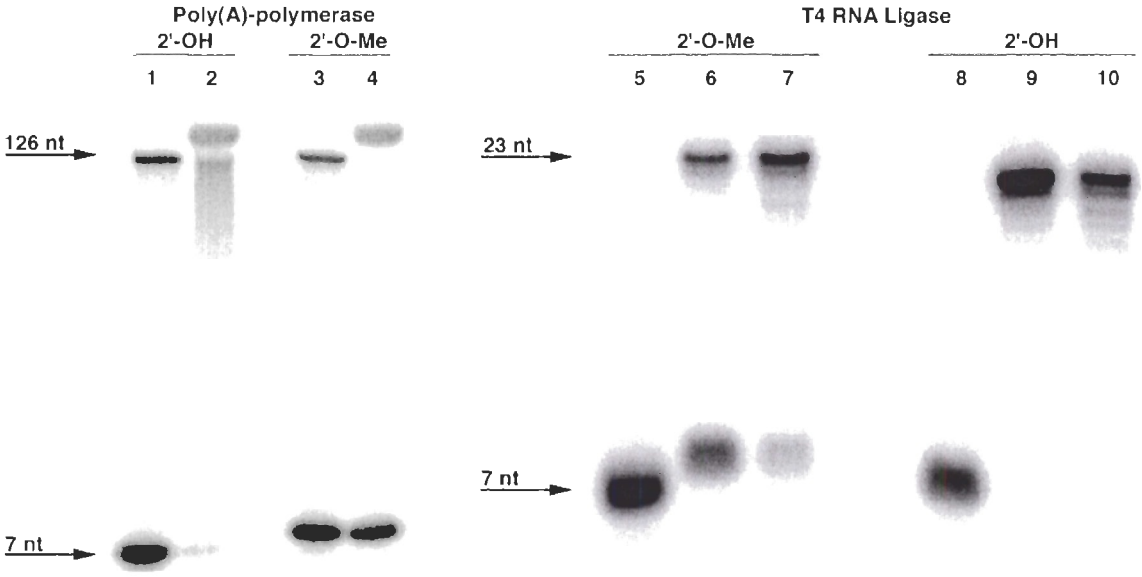
The experiments presented in this chapter clearly demonstrate that all small RNAs in plants are modified on the 2' hydroxyl of their 3' terminus. Together with *in vitro* evidence (Yu, et al., 2005) this modification is most likely a 2'-O-Me. This discovery was made while investigating the small RNA populations from HC-Pro expressing CMV/Y-Sat infected tobacco plants in a four point comparison study. As the majority of mature small RNAs are methylated on their 2' hydroxyl, the cloning efficiency is lower when compared to an unmodified 2' hydroxyl. When HC-Pro interfered with a sub-population of small RNAs, namely the viral siRNAs, the ligation efficiency of the unmodified small RNAs was greater than the ligation efficiency of modified small RNA. Thus, the discrepancy between the histogram of HC-Pro(+) CMV/Y-Sat(+)

and the autoradiogram of the 5' radiolabelling. Further implementations of this discovery are discussed in Chapter 5.

Thus far, the biochemical aspects of small RNA populations were discussed and the small RNA cloning data analyzed globally. The following chapter will discuss the results from the small RNA cloning project in more detail.

Figure 3.8: RNA extension assays.

(Left) RNA containing a terminal 2'-O-Me is resistant to extension by poly(A) polymerase. A synthetic 7-nt RNA containing a 2'-O-Me was completely resistant to extension when incubated with ATP and poly(A) polymerase (0- and 1-h time points, lanes 3 and 4); an unmodified control RNA having the same sequence was nearly completely extended in 1 h (0- and 1-h time points, lanes 1 and 2). A control 126-nt unmodified RNA is efficiently extended by poly(A) polymerase in both experiments. (Right) T4 RNA ligase recognizes 2'-O-Me modified RNA as an acceptor substrate. T4 RNA ligase ligates a 7-nt 2'-O-Me RNA to an adenylated DNA oligonucleotide, resulting in a 23-nt RNA-DNA hybrid (0-, 1-, and 19-h time points, lanes 5-7). This ligation proceeded 2-3 times slower than with the 7-nt unmodified control RNA having the same sequence (lanes 8-10).



4 SMALL RNA GROUPS REVEALED BY *EBBIE* ARE STUDIED USING A NOVEL STREPTAVIDIN GEL-SHIFT ASSAY

As discussed in Chapter 2, one of *Ebbie's* strength is a dynamic BlastN database, which allows for rapid comparison of newly entered small RNA sequences to all previously cloned small RNAs of the same cloning project (Ebhardt, et al., 2006). This feature allows for rapid identification of high-abundant small RNA sequences termed groups. Following the discovery of small RNA groups, the question remains if these groups are degradation products of larger RNAs or if they are small RNAs associated with RNA silencing. As discussed in Chapter 3, the majority of siRNAs and miRNAs are modified on their 3' terminus with a 2'-O-Me. Historically, it was believed that 21 to 25 nt long RNAs are degradation products of endo- or exonucleolytic activity.

Today we know that both statements are true and, at least in plants, one can distinguish small RNAs associated with the RNA silencing pathway and small RNAs originating from endo- or exonucleolytic nuclease activity by examining their 3' terminus modification. It should be noted that even if the small RNA of interest is not modified on its 3' end, it only means that it is not generated by any RNA silencing pathway. It does not exclude any biological role for these small RNAs, especially if they occur in large quantities, e.g. *rRNA group 01*.

To elucidate the terminal modification of some of these groups, I developed a novel streptavidin gel-shift assay, which isolates specific small RNAs from a pool of small RNAs and allows further biochemical analysis. Compared to Northern hybridization, the streptavidin gel-shift assay has three main advantages. First, the number of small RNA probes per experiment is limited by the lanes in the gel and the available probes themselves. Northern hybridization experiments limit the number of probes to one per hybridization. Secondly, one streptavidin gel-shift assay is required to quantify relative ratios of a particular small RNA to the total population of small RNAs. In a Northern hybridization experiment, at least two consecutive hybridizations are required to estimate a relative amount to a known non-coding RNA, e.g. 5S rRNA. Thirdly, the streptavidin gel-shift assay does not require any transfers of gels to nylon membranes. The autoradiogram is directly generated by the native gel minimizing signal diffusion effects typical of Northern autoradiograms.

In this chapter I will first discuss various groups discovered by *Ebbie*, then discuss the streptavidin gel-shift assay and finally present small RNAs associated with CMV/Y-Sat infection of tobacco plants.

4.1 Small RNA groups identified by *Ebbie*

Once a small RNA sequence was isolated from a DNA sequencing file, the clone was compared to the dynamic database of previously cloned small RNAs using BlastN. The BlastN output was printed onto the screen of the dynamic web page (see Chapter 2, Figure 2.4) and the user is able to interpret the comparison search results. To minimize the annotation of data and maximize the information content, small RNA clones with significant overlap were annotated as groups. A small RNA group was defined to be at least two sequences overlapped by 12 or more consecutive base pairs and resulted in 31 groups. This empirical cut-off proved very effective, as a 16 bp cut off would have resulted in 30 groups. The percentage overlap score from BlastN was not chosen as cut-off criteria, as BlastN might only align parts of two sequences but still give a 100 % score for this partial alignment.

Out of 698 small RNAs 213 were of viral origin and are discussed below. Out of 485 non-viral sequences, 120 were clustered into 31 groups. These 31 groups can be further classified into five clades: NULL-clade, micro RNA or silencing RNA clade, rRNA clade, tRNA clade and a single group of a messenger RNA clade.

As seen in Table 4.1, there is a large group of small RNA sequences that could not be annotated. Even more fascinating is the fact that in this NULL-clade nine of ten groups are comprised of identical small RNA sequences, except group 22. For example, the 18 nt long sequence termed group 08 was cloned 12 times out of 485 non-viral sequences (2.5 %) yet failed to show any significant hit when compared to the GenBank database from NCBI (BlastN search performed in February 2007; low complexity filter OFF, Viridiplantae [ORGN], otherwise default parameters). It would be interesting to compare this NULL-clade to other large scale sequencing projects using 454-sequencing (Margulies, et al., 2005), as some of these groups are only comprised of 2 members. Even when only considering the larger groups however, four out of five groups in the NULL-clade are comprised of identical sequences, which can not be said about most silencing RNAs.

Most groups within the silencing and micro RNA clade were identified by *Ebbie* using a locally installed database of miRNAs (version 8.0). Most members of the silencing RNA clade were assigned this annotation recently as large-scale sequencing data became available from Qi

and colleagues (Qi, et al., 2006). Qi and colleagues sequenced small RNAs associated with *Arabidopsis* ARGONAUTE4 protein, a key component of RITS, thus linking it to a RNA silencing pathway which silences transposable and repetitive DNA elements.

Reading other small RNA cloning project reports, most research groups discard any small RNA originating from ribosomal or transfer RNA (Lau, et al., 2001; Rajagopalan, et al., 2006; Xie, et al., 2004). The first group ever identified by *Ebbie*, was rRNA group 01. rRNA group 01 is the 5' start of 18S rRNA and was cloned 14 times of which 7 are identical sequences and the remainder were heterogeneous in length. 12 of 14 sequences start at the 5' position of the 18S rRNA which is an intriguing fact, as the 18S rRNA is 1,768 bp long (GenBank: AJ236016). Is this a pathway for ribosomal degradation or regulation independent of the RNA silencing machinery? Two other groups annotate with 18S rRNA: rRNA group 18 and rRNA group 27 (see Table 4.1). Other rRNA groups, such as group 11 was annotated as the 5' start 5S rRNA, whereas groups 16 and 19 were identified as the 3' end of 25S and 5S rRNA, respectively.

One sequence that was cloned twice is annotated as a piece of messenger RNA – mRNA group 06. This is a heterogeneous group and might be a mRNA degradation product. Alternatively, while analyzing 454 deep sequencing data the Bartel lab (Massachusetts Institute of Technology) found a large number of small RNAs matching mRNAs that overlap each other, which are being investigated further (Bartel, DP personal communications).

Table 4.1: Small RNA groups.

This table summarizes all groups identified by *Ebbie*. The first column describes the name of the group and subgroups, e.g. all rRNA small RNA groups were sub-classified as rRNA-GroupN. The second column called Members provides a number of how many members each group has and if the members are homogeneous (hom) or heterogeneous (het) in length. Cloned sequence is typically the longest cloned sequence of the group and the last column summarizes the annotation of the small RNA group. The table is sorted by classes, such as rRNA-groups and siRNA-groups. Group 1 and 11 represent the 5' of 16S rRNA and 5S rRNA, respectively. Group 16 and 19 are the 3' terminus of 25S rRNA and 5S rRNA, respectively. Note: 'NULL' means no 100 % full length annotation was possible.

Name	Members	Cloned sequence	Annotation
group 02	2 – hom	TGCTTGTTTTCATCAAATGTCCAC TGCTTGTTTTCATCAAATGTCCAC	NULL
group 03	5 – hom	TTGATACGCACCTGAATCGGC TTGATACGCACCTGAATCGGC TTGATACGCACCTGAATCGGC TTGATACGCACCTGAATCGGC TTGATACGCACCTGAATCGGC	NULL, possibly mRNA, antisense
group 04	2 – hom	GTGTATTTTTGTCGTAATTGCTCC GTGTATTTTTGTCGTAATTGCTCC	NULL

Name	Members	Cloned sequence	Annotation
group 05	2 – hom	AACACAGGTAACCGATAACTATCT AACACAGGTAACCGATAACTATCT	NULL
group 07	3 – hom	TGGCAACTTCTTTATCATGCC TGGCAACTTCTTCATCATGCC TGGCAACTTCTTCATCATGCC	NULL
group 08	12 – hom	GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGGTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG GTCGGAGATAGGCTGAAG	NULL
group 13	2 – hom	ATACTTCGTACACGGGTCTACCC ATACTTCGTACACGGGTCTACCC	NULL
group 22	3 – het	TCAAATCTGGTTAAGCTAGCACGA TCCATCTGGTTAAGCTAGCACGA TCCACTCTGGCTAAGCTAGCACGA	NULL
group 24	2 – hom	TCTTCTTTGCAAAATCAAGAA TCTTCTTTGCAAAATCAAGAA	NULL
group 26	2 – hom	CTCGTCTGTTCCGCACTTGATAAG CTCGTCTGTTCCGCACTTGATAAG	NULL
mRNA-group 06	2 – het	CTGTGGTCGTGAAGCACCGAC CTGTGGTCGTGAAGCAC	mRNA
miRNA-group 29	2 – hom	TGCCTGGCTCCCTGTATGCCA TGCCTGGCTCCCTGTATGCCA	MIR160
miRNA-group 30	12 – het	TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCCTC TCGGACCAGGCTTCATTCCCC TCGGACCAGGCTTCATTCC TCGGACCAGGCTTCATTCC TCGGACCAGGCTTCATTCC GGACCAGGCTTCATTCC	MIR166
miRNA-group 31	5 – hom	TCGCTTGGTGCAGGTCGGGAC TCGCTTGGTGCAGGTCGGGAC TCGCTTGGTGCAGTTCGGGAC TCGCCTGGTGCAGGTCGGGAC TCGCTTGGTGCAGGTCGGGAC	MIR168
mRNA-group 06	2 – het	CTGTGGTCGTGAAGCACCGAC CTGTGGTCGTGAAGCAC	mRNA, sense

Name	Members	Cloned sequence	Annotation
rRNA-group 01	14 – het	TACCTGGTTGATCCTGCCAGTAGTC TACCTGGTTGATCCTGCCAGTAGTC TACCTGGTTGATCCTGCCAGTAGTC TACCTGGTTGATCCTGCCAGTAGTC TACCTGGTTGATCCTGCCAGTAGTC TACCTGGTTGATCCTGCCAGTAGTC TACCTGGTTGATCCTGCCAGTAGT TACCTGGTTGATCCTGCCAGTAG ACCTGGTTGATCCTGCCAGT TACCTGGTTGATCCTGCCAG TACCTGGTTGATCCTGCCA TACCTGGTTGATCCTGCC TACCCGGTTGATCCTGCC TGATCCTGCCAGTA	18S rRNA (pos.: 1..24)
rRNA-group 09	2 – hom	GAGTAACTATGACTCTCTTC GAGTAACTATGACTCTCTTC	25S rRNA (pos.: 2235..2254)
rRNA-group 11	4 – het	GGATGCGATCATACCAGC GGATGCGATCATACCAGC GGATGCGATCATACCAG GGATGCGATCATACCAGCACTAAC	5S rRNA (pos.: 1..20)
rRNA-group 16	7 – het	CAGCCCTTTGTCGCTCCGATTCGT AGCCCTTTGTCGCTCCGATTCGT AGCCCTTTGTCGCTCCGATTCGT AGCCCTTTGTCGCTCCGATTCGT TTTGTCGCTCCGATTCGT GCCCTTTGTCGCTCCGATTCGT	25S rRNA (pos.: 3359..3381)
rRNA-group 18	5 – het	TCTCATGGAGAGTTCGATCCTGGCT TCTCATGGAGAGTTCGATCCTGGCT TCTCATGGAGAGTTCGATCCTGGCT TCTCATGGAGAGTTCGAT GTTCGATCCTGGCTCAGGATGAA	18S rRNA (pos.: 116..137)
rRNA-group 19	4 – het	GGAAGTCCTCGTGTTGCATCCCT AGTCCTCGTGTTGCATCCCT TCCTCGTGTTGCATCCCT GAAGTCCTCGGGTTGCATCCCT	5S rRNA (pos.: 295..316)
rRNA-group 27	2 – hom	CCAGACATAGTAAGGATTGAC CCAGACATAGTAAGGATTGAC	18S rRNA (pos.: 1210..1230)
siRNA-group 10	6 – het	TATTCTGGTGTCCTAGGCGTAGAG TATTCTGGTGTCCTAGGCGTAGA TATTCTGGTGTCCTAGGCGTA TATTCTGGTGTCCTAGGCGT TATTCTGGTGTCCTAGGCGTA TATTCTGGTGTCCTAGGCGTAG	conserved intergenic sequence
siRNA-group 14	2 – het	TTCCACAGCTTTCTTGAACTG TTCCACAGCTTTCTTGAAC	siRNA associated with AGO1
siRNA-group 15	2 – hom	CGTTTCCGGACTCGTCC CGTTTCCGGACTCGTCC	Internal Transcribed Spacer 1
siRNA-group 17	3 – het	GTCTGGGTGGTGTAGTTGGTT GTCTGGGTGGTGTAGTTGGTT GTCTGGGTGGTGTAGTC TCTGGGTGGTGTAG	conserved intergenic sequence

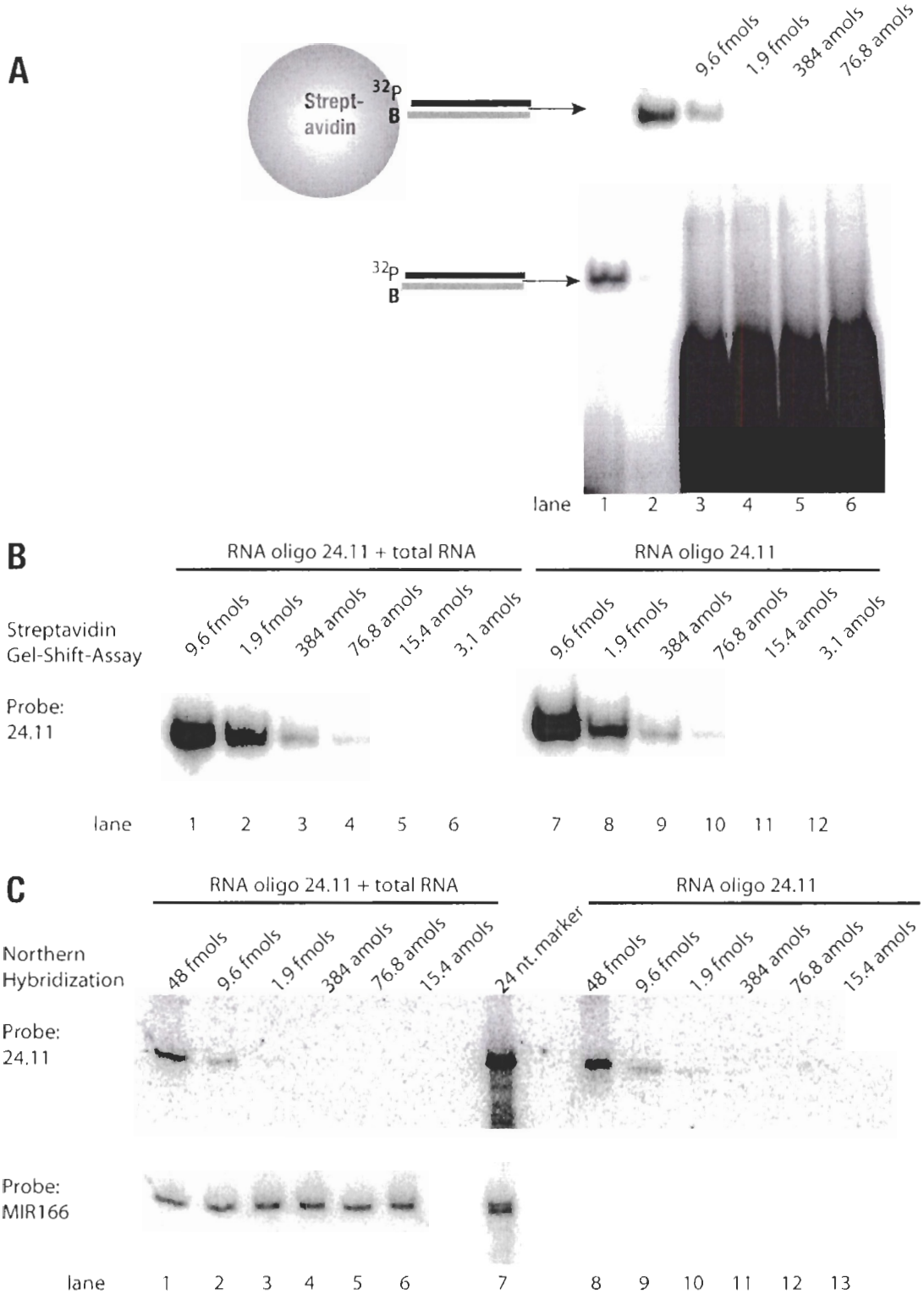
Name	Members	Cloned sequence	Annotation
siRNA-group 20	4 – het	GGGGATGTAGCTCAAATGGTAGG GGGGATGTAGCTCAGATGG GGGGATGTAGCTCAGATGG GGGGATGTAGCTCAGATGG	conserved intergenic sequence
siRNA-group 21	3 – het	GGGGTGGTGGCGCAGTTGGC GGGGTGGTGGCGCAG GGGGTGGTGGCGCAG	ago associated siRNA, conserved intergenic sequence
siRNA-group 25	2 – het	AACTTTACCAGTTGCGCCAAGGCT ATTTTACCAGTTGCTCACCTTC	conserved intronic sequence
tRNA-group 12	8 – het	GTCGTTGTAGTATAGTGG GTCGTTGTAGTATAGTGGTGA GTCGTTGTAGTATAGTGGTA GTCGTTGTAGTATAGTGGT GTCGTTGTAGTATAGTGGT GTCGTTGTAGTATAGTGGTGAGT GTCGTTGTAGTATAGTGGTAAGTAT GTCGTTGTAGTATAGTGGTAAGTA	5' start of tRNA- Asp
tRNA-group 23	2 – het	GGTTCTATGGTGTAGTGGTTAGCA AGTTCTATGGTGTAGTGG GCGCATCTGGTGTAGTGG	5' start of tRNA-Gln
tRNA-group 28	2 – het	GCGCTCTTAGTTCAGTTCGGTAG GCGCTCTTAGTTCAGTTCGG	5' start of tRNA-Trp

4.2 Streptavidin gel-shift assay

To isolate specific small RNAs from a pool of small RNAs, a 3' biotin-labelled DNA with a complementary sequence was synthesized. The probe and a total population of 5' radiolabelled small RNAs were heated to 90 °C for 90 seconds in a neutral annealing buffer containing 50 mM NaCl. This mixture was slowly cooled-down allowing the probe to anneal to the specific small RNA species of interest. Once room temperature was reached, streptavidin was added. Streptavidin and biotin form the strongest natural non-covalent bond with a K_d of 48 fM. In a native PAGE, this RNA-DNA-streptavidin complex migrates significantly slower than double-stranded RNA. Depending on the purity of the size fractionated small RNA preparation, a 1 fmole detection limit is easily achievable (see Figure 4.1 panel A). In an ideal situation where only one species of RNA is 5' radiolabelled, the gel-shift assay can reach a sensitivity of 15 atto moles, regardless of additional nucleic acids present in the mixture (compare Figure 4.1, panel B, lanes 5 *versus* lane 11). The sensitivity of a Northern hybridization under the same conditions in our hands show a detection limit of 10 fempto moles, roughly an order of magnitude less sensitive (compare Figure 4.1, panel B, lanes 5 and 11 *versus* Figure 4.1, panel C, lanes 2 and 9).

Figure 4.1: Comparing sensitivity of Streptavidin gel-shift assay to Northern hybridization.

(A) In an autoradiogram of a 10 % native PAGE, a complementary 3' biotin-labelled probe anneals to a 5' radiolabelled RNA (lane 1). The annealed duplex migrates significantly slower when it is bound to streptavidin (lane 2). Lanes 3 to 6 are serial dilutions of a synthesized RNA that is added to a pool of size fractionated small RNAs prior to appending the 5' radiolabel. Following the phosphorylation reaction, the 3' biotin-labelled probe is annealed and streptavidin added. The assay under these conditions detects at least 1.9 fmoles of radiolabelled RNA (lane 4). (B) Autoradiogram of a 10 % native PAGE where 5' radiolabelled RNA is titrated in 5 fold dilutions with (lanes 1 to 6) or without (lanes 7 to 12) non-radiolabelled RNA in the test tube. The sensitivity of the streptavidin gel-shift assay under these ideal conditions is limited to the detection limit of the phosphorimaging system, which is between 76 and 15 atto moles. (C) Autoradiogram showing a Northern hybridization experiment under identical conditions to (A) and (B). In panel (C) on top, a probe for the artificial RNA is added and in the bottom panel a probe for MIR166 was added. The sensitivity of the Northern is at least 10 to 50 fold lower when compared to the streptavidin gel-shift assay.


A
Strept-
avidin
32P
B
9.6 fmols
1.9 fmols
384 amols
76.8 amols
lane 1 2 3 4 5 6
B
RNA oligo 24.11 + total RNA
RNA oligo 24.11
Streptavidin
Gel-Shift-Assay
9.6 fmols
1.9 fmols
384 amols
76.8 amols
15.4 amols
3.1 amols
Probe:
24.11
lane 1 2 3 4 5 6 7 8 9 10 11 12
C
RNA oligo 24.11 + total RNA
RNA oligo 24.11
Northern
Hybridization
48 fmols
9.6 fmols
1.9 fmols
384 amols
76.8 amols
15.4 amols
24 nt. marker
Probe:
24.11
Probe:
MIR166
lane 1 2 3 4 5 6 7 8 9 10 11 12 13

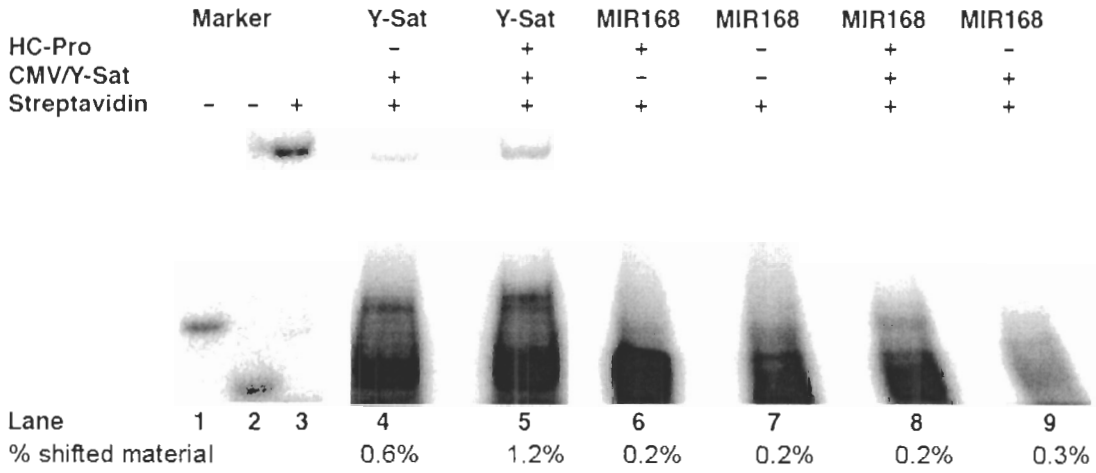
4.2.1 Validating small RNA ratios using streptavidin gel-shift assay

While cloning small RNAs provides the researcher with primary sequence information about the small RNA, the cloning protocol dictates the use of poly(A)polymerase (PAP) or T4 RNA ligase to extend the 3' terminus. As discussed in Chapter 3, PAP is not suitable if the small RNA has a modified 3' terminus. Even T4 RNA ligase requires longer incubation times to achieve the same ligation efficiency when the 2' hydroxyl is modified with a methyl group. For these reasons, it is essential to validate the cloning data and compare the ratio of cloned small RNAs of any given species to an assay that is independent of the 3' terminus modification.

Figure 4.2 shows the autoradiogram of a streptavidin gel-shift assay probing for exogenous Y-Sat small RNA (pos.: 173-193) and the endogenous micro RNA (MIR168) in tobacco plants using in our study. Once Typhoon Strom 820 scanned in the autoradiogram, it was analyzed using Molecular Dynamics Image Quant. The results show that the endogenous MIR168 was evenly expressed in all four tobacco plants in the four-point comparison study at 0.2 % of the total population of small RNAs (Figure 4.2, lanes 6 to 9). In CMV/Y-Sat infected tobacco plants, Y-Sat small RNAs from the yellowing region (Y-Sat pos.: 173-193) were detected. In HC-Pro(–) CMV/Y-Sat(+) plants, 0.6 % of the total population of small RNAs from the Y-Sat region were probed (Figure 4.2, lane 4). In HC-Pro(+) CMV/Y-Sat(+), this ratio doubles to 1.2 % (Figure 4.2, lane 5). The observation of more small RNAs from Y-Sat in HC-Pro(+) plants than HC-Pro(–) plants in the Streptavidin gel-shift assay was supported by earlier experiments by Wang and colleagues (Wang, et al., 2004).

Figure 4.2: Comparing small RNA ratios of exogenous and endogenous small RNAs.

Autoradiogram of a 10 % native gel: lanes 1 to 3 show a 5' radiolabelled RNA oligo with its complementary DNA (lane 1), single stranded 5' radiolabelled RNA oligo (lane 2) and 5' radiolabelled RNA with its complementary 3' biotin labelled DNA oligo plus streptavidin (lane 3). Lanes 4 and 5 probe for sense of Y-Sat (pos. 173-193) in HC-Pro(–) and HC-Pro(+) plants. Lanes 6 to 9 probe for MIR168 in all tobacco plants in the four-point comparison study. The percentages of material shifted to the total material are given below in percent.



4.2.2 Relative ratio of a particular small RNA to the total population of small RNAs

The streptavidin gel-shift assay has two major advantages over Northern hybridization experiments. In Northern hybridization experiments each autoradiogram only gives information about a single hybridization probe. In contrast, in a streptavidin gel-shift assay autoradiogram the number of probes is limited to the number of lanes in a polyacrylamide gel. In Figure 4.3, for example, ten small RNA probes were simultaneously analyzed. The same experiment using the Northern hybridization experiment would have taken at least ten times longer, as only one radiolabelled probe per hybridization can be used.

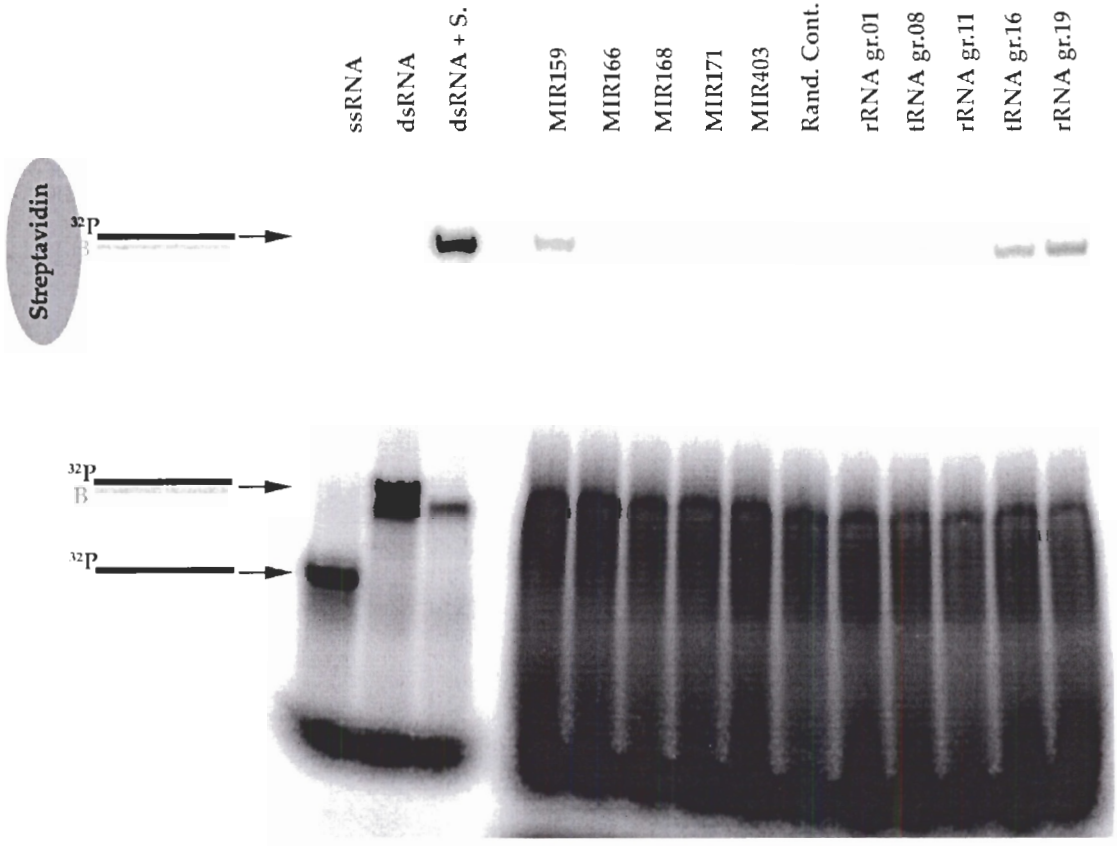
Using the streptavidin gel-shift assay ratios of all probed small RNA the total amount of small RNA can be quantified directly from the autoradiogram. This allows for a rapid and more precise determination of relative amounts of certain small RNAs. Traditionally, in a Northern hybridization experiment, the Northern membrane would be probed for a known RNA, e.g. 5S rRNA. After this hybridization, the same membrane would be stripped and re-probed with a small RNA probe. Finally, the intensities of the two hybridization experiments are compared to each other and a relative concentration inferred.

Figure 4.3 shows in lanes 1 the single stranded radiolabelled synthetic RNA, annealed to

its complementary 3' biotin labelled probe (Figure 4.3, lane 2) and the complex is shifted by streptavidin (Figure 4.3, lane 3). Lanes 4 to 8 show gel-shifts with various endogenous micro RNA probes in adult leaves from wild-type tobacco. MIR166 does not show a strong signal in mature leaves when compared to RNA isolated from young leaves in our four-point comparison study. This observation is in concert with Jung and colleagues who established that the MIR165/MIR166 family is involved in lateral meristem formation, leaf polarity, and vascular development (Jung, Park, 2006). Lane 9 probes the total small RNA with a randomized control. Lanes 10 to 14 show the streptavidin-shift when using complementary probes for various small RNA groups identified by *Ebbie*, as discussed above.

Figure 4.3: Validating ratios of individual small RNAs to the total small RNA population.

Autoradiogram of a 10 % native gel: lanes 1 to 3 demonstrating the stages of the streptavidin gel-shift assay: single stranded radiolabelled RNA, RNA annealed to biotin-labelled probe and streptavidin annealed to the DNA-RNA duplex. Using total size fractionated 5' radiolabelled RNA various probes (micro RNAs, a negative control and various rRNA groups) were annealed and streptavidin added.



4.3 Small RNAs from CMV and its Y-Satellite

From 422 small RNAs cloned from CMV/Y-Sat(+) plants, 213 were of viral origin, of which 48 from the tripartite CMV genome and 165 from Y-Satellite. As mentioned in the introduction, the CMV helper virus provides the Y-Sat with essential proteins required for replication, e.g. RdRp, whereas the Y-Sat genome does not contribute any proteins for viral amplification. Thus, it might not be surprising that only 48 small RNAs originated from the 8,623 nt long CMV tripartite genome. In contrast, Y-Sat does not encode any protein coding regions and 165 cloned small RNA were mapped to its 369 nts long genome. To underline the ratios of small RNAs mapping to the helper virus genome *versus* the satellite genome a simple calculation could be made: if each of the 48 or 165 small RNAs were 21 nts in length, then the CMV tripartite genome was only covered 0.11 times, whereas the Y-Sat genome has a 9.4 times coverage.

4.3.1 Small RNAs from CMV RNA

All small RNAs annotated to CMV have a length between 20 and 24 nts (Figure 3.2). The majority, 36 of 48 CMV small RNAs, originated from CMV RNA III. 19 of 36 small RNAs originate from the (+) strand of CMV RNAIII, whereas 7 of 36 originate from the (–) strand. The fact that there are small RNAs from both strands suggests that both (+) and (–) CMV RNA are generated within the infected plant cell. The model of rolling-cycle-replication postulates that the original (+) strand is converted into a circular (–) strand from which the RdRp generates a continuous (+) strand, which is subsequently cleaved into single units of (+) strands. Since more than 1/3 of all small RNAs from CMV RNAIII originates from the (–) strand suggests a greater role for the (–) strand.

When aligning all 36 small RNAs to the CMV RNAIII, it becomes apparent that ten small RNAs from the (+) and (–) strand do not align 100 % to a given CMV RNAIII isolate. Six of ten are misaligned at the end, whereas four of ten have a mismatch elsewhere. A possible technical explanation is that during the cloning process the terminal bases were exchanged or damaged, thus the final DNA sequencing read does not reflect the actual sequence. More interesting are the examples of small RNAs with internal mismatches. A simple technical argument: the tobacco plants were not infected with a pure infectious agent, but rather with a mixture of two or more CMV strains. A more intriguing biological argument could be made for RNA editing of these small RNAs from CMV. Both arguments are plausible, the latter more intriguing. Other large-scale sequencing projects also have difficulties aligning all small RNAs to

the genome (Qi, et al., 2006; Rajagopalan, et al., 2006).

When comparing CMV small RNAs to each other, four groups of two show a complete sequence overlap. Five groups of two and a group of three sequences show a partial overlap in their sequence. As these are very poor statistics, it is not immediately clear if these are functional or non-functional degradation products from CMV RNA cleavage by a Dicer-like RNase III.

Figure 4.4: Distribution of small RNAs from CMV RNAIII.

This figure is drawn to scale. The long black line represents the 2,216 nts long CMV RNAIII with the two green lines with arrows representing the capsid and 3a proteins. The red lines are the small RNAs that match the CMV RNAIII, where lines drawn on the top are sense; the lines on the bottom are antisense. The data suggests that the small RNAs are distributed almost evenly across the CMV RNAIII genome, with a slight enrichment in the region between 500 and 800.

4.3.2 Small RNAs from Y-Sat

As expected in CMV/Y-Sat infected tobacco plants, a large fraction of small RNA mapped the 369 nt long Y-Sat genome, as seen in Figure 4.5 A. The Y-Sat genome and numbering is shown in black letters. Small Y-Sat RNA from HC-Pro(+) CMV/Y-Sat(+) plants are blue. In red and orange are small Y-Sat RNA from HC-Pro(–) CMV/Y-Sat(+) plants. Bars above the Y-Sat genome represent sense small RNA, whereas bars below represent antisense small RNA sequences. 57 % (131/230) of all cloned small RNA are from Y-Sat in HC-Pro(+) CMV/Y-Sat(+) plants, in HC-Pro(–) CMV/Y-Sat(+) plants this number is 23 % (34/150). These cloning data agree well with the previous Northern hybridization data (Wang, et al., 2004) as well as with the data presented in this thesis.

Y-Sat small RNA from HC-Pro(–) infected plants are clustered along the Y-Sat genome. 76.5 % (26/34) of these small RNA are from the sense strand of the satellite. There is an intriguing triplicate 21 nt clone annotated to position 129 to 140, suggesting a high expression of this particular small RNA. There are also five cloned small RNA sequences spanning the region (181 to 202) of Y-Sat, which has been mapped to cause yellowing in tobacco (Roossinck, et al., 1992). The largest cluster of sense small RNA with nine members is located between position 110 and 150. There are certain regions of the Y-Sat genome from which no small RNA were cloned. This could be due to the sample size of only 34 small RNA from Y-Sat (with an average length of 21.58 nt, the 369 nt Y-Sat genome was only covered twice).

It is also interesting to note that no small RNA overlaps the end/start of the 369 nts long

genome. Two small RNAs start at position 1 whereas a small RNA's 3' terminus matches the satellite's 3' terminus. With a small RNA cloning intensity covering the Y-Sat genome 9.4 times, it is very likely that there is a biological explanation for this finding. Our small RNA cloning data suggests that the Y-Sat small RNAs are derived from the matured and correctly folded single stranded Y-Sat RNA and not from the double stranded RNA during the rolling-cycle amplification. The secondary structure of the Y-Sat RNA might be important for the correct generation of viral siRNA by DCL4 or DCL2 cleavage.

In HC-Pro(+) plants Y-Sat small RNA are clustered among the whole Y-Sat genome, covering the entire genome. This could be due to the higher coverage (average length 21.47 nt, the 369 nt Y-Sat genome was covered 7.6 times) or other secondary effects (as discussed later). There are two small RNA starting at position 1 of the Y-Sat genome and a small RNA ending in position 369. Thus, it appears that these small RNA are generated from a single-repeat Y-Sat genome. In HC-Pro(+) CMV/Y-Sat(+) plants, more small RNA were from the antisense strand 33.6 % (44/131) when compared to 23.5 % in HC-Pro(–) infected plants. The yellowing region (position 181 to 202) is covered by a cluster of 16 sense strand small RNA clones mostly 21 nt in size and four antisense 21 nt RNA. All clusters appear to be staggered with the starting position off set by 1 or 2 nts.

When comparing both sets of small RNA from the Y-Sat genome, clusters often overlap. Also, the staggering is present in both samples, making it intriguing to speculate if these small RNA were amplified by a viral or the host RNA dependent RNA polymerase (RdRp or RDR). There are also several examples whereby a 21 nt small RNA from HC-Pro(–) plants is found in one position (e.g. 24-5, 137-117, 237-217) and a cascade of small RNA from HC-Pro expressing plants is found 5' to the original 21 nt small RNA on the (–) strand. This observation can not easily be explained by a four fold higher coverage of the Y-Sat genome in HC-Pro(+) plants. The small RNA cluster at position 24-4 shows a ratio of 1:9. It could, however, be explained with an amplification effect by which the small RNA anneal to the Y-Sat genome and serve as primers for RdRp or RDR extension. The now double stranded Y-Sat genome could be digested by an endonuclease into small RNA fragments. If this process is repeated frequently, it could lead to the pattern seen in Figure 4.5 A.

To ensure the cloning data accurately reflects the distribution of Y-Sat small RNAs found in the total small RNA population, six 3' biotin-labelled DNA oligos were designed to probe for sense and antisense of three Y-Sat genome regions (positions 76 to 99, 173 to 193 and 225 to 247). Judging from the cloning data, the sense and antisense probe for Y-Sat position 76 to 99

should have equal intensities in the autoradiogram of the native gel shift. This is indeed the case (Figure 4.5, panel B, lanes 1 and 2). For the yellowing region of Y-Sat, position 173 to 193, a ratio of 3 antisense to 13 sense was observed from the cloned data. The intensities in the gel shift have a ratio of 1 to 9 (Figure 4.5, panel B, lanes 3 and 4). The Y-Sat region position 225 to 247 have a ratio of 13 antisense and 7 sense is seen in the cloning data as well as in the native gel shift assay – ratio is 2.3 to 1 (Figure 4.5, panel B, lanes 5 and 6). The samples were probed with three non-viral 3' biotin labelled oligos (pr1: group 01 pr2: MIR166 pr3: group 08) as positive control (Figure 4.5, panel B, lanes 13 to 18). The gel shift assay and the cloning data agree very well, suggesting that no bias was introduced during the cloning procedure.

4.4 Concluding remarks

Ebbie was able to assist in annotating the small RNA sequences by comparing each newly entered sequence to a dynamic database of previously cloned small RNAs (Ebhardt, et al., 2005). This process allows for rapid identification of frequently cloned small RNAs.

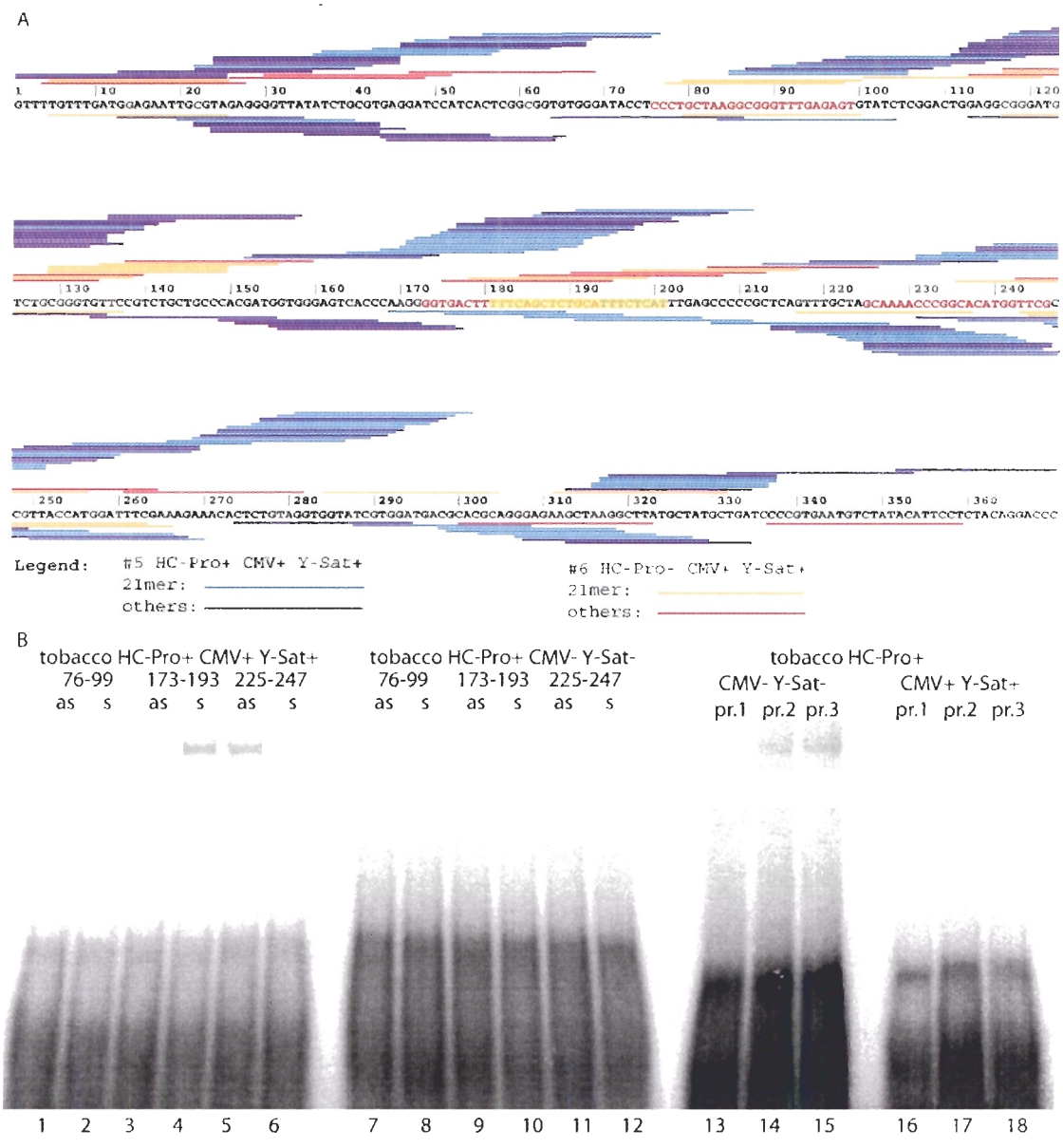
The streptavidin gel-shift assay provides the researcher with a valuable tool to compare the previously identified small RNAs and their relative expression levels when compared to total small RNA population. As discussed in Chapter 3, this might be important when comparing RNA sequencing data from small RNA cloning projects, which typically relies on T4 RNA ligase and an unmodified 3' terminus, with an unbiased assay that only requires hydrogen bonding between the two nucleic acid strands regardless of the sugar backbone modification.

Viral small RNAs from our cloning project support previous findings that the Y-Sat functions through small RNA production and not *via* expressing proteins or peptides. It is even more intriguing that no small RNA was found overlaying the 5' and 3' terminal positions, suggesting that the Y-Sat small RNAs are derived from a single stranded Y-Satellite. The single stranded Y-Sat folds into a characteristic secondary structure (see Figure 1.8) which either means that DCL4 and DCL2, the two dicer-like plant enzymes in the viral defence pathway, not only cut perfectly double stranded RNA but also hair pins. It could also mean that other dicer-like enzymes are involved in generating Y-Sat small RNAs from hairpin structures, such as DCL1, which naturally cleaves hairpin micro RNAs.

Figure 4.5: Y-Satellite small RNA aligned to the Y-Sat genome.

(A) In black letters is the Y-Sat genome. Blue bars represent small RNA cloned from HC-Pro(+) plants, whereas yellow/red bars are small RNA cloned from HC-Pro(–) plants. Coloured bars above the Y-Sat genome are from the sense strand; coloured bars below the Y-Sat genome align to the antisense strand.

(B) Autoradiogram of a streptavidin gel-shift assay probing with six Y-Sat small RNA 3' biotin labelled DNA oiogos from HC-Pro(+) CMV/Y-Sat(+) plants (lanes 1 to 6) and HC-Pro(–) CMV/Y-Sat(+) plants (lanes 7 to 12). As positive control, both plant samples were also probed with pr1 (group 01), pr2 (MIR166) and pr3 (group 08) lanes 13 to 18).



5 CONCLUSIONS AND IMPLICATIONS

For nearly half a decade of cloning small RNAs from plants, it was not obvious that plant small RNAs are modified on their 3' terminus. The work presented in this thesis shows that the majority of small RNAs in plants have a modification on their 3' terminal 2' hydroxyl, a major difference to 3' terminally unmodified animal small RNAs. This modification is influenced by P1/HC-Pro in a DCL pathway dependent manner.

5.1 How does HC-Pro prevent methylation of small RNAs?

In CMV/Y-Sat infected plants there were two general populations of small RNAs: endogenous or housekeeping small RNAs including micro RNAs and exogenous small RNAs originating from the viral infection. P1/HC-Pro interfered only with the 2'-O-methylation of the exogenous small RNAs, whereas the great majority of endogenous small RNAs were modified on their 2'-hydroxyl. In P1/HC-Pro(+) CMV/Y-Sat(+) plants the endogenous small RNAs were modified and the exogenous small RNAs were not. This discrepancy lead to the rapid ligation of the 3' adaptor molecule to the unmodified RNA by T4 RNA ligase whereas the 3' terminal 2'-O-methyl modified endogenous RNAs were cloned with a lesser efficiency. This unbalanced 3' adaptor ligation lead to the discrepancy between the 5' radiolabelling and the histogram of the cloning data.

5.1.1 Alternative models how HC-Pro acts as a RNA silencing suppressor

The model presented in the work supports the idea that P1/HC-Pro interferes with the biogenesis of small RNAs in plants at a crucial step when a 2'-O-Me is attached to the 3' terminal ribose by HEN1 (Ebhardt, et al. 2005; Yu, et al., 2005). It should be noted, that recent literature report alternative models how HC-Pro achieves its RNA silencing activity.

5.1.1.1 HC-Pro binds small RNAs

In the introduction chapter of this thesis, the viral RNA silencing suppressor p19 was chosen as a model RNA silencing suppressor that preferably binds to 19 bp RNA helixes with a 3' 2 nt overhang on each side, termed '19rh2' from here on. This mode of binding 19rh2 was

supported by native gel-shift assays and two independent crystal structures clearly showing the tight binding of p19 to 19rh2 (Vargason, et al., 2003; Ye, et al., 2003b). Two recent publications also claim that the viral RNA silencer HC-Pro also binds 19rh2 (Lakatos, et al., 2006; Merai, et al., 2006). To the best of my knowledge, no HC-Pro–19rh2 crystal structure has been solved. (A former post-doc from the Patel-lab is attempting to crystallize the HC-Pro–19rh2 complex [Patel DJ personal communications].) In the past, HC-Pro was shown to interact weakly with nucleic acid (Maia, Bernardi, 1996; Pirone, Thornbury, 1984).

Thus far, *in vitro* native gel shift assays are the only evidence for the claim that HC-Pro binds 19rh2. Merai and colleagues state that HC-Pro binds double-stranded small RNAs specifically and size-selectively (Merai, et al., 2006). Their assay constitutes of adding various suppressor proteins to radiolabelled small RNA, e.g. 19rh2. Three lines of evidence in their publication argue against the same mode of binding as RNA silencing suppressor p19. First, HC-Pro isolated from potato virus Y (PVY) or tobacco etch virus (TEV) infected plants do not show the same binding activity to 19 bpR. In Merai *et al.* Figure 2, panel B lanes 8 and 9, HC-Pro from PVY does not show binding to 19rh2 whereas HC-Pro from TEV shows quantitative binding to the same construct (with identical results shown in Merai *et al.* Figure 4, panel A lanes 8 and 9). Secondly, the mode of binding when comparing p19 with HC-Pro from TEV is not identical using the identical 19rh2. The autoradiogram of their native gel-shift assays show that p19–19rh2 shifts to about half the height of the HC-Pro-TEV–19rh2 (Merai *et al.* Figure 4, panel A lane 2 *versus* lane 9). Thirdly, in a competition experiment where radiolabelled 19rh2 was substituted with unlabelled 19rh2, two species of RNA–HC-Pro complexes are visible versus only one species of RNA–p19 complex. In summary: HC-Pro isolated from two different sources lead to two distinct results. The p19–19hr2 and the HC-Pro–19hr2 complexes do not show the same mobility in any of the native gel-shift assays, thus it can be inferred from this data that p19 and HC-Pro have a differential mode of action as RNA silencing suppressors *in vitro*.

Lakatos and colleagues studied HC-Pro and p19 more in-depth (Lakatos, et al., 2006). In their first experiment Lakatos and colleagues test p19, p21 and HC-Pro's ability to suppress cleavage of a tested mRNA, which HC-Pro does with a 10 times lower efficiency than p19. This assay does not test for the ability of HC-Pro to block methylation of siRNAs, as this *in vitro* assay is done using insect embryo extracts (*Drosophila*). In a very elegant second experiment, Lakatos and colleagues incubate RISC with the RNA silencing suppressors and monitor the radiolabelled siRNA as a function of RNA silencer concentration. Even at very sub nano-Molar protein concentrations p19 is able to extract siRNAs from the assembled RISC and form its own siRNA-

p19 complex. HC-Pro does not extract siRNAs from the RISC, instead at about 150 nM forms a complex comprised of siRNA–DCR2–R2D2 and HC-Pro. This experiment actually supports our findings that HC-Pro inhibits a crucial step in the formation of the 3' terminal 2'-O-Me, possibly by binding to the pre-siRNA complex inhibiting HEN1 from methylating the siRNA duplex. In a third experiment, Lakatos and colleagues show gel-shift assays similar to Merai *et al.* with similar results: HC-Pro binds weakly 19hr2 at a concentration of 500 nM. Lakatos and colleagues showed in an earlier publication that p19 binds to 19hr2 with an affinity of 5 nM (Lakatos, et al., 2004).

In summary, there is no compelling evidence that HC-Pro and p19 have the same mode of binding 19hr2 *in vitro*. On the contrary, instead of extracting siRNA from a loaded RISC, like p19, HC-Pro binds to a pre-RISC complex supporting our finding that HC-Pro interferes with the siRNA biogenesis preventing HEN1 to methylate the 3' terminus of siRNAs.

5.1.1.2 P1 of P1/HC-Pro enhances HC-Pro

Valli and colleagues studied the potyvirus family and found that P1 enhances the RNA silencing effects of HC-Pro (Valli, et al., 2006). The potyvirus is a monopartite genome and the translated N-terminus encodes P1/HC-Pro. Using a green fluorescent protein reporter (35S:GFP), the authors were able dissect the influence of various constructs by observing the fluorescence of GFP. Using an empty control vector, the 35S:GFP shows its highest levels of fluorescence 2 to 3 days post induction (dpi), before the exogenous GFP message is silenced by the plant's RNA silencing machinery. Using a 35S:GFP:P1 construct, the authors saw no difference between the additional P1 expression and the empty vector. When 35S:GFP:HC-Pro was expressed in cells, the plants showed a bright fluorescence 6 dpi. The fluorescence signal was even stronger when the 35S:GFP:P1/HC-Pro construct was used, indicating a collaborative effort of both P1 and HC-Pro to suppress the plant's RNA silencing machinery. Unfortunately, these experiments were not able to establish a model by which these two proteases act in concert to suppress the host's RNA silencing machinery.

5.1.1.3 HC-Pro affects micro RNA biogenesis

Kasschau and colleagues expressed the P1/HC-Pro RNA silencing suppressor of the *Turnip mosaic virus* under the strong 35S promoter in *Arabidopsis* (Kasschau, et al., 2003). Out of six transformed plants with P1/HC-Pro, three showed very high levels of expression and another one showed lower levels of P1/HC-Pro expression. The plants with the higher levels of

P1/HC-Pro showed the greatest developmental defects. The plant with a lower level of P1/HC-Pro expression showed less severe levels of expression. Kasschau traced the developmental defects back to the interference of P1/HC-Pro with endogenous miRNAs causing the levels of miRNA targeted mRNAs to rise (Kasschau, et al., 2003).

5.2 A uniting model for HC-Pro's affects on small RNA biogenesis?

When reading the literature on HC-Pro, it is not clear if HC-Pro affects exogenous small RNAs to the same degree as endogenous RNAs. On first glance, the results in Ebhardt *et al.*, 2005 claiming that HC-Pro predominantly affects methylation of exogenous small RNAs and Kasschau, *et al.*, 2003 claiming developmental defects due to HC-Pro's interference with endogenous miRNAs contradict each other. Examining the results of the latter publication closer, it becomes obvious that the observed developmental defects are directly linked to the amount of HC-Pro expressed in the plant: only plants expressing high levels of P1/HC-Pro also show high levels of developmental defects due to HC-Pro's interference with miRNA biogenesis. Lower levels of P1/HC-Pro had a lesser effect on development.

Therefore, a uniting model for HC-Pro's affects on small RNA biogenesis must include the concentration of the RNA silencing suppressor in the host cell. To generate transgenic plants expressing a protein of interest, plant promoters are used to enhance the yield of the protein of interest in certain plant tissues (Schunmann, et al., 2003a; Schunmann, et al., 2003b). The expression levels of the same protein can easily show three orders of magnitude difference using various promoters (Horstmann, et al., 2004).

Another factor influencing HC-Pro's affect on small RNA biogenesis is the sub-cellular localization. From GFP-tagged DCL microscopy observations we know that DCL1 resides in the nucleus (Papp, et al., 2003). As methylation occurs rapidly after cutting the precursor RNA with DCL, HC-Pro must co-localize with the respective DCL.

A uniting model? HC-Pro interferes with small RNA biogenesis at the step of 2'-O-methylation. During viral infection, low levels of HC-Pro in the cytoplasm will predominantly interfere with viral small RNAs (Ebhardt, et al., 2005). A strong promoter such as the 35S will lead to higher levels of HC-Pro throughout the cell, including the nucleus, and affect not only the 2'-O-methylation of exogenous small RNAs, but also endogenous small RNAs (Kasschau, et al., 2003). The differential targeting of small RNAs is achieved by localization of HC-Pro, which should predominantly accumulate in the cytoplasm. At higher concentrations of HC-Pro, the RNA silencing suppressor also enters the nucleus where it interferes with the biogenesis of small

RNAs. To prove this hypothesis, I would design a P1/HC-Pro construct with a fusion GFP using different promoters. Using fluorescence microscopy, the sub-cellular localization of HC-Pro could be determined. It would also be interesting to determine if the sub-cellular localization of HC-Pro changes upon viral infection.

5.3 Implications of HC-Pro's effects on small RNA biogenesis

It was previously shown by genetic experiments that HEN1 is a methyltransferase responsible for methylating the 3' terminus of miRNAs (Yu, et al., 2005). The results presented in this work clearly show that HC-Pro interferes with the methylation of viral small RNAs and not endogenous micro RNAs. Due to this discrimination of HC-Pro towards viral small RNAs an imbalance during the cloning of small RNAs using T4 RNA ligase became obvious. The viral unmodified siRNAs were appended more readily with the 3' adenylylated DNA oligo than 3' terminal modified endogenous small RNAs. As no other small RNA cloning in our four-point comparison study showed any discrepancy between the 5' radiolabelling and the actual cloning data, it can be inferred that all small RNAs are modified on their 3' terminus. These initial observations were supported by periodate cleavage assays, showing a sensitivity of viral siRNAs in HC-Pro(+) plants, but not in HC-Pro(–) plants. Also, endogenous miRNAs were also not sensitive to periodate cleavage, supporting the hypothesis of a 3' terminal modification.

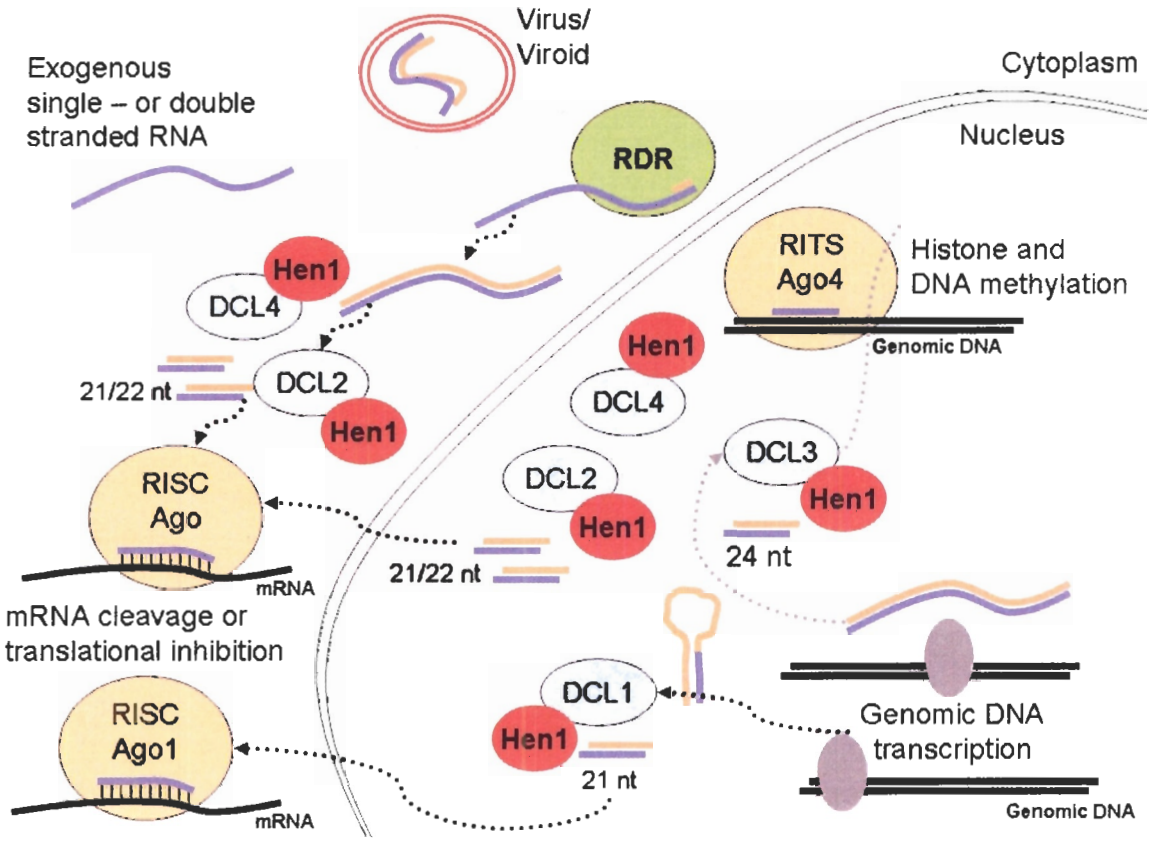
This ubiquitous modification of plant small RNAs is not found in animals, a significant difference between the two kingdoms. Also, the modification is on the 2' hydroxyl (a 2'-O-Me) of the 3' terminus (Ebhardt, et al., 2005) preventing 3'-end uridylation activity (Li, et al., 2005). As the 3' terminal modification is appended rapidly after cleavage by Dicer-like RNaseIII ribonucleases, it is very likely that the methyltransferase HEN1 is associated with the Dicer-like enzymes. Figure 5.1 shows an updated model for small RNA biogenesis in plants. In this model not all known RNA silencing pathways are shown, which were discussed in the Introduction chapter, but the model emphasizes the association of HEN1 to Dicer-like enzymes (DCLs).

There are multiple scenarios how HC-Pro could interfere with appending the methyl group. It is unlikely that HC-Pro directly binds to Hen1, as Yu and colleagues did not observe binding of Hen1 to HC-Pro (Yu, et al., 2006). HC-Pro could bind to cytoplasmic DCL4 or DCL2 and prevent HEN1 from interacting with the complex. HC-Pro could weakly bind siRNA duplexes and prevent HEN1 from methylating RNA. Or it could prevent a cofactor binding to HEN1 thus inhibiting HEN1 activity. A mode by which HC-Pro binds to a pre-RISC complex, thus preventing or inhibiting HEN1 from methylating RISC associated small RNAs, is very likely

and supported by *in vitro* data (Lakatos, et al., 2006).

Figure 5.1: Updated model for small RNA biogenesis in plants.

This schema shows some of the RNA silencing pathways in a plant cell, emphasizing the ubiquitous methylation of small RNAs associated with RNA silencing pathways. As the methylation occurs rapidly after cleavage of small RNA precursors by Dicer-like enzymes, HEN1 is most likely associated with the Dicer-like enzymes in plants.



Each Dicer-like ribonucleases (DCL) in plants is responsible for processing specific small RNAs. Our data suggests that HEN1 is associated with each DCL and that methylation occurs immediately after the cleavage event. It also shows that each RNA silencing pathway shows a spacial separation within the cell. For example, DCL1 – responsible for miRNA cleavage – is localized to the nucleus; HC-Pro does not affect the methylation state of miRNAs. Thus, HC-Pro most likely does not reside in the nucleus but more likely to be associated with DCL4 and DCL2 in the cytoplasm where it interferes with HEN1 activity.

Other researchers provide evidence for HC-Pro to bind small RNAs – if HC-Pro was to bind all small RNAs equally, growth defects would be observed, as miRNAs are essential for

development of organs. In our hands, no growth defects were observed with the expression of HC-Pro in tobacco plants. Therefore, if HC-Pro only negatively affects siRNAs and not endogenous miRNAs, these two groups of small RNAs are in two different complexes, as would be expected. The complex which harbours miRNAs is harder to access by HC-Pro than the siRNA complex. If we recall that both are incorporated into an argonaute containing protein complex (called RISC), it opens the question why one complex is more stable or more resistant to HC-Pro binding. At this point a speculative answer is the association of miRNA-RISC with a larger protein complex, making it harder for HC-Pro to access the miRNAs, whereas the viral siRNA-RISC is without the added protection of a larger protein complex.

5.4 Paradox solved?

Wang and colleagues (Wang, et al., 2004) established that the phenotype of CMV/Y-Sat could be rescued by expressing a strong RNA silencing suppressor HC-Pro. Usually, a CMV/Y-Sat infection in tobacco leads to chlorosis and the tobacco leafs yellow. A transgenic tobacco plant expressing HC-Pro is protected from chlorosis after CMV/Y-Sat infection; the plant leafs were green. This phenotypic result would not have been expected from a Northern hybridization experiment when probing RNA from both infected plants with a Y-Sat probe. The HC-Pro(–) plants showed very little amounts of Y-Sat small RNAs, whereas in HC-Pro(+) plants high levels of Y-Sat small RNAs were detected. There was an obvious paradox from these results: how could small amounts of Y-Sat small RNAs lead to leaf yellowing in HC-Pro(–) plants but large amounts of Y-Sat small RNAs in HC-Pro(+) plants showed no apparent yellowing?

From the phenotype observation we know that more small RNAs from Y-Sat do not translate into more biological active siRNAs from Y-Sat – the HC-Pro(+) infected tobacco plants are green. Thus, the quantity does not matter. But how is quality achieved by biological active small RNAs from Y-Sat in HC-Pro(–) CMV/Y-Sat(+) plants? A possible explanation is the importance of Y-Sat's secondary structure. The secondary structure of the single stranded Y-Sat could guide DCL4 and DCL2 to make the correct cut generating a biological active siRNA. It might also stop DCL4 and DCL2 from cutting more than once, as demonstrated by Gasciolli and colleagues (Gasciolli, et al., 2005a). The result of a successful cleavage is the biological active siRNA from Y-Sat, which causes yellowing in the plant by either directly targeting an mRNA, e.g. chlorophyll synthetase, or by neutralizing endogenous small RNAs by forming a complementary duplex.

Secondary small RNAs from Y-Sat in HC-Pro(+) CMV/Y-Sat(+) plants could arise from

a primer dependent and a primer independent amplification of the Y-Sat genome. In the primer dependent amplification, the Y-Sat siRNAs lacking the terminal methyl group could serve as primers for the primer dependent amplification of the Y-Sat genome by RdRp or RDRs. In the case of primer independent amplification, a Y-Sat siRNA actually cuts the Y-Sat single stranded RNA genome and the truncated molecule is amplified by RdRp or RDRs. Once the Y-Sat genome is double stranded, DCL4 and DCL2 will recognize this double stranded RNA as substrate and initiate cleavage into 21/22 nt long RNA duplexes. The resulting small RNAs are not cut in registry with the biological active primary Y-Sat siRNAs with their seed sequence leading them either to the wrong mRNA, or to no mRNA at all. Thus, all si-RISC loaded with the secondary small RNAs from Y-Sat will not be active anymore. As there are substantially more secondary small RNAs from Y-Sat than primary Y-Sat siRNAs in HC-Pro(+) CMV/Y-Sat(+) plants, it is more likely that any given si-RISC loads the incorrect small RNA, thus is not biological active. Therefore, the plant leafs are still green, even with a large amount of Y-Sat and Y-Sat small RNAs detected by Northern hybridization and cloning.

Support for the primer independent amplification comes from the fact that some siRNAs seem to cleave Y-Sat RNA: e.g. Figure 4.5. There are two siRNAs position 1 to 24; position 12 would be the cleavage point and at position 13 starts a new Y-Sat small RNA. Another example is a small RNA from position 152 to 173 where at position 164 a new Y-Sat small RNA starts, one nucleotide position past the expected cleavage site of the 152-173 siRNA. All cascades of small RNA from Y-Sat could be attributed to primer dependent amplification of the Y-Sat genome and subsequent cleavage of double stranded RNA by DCL4 or DCL2.

5.5 Concluding remarks

The discovery of a ubiquitous modification on the 2'-hydroxyl of the 3' terminal ribose of RNA silencing associated small RNAs in plants was a major discovery and is significant for the biological function of these small RNAs in plants. This is a major difference to the kingdom of animals where small RNAs associated with RNA silencing are not modified on their 2'-hydroxyl of the 3' terminal ribose.

APPENDIX

Buffers

Buffer	Ingredients
1xPNK	50 mM Tris·HCl (pH 7.9), 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT **Note:** A 10xPNK buffer stock without DTT was prepared, 1 mL was aliquoted into blue 1.5 mL eppendorf tubes, stored at -20 °C and 10 µL of 1 M DTT added before the tube was used.
1xPCR	100 mM Tris, 500 mM KCl, 15 mM $MgCl_2$, 0.1 % gelatin, pH 8.3
1xTBN	90 mM Tris-base, 45 mM boric acid, 75 mM NaCl, pH 8.5 **Note:** This buffer was stored as a 5x stock at room temperature for 1 – 2 months only.
1x T4 RNA ligase buffer	50 mM Hepes, pH 8.3, 10 mM $MgCl_2$, 3.3 mM DTT, 10 µg/ml BSA, 8.3% glycerol **Note:** T4 RNA ligase was purchased from Ambion (Ambion, Austin, TX), but the ligation reaction was not performed in the supplied buffer.
Streptavidin Binding Buffer	140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 10 % glycerol, 2 mM DTT,

dNTPs and NTPs

5X NTPs:

From powder, resuspend 1g in 2 ml of H_2O. Emperically adjust pH with 1M NaOH to pH 7. (Estimate equivalents of NaOH needed based on salt form of the NTP and charge of 3.5 at pH 7) Measure concentration of these stocks. Use the following molar absorption coefficient (260 nm): ATP = 15,000 M^{-1} cm^{-1}, CTP = 7,600 M^{-1} cm^{-1}, GTP = 11,400 M^{-1} cm^{-1}, UTP = 10,000 M^{-1} cm^{-1}. A 5x solution of NTPs contains 40 mM GTP, 25 mM CTP, 25 mM ATP and 10 mM UTP.

10x dNTPs for PCR:

For very large scale PCR, make from powder and don't deplete lab 10x. For smaller scale, can purchase 100 mM stocks from Pharmacia. Make up 10X stock 2mM of each dNTP, make sure final solution is neutral.

Phosphorylation of RNA and DNA

5' labelling of RNA and DNA is essential for visualizing RNA or DNA on a autoradiogram. As discussed in the introductory chapter and depicted in Figure 1.12, T4 Polynucleotide kinase performs two reactions: forward and exchange reaction. A typical forward reaction was composed of 1xPNK buffer, 5 pmoles 5'OH DNA or RNA, 6.6 pmoles [γ-^{32}P]ATP (PerkinElmer, Boston, MA) and 5 U of T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) in a 10 µL reaction volume. The reaction was incubated at 37 °C for no more than 7 minutes and the reaction stopped by either heat inactivation at 65 °C or adding of an equal volume of 2xformamide loading dye.

The exchange reaction was used for radiolabelling phosphorylated RNA extracted from plants, especially for visualizing small RNAs in a total RNA preparation. A typical exchange reaction from total RNA was composed of 1xPNK buffer, 1 µL total RNA (see 'Small RNA isolation from tobacco' below), 3.3 pmoles [γ-^{32}P]ATP (PerkinElmer, Boston, MA) and 5 U of T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) in a 10 µL reaction volume. The reaction was incubated at 37 °C for 25 minutes and the reaction stopped by either heat inactivation at 65 °C for 25 minutes or adding of an equal volume of 2xformamide loading dye. The reaction was loaded onto a 23.75 % 20 mL denaturing PAGE (Sambrook, Russell, 2001) and run for 90 minutes at 20 W using a BIORAD PowerPac 3000 (BIO-RAD Laboratories, Hercules, CA).

Figure I-1: Forward reaction speed.

Shown here is the speed of a forward reaction. (A) The phosphorylation reaction contained 3.5 pmoles 28.8 DNA primer, 3.3 pmoles [γ-^{32}P]ATP, 10 U polynucleotide kinase and 1xPNK buffer in a 20 µL reaction volume. The autoradiogram of the 23 % denaturing PAGE includes the incubation times at 37 °C above each lane. (B) The time dependency of DNA oligo 28.8, [γ-^{32}P]ATP and inorganic phosphorous (iP) are shown in a diagram.

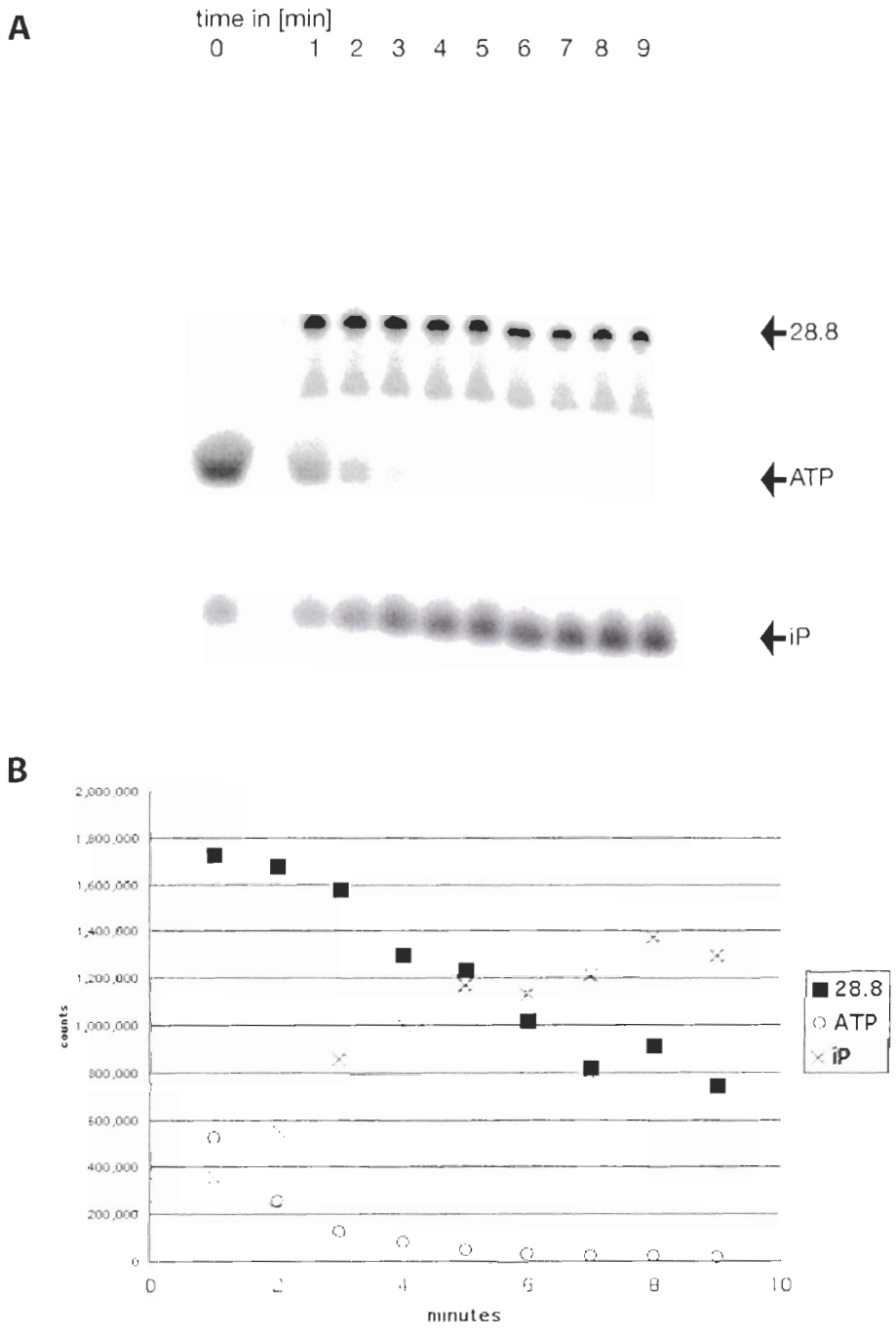
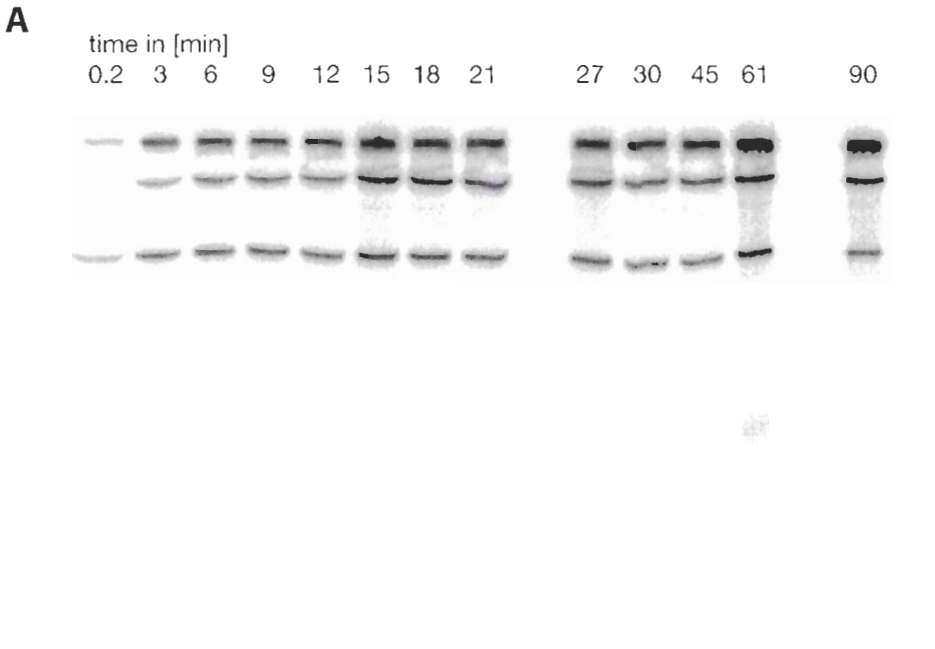


Figure I-2: Exchange reaction speed.

Shown here is the speed of an exchange reaction. (A) The phosphorylation reaction contained 2 µL total RNA, 6.6 pmoles [γ-^{32}P]ATP, 10 U polynucleotide kinase and 1xPNK buffer in a 20 µL reaction volume. The autoradiogram of the 23 % denaturing PAGE includes the incubation times at 37 °C above each lane. (B) Time dependency of 24 nt RNA shown in diagram.



Oligos

Oligo name	Sequence in 5' - 3' direction
B16.5	AppGAAGAGCCTACGACGA (DNA adaptor) App: 5'-Adenlylyated
r17.50	rAUCGUAGGCACCUGAAA (RNA adaptor)
21.22	ACTTATCGTAGGCACCTGAAA
32.21	TGATGTTGCGGGTGCCTCGTCGTAGGCTCTTC
21.28	CTTCAGCCTATCTCCGACTTB (complement to group 08) B: 3'-BiotinTEG-CPG (Glen Research, Sterling, VA)
24.27	GCCGATTCAGGTGCGTATCAATTB (complement to group 03)
25.9	TGTGCTCACTCTCTTCTGTCATTTB (complement to MIR156)
26.5	AGGGATGCAACACGAGGACTTCCTTB (complement to group 19)
26.6	CCTACCATTTGAGCTACATCCCCTTB (complement to group 20)
27.11	ACGAATCGGAGCGACAAAGGGCTGTTB (complement to group 16)
27.12	GTTAGTGCTGGTATGATCGCATCCTTB (complement to group 11)
28.37	GACTACTGGCAGGATCAACCAGGTATTB (complement to group 01)
28.39	ATACTTACCACTATACTACAACGACTTB (complement to group 12)

\Tobacco plants

Plant Transformation: N. tobacum Wisconsin 38 was transformed by using an antibiotic as the selective agent. Analysis of Transgenic Plants: GUS activity in transgenic tobacco was measured at 37 °C by the fluorometric 4-methylumbelliferryl-β-glucuronide assay by using 5 μg of leaf protein extract.

Virus/Satellite RNA Infection: The Y-Sat culture was initiated by mechanically inoculating CMV-infected young tobacco with transcript synthesized from an infectious Y-Sat clone. Extracts of the CMV/Y-Sat-infected tobacco leaves were then used for subsequent inoculation (Wang, et al., 2004).

Small RNA isolation from tobacco

Approximately 0.5 g of leaf tissue was collected from healthy tobacco plants or plants infected for 5 weeks with CMV or CMV plus Y-Sat, and ground into fine powder with pestle and mortar in the presence of liquid nitrogen and sand. The powder was transferred into a pre-cooled tube and suspended quickly with a mixture of 0.5 ml extraction buffer (0.1 M LiCl, 1 % SDS, 0.01 M EDTA, 0.1 M Tris, pH 9) with 0.5 ml phenol that was pre-heated in boiling water. After the tube cooled to room temperature, 0.5 ml chloroform was added and the mixture was extracted

for up to 20 minutes. The tube was centrifuged and the supernatant transferred to a fresh tube and extracted again with 0.5 ml chloroform. Nucleic acids were precipitated from the final supernatant using 1/10 volume of 3M NaOAc and 3 volumes of 100 % ethanol, air-dried, and dissolved in 0.73 ml H_2O. To remove large RNA molecules, the nucleic acid solution was mixed with 0.19 ml of 4M NaCl and 0.58 ml of 13 % PEG-8000, stored in ice for 30 minutes, and centrifuged at 4 °C for 20 minutes. The supernatant was transferred to a fresh tube and the small RNA species was precipitated with 1/10 volume of 3 M NaOAc and 3 volumes of ethanol. The pellet from PEG precipitation step was used for large RNA and DNA analysis.

Steps for RNA extraction

- Grind ~0.5 g plant tissue into powder in liquid N_2 with assistance of sand. Transfer the powder into 2 ml centrifuge tuber and keep at –80 °C until use.
- Heat a mixture of 0.5 ml extraction buffer (100 mM LiCl, 1 % SDS, 100 mM Tris pH 9, 10 mM EDTA) and 0.5 ml of phenol. Add the mixture into the powder, suspend as quickly as possible to avoid RNA degradation.
- Add 0.5 ml of chloroform and extract for up to 20 min.
- Spin. Transfer the supernatant with a wide-bore tip (blue tips with the ends cut) into a fresh 2 ml tube, extract with 0.5 ml of chloroform as above.
- Add 1/10 volume of 3M NaOAc, and 3 volumes of ethanol to the final supernatant, mix, and keep at -20 °C for 1 hour or longer.
- Spin at 4 °C for 20 min. Air-dried. Dissolve in H_2O.
- Removal of high-molecular-weight RNA
- Mix total RNA with polyethylenglycol 8000 (PEG-8000) and NaCl to make final concentrations of 5 % PEG and 0.5 M NaCl. Keep on ice for >10 min.
- Spin for 10 min and transfer the supernatant to fresh tubes.
- Add NaOAc to 0.3 M and 3 volumes of ethanol. Keep at -20 °C for 1 hour or -80 °C for 15 min.
- Spin at 4 °C for 20 min.
- Wash the pellet with cold 80 % ethanol and dry.

Notes:

- The protocol can be scaled up.
- Do not try to use too much tissue. Always try to use young tissue when possible.
- Try to avoid any cell debris.
- Never over-dry the pellet. In fact, undried pellet is much easier to dissolve than over-dried pellet.
- Dissolve the RNA in relatively large volume (e.g. ~0.5 ml for RNA from 1~2 g tissue).
- The final volume is about 1 ml for RNA from 1~2 g tissue.
- Again, try to avoid over-drying.

Cloning of smRNA

- 3' adaptor ligation
- 1.0 µL 5xRNA ligase buffer
- 1.0 µL RNA sample (isolated as described above)
- 1.0 µL AppB16.5 (100 µM stock – 20 µM final)
- 1.5 µL ddH_2O
- 0.5 µL T4 RNA ligase (GE Healthcare, Amersham 40 U/µL)
- Final volume: 5 µL.

Incubate at room temperature for 1 hour. Run on a 15 % denaturing sequencing polyacrylamide gel at 70 W for 120 minutes. Expose to a phosphorimager plate, then cut band corresponding to 34-mer (ligation product of 16mer + 28 = upper limit). Elute overnight in 500 µL 300 mM NaCl. Add 4 µL of 5 mg/mL glycogen (Ambion, TX) for a total concentration of 13.3 µL/mL and add 2x volume of ethanol and precipitate for 1 hour at -20 °C. Spin down at 13,200 rpm in table centrifuge for 15 minutes at 4 °C. Resuspend pellet in 10 µL ddH_2O.

- 5' adaptor ligation
- 2.0 µL 5xRNA ligase buffer
- 5.0 µL RNA sample (gel pure from 3' adaptor ligation)
- 4.3 µL r17.50 (93.05 µM stock – 31.5 µM final)

- 0.4 μL 10 mM ATP
- 1.0 μL T4 RNA ligase (GE Healthcare, Amersham 40 U/μL)
- Final volume: 12.7 μL.
- Incubate for 1 hour at room temperature. Repeat steps at for 3' adaptor ligation and resuspend in 10 μL ddH_2O.

Reverse Transcription of ligation products

- 5.0 μL RNA sample (gel pure from 5' adaptor ligation)
- 11.6 μL 32.21 (25.8 μM)
- Heat to 80 °C for 2 min, then spin to cool.
- 5.0 μL First Strand Buffer (GibcoBRL)
- 7.0 μL 10x dNTPs
- 0.3 μL 1 M DTT
- 3.0 μL SuperScript II Reverse Transcriptase (GibcoBRL)
- Final volume: 31.9 μL.

Heat to 48 °C for 3 minutes before adding reverse transcriptase. Take 2 μL before adding enzyme for negative control. Incubate both positive and negative controls at 48 °C for 1 hour. Add 1 μL of 0.1 M EDTA, 3.8 μL 1 M KOH. Incubate at 90 °C for 10 min. Add 8 μL Tris-HCl until pH ~8, then add 1 μL 0.2 M $MgCl_2$.

PCR amplification

- 10 μL 10x PCR buffer
- 10 μL 10x dNTPs
- 6.3 μL oligo 21.22 (15.9 μM)
- 1.9 μL oligo 31.21 (25.8 μM)
- 10 μL DNA from previous step (Note: for control PCR, add t_0 from RT step)
- 61.8 μL ddH_2O
- 2.0 μL Taq polymerase (10 U/μL)
- Final volume: 100 μL.

Run for 22 cycles, taking aliquots at 20 and 22 cycles. Add EDTA (to final concentration of 2 mM) and NaCl (to final of 300 mM), phenol/chlorophorm extract PCR and precipitate with 2x ethanol. Resuspend in 10 µL ddH_2O for a 10xPCR stock.

Fill in reaction for cloning

Before TA-cloning using pCR2.1 vector from Invitrogen, extend PCR products to contain an overhang T in a 30 µL PCR reaction:

- 3.0 µL 10xPCR buffer
- 3.0 µL 10xdNTPs
- 5.0 µL 10xPCR stock from step above
- 0.6 µL Taq polymerase
- 8.4 µL ddH_2O
- Final volume: 30 µL.

Use the fresh PCR product for ligation into pCR2.1 vector, following precisely the instructions from Invitrogen (Carlsbad, California 92008, U.S.A.).

Analysis of small RNA sequence

Applied Biosystems sequencing chromatograms were manually reviewed by using CHROMAS (www.technelysium.com.au/chromas.html). All files having unambiguous sequence within the cloning region were analyzed by *Ebbie*, a dynamic web server that searches for single or multiple instances of primer pairs within the DNA sequencing file. *Ebbie*'s strength lies in the reliable storage of sequencing data and instant sequence analysis of found smRNA clones to previously found smRNA clones, allowing for uncovering of frequent clones. Ebbie is described in more detail in Chapter 2 of this thesis.

Gel shift isolation of small RNA

RNA 15 to 30 nts in length were gel purified on a 12 % denaturing PAGE using 18 and 24 nt 5' radiolabelled markers, eluted in 300 mM sodium chloride overnight and precipitated with 12.5 µg/ml glycogen (Ambion, Austin, TX) in 2.5x Ethanol. The gel purified RNA was resuspended in dH_2O to a final estimated concentration of 1 µM. 5 pmoles of gel purified small RNA was radiolabelled using an excess of [γ-^{32}P]ATP (PerkinElmer, Boston, MA) and 5 U of T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) in 1x PNK buffer. The reaction was

incubated at 37 °C for 25 minutes, followed by 65 °C for 25 minutes to heat inactivate T4 polynucleotide kinase. To the 5' radiolabelled 5 pmoles of RNA, 10 pmoles of anti-sense probe (25.9, 26.5, 27.12 or 28.37) was added and the mixture incubated at 1xTBN at 90 °C for 90 seconds. After slowly cooling down to room temperature, 4 µg Streptavidin (SIGMA S0677) were added, loaded onto a 1x TBN 10 % native PAGE and run for 90 minutes at 10 W at 4 °C.

Streptavidin gel shift reaction step-by-step:

- 1 µL 5xTBN
- 1 µL radiolabelled RNA
- 1 µL complementary biotin labelled oligo
- incubate for 90 seconds at 90 °C in PCR machine
- slowly cool down reaction
- 1 µL streptavidin; SIGMA S0677, 1 g resuspended in 250 µL Binding Buffer
- incubate for 10 min at room temperature
- 1 µL 6xnative loading dye (Fermentas)

This method has the following advantages over the conventional Northern blot hybridization approaches: (i) The relative amount of a particular smRNA can be directly measured by comparing the signal intensity of the RNA:DNA duplex with that of the total smRNA population. (ii) The potential difficulty in detecting short RNA species by Northern hybridization is overcome by performing the hybridization in a solution containing a large excess of biotin-labelled DNA probe. For instance, as little as 1-3 fmol of smRNA can be detected by a conventional Northern blot, but the streptavidin gel shift assay outlined here is 10-100 times more sensitive. (iii) The smRNA in the shifted RNA:DNA duplex can be recovered for further biochemical analysis.

Periodate treatment

Radiolabelled RNA was incubated in 25 mM HEPES (pH 7.0) and 50 mM sodium periodate at room temperature for 10 minutes in the dark. The β-elimination of the unstable dialdehyde was performed in formamide/30 mM EDTA (ratio 1 vol. aqueous to 1 vol. formamide) at 99 °C for 30 minutes.

Reagents:

- 2x formamide loading dye (5 mM EDTA, Bromophenol Blue [and/or other colors], the rest is formamide)
- 100 mM HEPES buffer, adjusted to pH 7.0
- 200 mM Sodium Periodate ($NaIO_4$)
- 200 mM Sodium Chloride (NaCl)
- 5' radiolabelled RNA
- Protocol:
- 5' end label RNA 10 µL reaction
- take 1 µL of phosphorylated RNA and add 9 µL 100 mM HEPES, pH 7.0
- Note: the amount of RNA can be adjusted, depending how much RNA was labelled and how "fresh" your isotope is.
- take out 5 µL for t_0 time point, adding to t_0 5 µL 200 mM NaCl and 10 µL 2x formamide loading dye
- Note: adding NaCl to the t_0 time point is nice so that the Na^+ concentrations are the same in both samples.
- to remaining RNA, add 5 µL of 200 mM $NaIO_4$
- incubate at 22 °C in the dark for 10 min (e.g. PCR machine)
- add 10 µL formamide loading dye
- heat up sample to 99 °C for 30 minutes (e.g. PCR machine)
- run into high percentage denaturing PAGE, e.g. 23 % for 7mer/24mer,

Northern hybridization

A very useful resource was the Bartel lab's Northern hybridization protocol (Bartel, D.P., MIT http://web.wi.mit.edu/bartel/pub/protocols/miRNA_Nrthrns_Protocol.pdf).

A 15 % denaturing PAGE was run and then transferred onto a Nylon membrane as described by Ming and colleagues (Ming, et al., 1994). The membrane was then UV-cross-linked twice and baked at 80 °C for 45 minutes.

To prevent non-specific binding of probes the membrane was incubated in pre-

hybridization buffer plus 10 µL of 1 g/mL salmon sperm DNA for at least 1 hour. A DNA probe was radiolabelled using the forward phosphorylation methodology and added to the pre-hybridization solution.

(Pre-)Hybridization solution

Final	Start	100 ml (two blots)	800 ml (16 blots)
2.5X SSC	20X	12.5 ml	100 ml
10 mM Na_2HPO_4 pH 7.2	1M	1 ml	8 ml
3.5 % SDS	10%	35 ml	280 ml
1X Denhardt's Solution (Sambrook, Russell, 2001)	100X	1.5 ml	12 ml

The hybridization probe was incubated overnight with the Nylon membrane and washed the next morning. Typically, there were at least two washes using a non-stringent wash. If needed, additional washes were applied, including a more stringent wash.

Non-Stringent Wash Solution

Final	Start	200 ml (>two blots)	1 L (>six blots)
1.5X SSC	20X	30 ml	75 ml
12.5 mM NaH_2PO_4 pH 7.5	1M	5 ml	12.5 ml
2.5 % SDS	10%	100 ml	250 ml
5X Denhardt's Solution	100X	20 ml	50 ml

Stringent Wash Solution

Final	Start	200 ml (>four blots)	500 ml (>12 blots)
0.5X SSC	20X	10 ml	25 ml
0.5% SDS	10%	20 ml	50 ml

After the washes, the membranes were packaged into SARAN-wrap and exposed to Molecular Dynamics phosphorimager plates. Depending on the strength of the radiolabel, an

initial scan of the phosphorimager plate on the Molecular Dynamics Strom 820 (GE Healthcare Life Sciences) could be performed an hour after exposure. For better quality scan, exposures overnight were appropriate.

BIBLIOGRAPHY

Allison LA (2007) Fundamental Molecular Biology, Ed First. Blackwell Publishing Ltd., Malden, MA, U.S.A.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389-3402

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A **95:** 13079-13084

Aranda MA, Fraile A, Dopazo J, Malpica JM, Garcia-Arenal F (1997) Contribution of mutation and RNA recombination to the evolution of a plant pathogenic RNA. J Mol Evol **44:** 81-88

Astier-Manifacier S, Cornuet P (1971) RNA-dependent RNA polymerase in Chinese cabbage. Biochim Biophys Acta **232:** 484-493

Atreya CD, Pirone TP (1993) Mutational analysis of the helper component-proteinase gene of a potyvirus: effects of amino acid substitutions, deletions, and gene replacement on virulence and aphid transmissibility. Proc Natl Acad Sci U S A **90:** 11919-11923

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell **116:** 281-297

Beclin C, Boutet S, Waterhouse P, Vaucheret H (2002) A branched pathway for transgene-induced RNA silencing in plants. Curr Biol **12:** 684-688

Berger PH, Hunt AG, Domier LL, Hellmann GM, Stram Y, Thornbury DW, Pirone TP (1989) Expression in transgenic plants of a viral gene product that mediates insect transmission of potyviruses. Proc Natl Acad Sci U S A **86:** 8402-8406

Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F,Jr, Hohn T, Pooggin MM (2006a) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233-6246

Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F,Jr, Hohn T, Pooggin MM (2006b) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233-6246

Bohnsack MT, Czaplinski K, Gorlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA **10:** 185-191

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell **123:** 1279-1291

Boutet S, Vazquez F, Liu J, Beclin C, Fagard M, Gratias A, Morel JB, Crete P, Chen X, Vaucheret H (2003) Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr Biol **13:** 843-848

Brigneti G, Voinnet O, Li WX, Ji LH, Ding SW, Baulcombe DC (1998) Viral pathogenicity determinants are suppressors of transgene silencing in Nicotiana benthamiana. EMBO J **17:** 6739-6746

Brown TA (2002) Genomes. BIOS Scientific Publishers Ltd, Oxford OX4 1RE, UK

Bryant FR, Benkovic SJ (1982) Phosphorothioate substrates for T4 RNA ligase. Biochemistry **21:** 5877-5885

Bushnell DA, Cramer P, Kornberg RD (2002) Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. Proc Natl Acad Sci U S A **99:** 1218-1222

Cameron V, Uhlenbeck OC (1977) 3'-Phosphatase activity in T4 polynucleotide kinase. Biochemistry **16:** 5120-5126

Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A **99 Suppl 4:** 16491-16498

Carrington JC, Cary SM, Parks TD, Dougherty WG (1989a) A second proteinase encoded by a plant potyvirus genome. EMBO J **8:** 365-370

Carrington JC, Freed DD, Sanders TC (1989b) Autocatalytic processing of the potyvirus helper component proteinase in Escherichia coli and in vitro. J Virol **63:** 4459-4463

Castro E (2001) Perl and cgi for the world wide web, Ed Second Edition. Peachpit Press, Berkeley, CA

Cedar H, Razin A (1990) DNA methylation and development. Biochim Biophys Acta **1049:** 1-8

Cherepanov AV, de Vries S (2002) Kinetic mechanism of the Mg2+-dependent nucleotidyl transfer catalyzed by T4 DNA and RNA ligases. J Biol Chem **277:** 1695-1704

Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. Bioinformatics **17:** 1093-1104

Chu CY, Rana TM (2006) Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol **4:** e210

Cronin S, Verchot J, Haldeman-Cahill R, Schaad MC, Carrington JC (1995) Long-distance movement factor: a transport function of the potyvirus helper component proteinase. Plant Cell **7:** 549-559

Crotty S, Cameron CE, Andino R (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci U S A **98:** 6895-6900

Cunningham PR, Richard RB, Weitzmann CJ, Nurse K, Ofengand J (1991) The absence of modified nucleotides affects both in vitro assembly and in vitro function of the 30S ribosomal subunit of Escherichia coli. Biochimie **73:** 789-796

Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC, Voinnet O (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science **313:** 68-71

Devic M, Jaegle M, Baulcombe D (1989) Symptom production on tobacco and tomato is determined by two distinct domains of the satellite RNA of cucumber mosaic virus (strain Y). J Gen Virol **70:** 2765-2774

Devic M, Jaegle M, Baulcombe D (1990) Cucumber mosaic virus satellite RNA (strain Y): analysis of sequences which affect systemic necrosis on tomato. J Gen Virol **71:** 1443-1449

DuBois P (2001) MySQL and Perl for the Web, Ed 1. New Riders Publishing, Indianapolis, IN

Dunoyer P, Himber C, Voinnet O (2006) Induction, suppression and requirement of RNA silencing pathways in virulent Agrobacterium tumefaciens infections. Nat Genet **38:** 258-263

Dunoyer P, Lecellier CH, Parizotto EA, Himber C, Voinnet O (2004) Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. Plant Cell **16:** 1235-1250

Eastberg JH, Pelletier J, Stoddard BL (2004) Recognition of DNA substrates by T4 bacteriophage polynucleotide kinase. Nucleic Acids Res **32:** 653-660

Ebhardt HA, Thi EP, Wang MB, Unrau PJ (2005) Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. Proc Natl Acad Sci U S A **102:** 13398-13403

Ebhardt HA, Wiese KC, Unrau PJ (2006) *Ebbie*: automated analysis and storage of small RNA cloning data using a dynamic web server. BMC Bioinformatics **7:** 18

Eick D, Kohlhuber F, Wolf DA, Strobl LJ (1994) Activation of pausing RNA polymerases by nuclear run-on experiments. Anal Biochem **218:** 347-351

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J **20:** 6877-6888

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res **8:** 186-194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8:** 175-185

Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science **310:** 1817-1821

Filipowicz W (2005) RNAi: the nuts and bolts of the RISC machine. Cell **122:** 17-20

Finnegan EJ, Margis R, Waterhouse PM (2003) Posttranscriptional gene silencing is not compromised in the Arabidopsis CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of Dicer-1 from Drosophila. Curr Biol **13:** 236-240

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature **391:** 806-811

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A **89:** 1827-1831

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005a) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr Biol **15:** 1494-1500

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005b) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr Biol **15:** 1494-1500

Gordon KH, Symons RH (1983) Satellite RNA of cucumber mosaic virus forms a secondary structure with partial 3'-terminal homology to genomal RNAs. Nucleic Acids Res **11:** 947-960

Gould AR, Palukaitis P, Symons RH, Mossop DW (1978) Characterization of a satellite RNA associated with cucumber mosaic virus. Virology **84:** 443-455

Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol **57:** 761-780

Gustafson AM, Allen E, Givan S, Smith D, Carrington JC, Kasschau KD (2005) ASRP: the Arabidopsis Small RNA Project Database. Nucleic Acids Res **33:** D637-40

Hamilton A, Voinnet O, Chappell L, Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. EMBO J **21:** 4671-4679

Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286:** 950-952

Hidaka S, Hanada K, Ishikawa K (1990) In vitro messenger properties of a satellite RNA of cucumber mosaic virus. J Gen Virol **71 (Pt 2):** 439-442

Hidaka S, Hanada K, Ishikawa K, Miura K (1988) Complete nucleotide sequence of two new satellite RNAs associated with cucumber mosaic virus. Virology **164:** 326-333

Hill KR, Hajjou M, Hu JY, Raju R (1997) RNA-RNA recombination in Sindbis virus: roles of the 3' conserved motif, poly(A) tail, and nonviral sequences of template RNAs in polymerase recognition and template switching. J Virol **71:** 2693-2704

Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, Koiwa H, Seki M, Shinozaki K, Fukuhara T (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. Plant Mol Biol **57:** 173-188

Ho CK, Wang LK, Lima CD, Shuman S (2004) Structure and mechanism of RNA ligase. Structure **12:** 327-339

Horstmann V, Huether CM, Jost W, Reski R, Decker EL (2004) Quantitative promoter analysis in Physcomitrella patens: a set of plant vectors activating gene expression within three orders of magnitude. BMC Biotechnol **4:** 13

Itaya A, Zhong X, Bundschuh R, Qi Y, Wang Y, Takeda R, Harris AR, Molina C, Nelson RS, Ding B (2007) A Structured Viroid RNA Is Substrate for Dicer-Like Cleavage to Produce Biologically Active Small RNAs but Is Resistant to RISC-Mediated Degradation. J Virol

Iyer LM, Balaji S, Koonin EV, Aravind L (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res **117:** 156-184

Jaegle M, Devic M, Longstaff M, Baulcombe D (1990) Cucumber mosaic virus satellite RNA (Y strain): analysis of sequences which affect yellow mosaic symptoms on tobacco. J Gen Virol **71:** 1905-1912

Janeway C, Travers P, Walport M, Shlomchik M (2004) **Immunobiology: The Immune System in Health and Disease**. Garland Publishing, New York, NY

Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol **57:** 19-53

Jung JH, Park CM (2006) MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. Planta

Kasschau KD, Carrington JC (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. Cell **95:** 461-470

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. Dev Cell **4:** 205-217

Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res **12:** 656-664

Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. Cell **115:** 209-216

Kurihara Y, Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A **101:** 12753-12758

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science **294:** 853-858

Lakatos L, Csorba T, Pantaleo V, Chapman EJ, Carrington JC, Liu YP, Dolja VV, Calvino LF, Lopez-Moya JJ, Burgyan J (2006) Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. EMBO J 25: 2768-2780

Lakatos L, Szittya G, Silhavy D, Burgyan J (2004) Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. EMBO J **23:** 876-884

Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science **294:** 858-862

Lee RC, Ambros V (2001) An extensive class of small RNAs in Caenorhabditis elegans. Science **294:** 862-864

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75:** 843-854

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J **23:** 4051-4060

Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol **15:** 1501-1507

Li S, Chou HH (2004) LUCY2: an interactive DNA sequence quality trimming and vector removal tool. Bioinformatics **20:** 2865-2866

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature **433:** 769-773

Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP (2003) The microRNAs of Caenorhabditis elegans. Genes Dev 17: 991-1008

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science 305: 1437-1441

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol **7:** 719-723

Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X (2003) R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science **301:** 1921-1925

Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science **297:** 2053-2056

Lot H, Marchoux G, Marrou J, Kaper JM, West CK, van Vloten-Doting L, Hull R (1974) Evidence for three functional RNA species in several strains of cucumber mosaic virus. J Gen Virol **22:** 81-93

Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ (2005) Elucidation of the small RNA component of the transcriptome. Science **309:** 1567-1569

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. Science **303:** 95-98

Maia IG, Bernardi F (1996) Nucleic acid-binding properties of a bacterially expressed potato virus Y helper component-proteinase. J Gen Virol **77 (Pt 5):** 869-877

Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM (2006) The evolution and diversification of Dicers in plants. FEBS Lett 580: 2442-2450

Margulies M, Egholm M, Altman WE, Attiya S, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:** 376-380

Marrou J (1973) Tomato necrosis. Ann Phytopathol **5:** 322

Marrou J, Marchoux G, Migliori A (1973) Contribution to the study of the cucumber mosaic virus infection progress in leaves of tobacco treated by actinomycin D. Ann Phytopathol **5:** 249-253

Masuta C, Suzuki M, Kuwata S, Takanami Y, Koiwai A (1993) Yellow mosaic symptoms induced by Y satellite RNA of cucumber mosaic virus is regulated by a single incompletely dominant gene in wild Nicotiana species. Phytopathology **83:** 411-413

Masuta C, Takanami Y (1989) Determination of sequence and structural requirements for pathogenicity of a cucumber mosaic virus satellite RNA (Y-satRNA). Plant Cell **1:** 1165-1173

Matzke MA, Primig M, Trnovsky J, Matzke AJ (1989) Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. EMBO J **8:** 643-649

Mendell JT (2005) MicroRNAs: critical regulators of development, cellular physiology and malignancy. Cell Cycle **4:** 1179-1184

Merai Z, Kerenyi Z, Kertesz S, Magna M, Lakatos L, Silhavy D (2006) Double-stranded RNA binding may be a general plant RNA viral strategy to suppress RNA silencing. J Virol **80:** 5747-5756

Mette MF, Matzke AJ, Matzke MA (2001) Resistance of RNA-mediated TGS to HC-Pro, a viral suppressor of PTGS, suggests alternative pathways for dsRNA processing. Curr Biol **11:** 1119-1123

Ming YZ, Di X, Gomez-Sanchez EP, Gomez-Sanchez CE (1994) Improved downward capillary transfer for blotting of DNA and RNA. BioTechniques **16:** 58-59

Muhlbach HP, Sanger HL (1979) Viroid replication is inhibited by alpha-amanitin. Nature **278:** 185-188

Nandakumar J, Shuman S (2004) How an RNA ligase discriminates RNA versus DNA damage. Mol Cell **16:** 211-221

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell **2:** 279-289

Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, et al. (2000) Genome sequence of Halobacterium species NRC-1. Proc Natl Acad Sci U S A **97:** 12176-12181

Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP (2000) Homologs of small nucleolar RNAs in Archaea. Science **288:** 517-522

Omer AD, Ziesche S, Ebhardt H, Dennis PP (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. Proc Natl Acad Sci U S A **99:** 5289-5294

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell **120:** 613-622

Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science **315:** 241-244

Papp I, Mette MF, Aufsatz W, Daxinger L, Schauer SE, Ray A, van der Winden J, Matzke M, Matzke AJ (2003) Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol **132:** 1382-1390

Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol **12:** 1484-1495

Parker JS, Barford D (2006) Argonaute: A scaffold for the function of short regulatory RNAs. Trends Biochem Sci **31:** 622-630

Petersen BO, Albrechtsen M (2005) Evidence implying only unprimed RdRP activity during transitive gene silencing in plants. Plant Mol Biol **58:** 575-583

Pirone TP, Thornbury DW (1984) The involvement of a helper component in nonpersistent transmission of plant viruses by aphids. Microbiol Sci **1:** 191-193

Rodriguez-Alvarado G, Roossinck MJ (1997) Structural analysis of a necrogenic strain of cucumber mosaic cucumovirus satellite RNA in planta. Virology **236:** 155-166

Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell **126:** 79-92

Pruss G, Ge X, Shi XM, Carrington JC, Bowman Vance V (1997) Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. Plant Cell **9:** 859-868

Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature **443:** 1008-1012

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev **20:** 3407-3425

Razin A, Cedar H (1991) DNA methylation and gene expression. Microbiol Rev **55:** 451-458

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature **403:** 901-906

Rodriguez-Alvarado G, Roossinck MJ (1997) Structural analysis of a necrogenic strain of cucumber mosaic cucumovirus satellite RNA in planta. Virology **236:** 155-166

Roossinck MJ, Sleat D, Palukaitis P (1992) Satellite RNAs of plant viruses: structures and biological effects. Microbiol Rev **56:** 265-279

Sambrook J, Russell DW (2001) **Molecular Cloning: A Laboratory Manual (Hardcover)** , Ed Third. Cold Spring Harbour, New York, NY

Santoro SW, Joyce GF (1997) A general purpose RNA-cleaving DNA enzyme. Proc Natl Acad Sci U S A **94:** 4262-4266

Schauer SE, Jacobsen SE, Meinke DW, Ray A (2002) DICER-LIKE1: blind men and elephants in Arabidopsis development. Trends Plant Sci **7:** 487-491

Schell J, Van Montagu M (1977) Transfer, maintenance, and expression of bacterial Ti-plasmid DNA in plant cells transformed with A. tumefaciens. Brookhaven Symp Biol **(29):** 36-49

Schiebel W, Haas B, Marinkovic S, Klanner A, Sanger HL (1993) RNA-directed RNA polymerase from tomato leaves. I. Purification and physical properties. J Biol Chem **268:** 11851-11857

Schwarz DS, Ding H, Kennington L, Moore JT, Schelter J, Burchard J, Linsley PS, Aronin N, Xu Z, Zamore PD (2006) Designing siRNA that distinguish between genes that differ by a single nucleotide. PLoS Genet 2: e140

Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. Cell **115:** 199-208

Schunmann PHD, Llewellyn DJ, Surin B, Boevink P, De Feyter RC, Waterhouse PM (2003a) A suite of novel promoters and terminators for plant biotechnology. Funct Plant Biol **30:** 443-452

Schunmann PHD, Surin B, Waterhouse PM (2003b) A suite of novel promoters and terminators for plant biotechnology - II. The pPLEX series for use in monocots. Funct Plant Biol **30:** 453-460

Shi H, Tschudi C, Ullu E (2006) An unusual Dicer-like1 protein fuels the RNA interference pathway in Trypanosoma brucei. RNA **12:** 2063-2072

Sleat DE, Palukaitis P (1990) Site-directed mutagenesis of a plant viral satellite RNA changes its phenotype from ameliorative to necrogenic. Proc Natl Acad Sci U S A **87:** 2946-2950

Sleat DE, Zhang L, Palukaitis P (1994) Mapping determinants within cucumber mosaic virus and its satellite RNA for the induction of necrosis in tomato plants. Mol Plant Microbe Interact **7:** 189-195

Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science **305:** 1434-1437

Sugino A, Snoper TJ, Cozzarelli NR (1977) Bacteriophage T4 RNA ligase. Reaction intermediates and interaction of substrates. J Biol Chem **252:** 1732-1738

Tabara H, Yigit E, Siomi H, Mello CC (2002) The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. Cell **109:** 861-871

Ullmann A, Jacob F, Monod J (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. J Mol Biol **24:** 339-343

Valli A, Martin-Hernandez AM, Lopez-Moya JJ, Garcia JA (2006) RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro. J Virol **80:** 10055-10063

Valoczi A, Hornyik C, Varga N, Burgyan J, Kauppinen S, Havelda Z (2004) Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res **32:** e175

van der Krol AR, Mur LA, Beld M, Mol JN, Stuitje AR (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell **2:** 291-299

Vanblokland R, Vandergeest N, Mol JNM, Kooter JM (1994) Transgene-mediated suppression of chalcone synthase expression in petunia-hybrida results from an increase in rna turnover. Plant Journal **6:** 861-877

Vargason JM, Szittya G, Burgyan J, Tanaka Hall TM (2003) Size selective recognition of siRNA by an RNA silencing suppressor. Cell **115:** 799-811

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev **20:** 759-771

Vaucheret H (2005) MicroRNA-dependent trans-acting siRNA production. Sci STKE **2005:** pe43

Walkey DGA (1991) **Applied Plant Virology** . Chapman & Hall, Thompson Scientific, Philadelphia, PA

Wang K, Herrera-Estrella L, Van Montagu M, Zambryski P (1984) Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from agrobacterium to the plant genome. Cell **38:** 455-462

Wang LK, Lima CD, Shuman S (2002) Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. EMBO J **21:** 3873-3880

Wang MB, Bian XY, Wu LM, Liu LX, Smith NA, Isenegger D, Wu RM, Masuta C, Vance VB, Watson JM, Rezaian A, Dennis ES, Waterhouse PM (2004a) On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. Proc Natl Acad Sci U S A 101: 3275-3280

Wang Y, Silverman SK (2006) Efficient RNA 5'-adenylation by T4 DNA ligase to facilitate practical applications. RNA **12:** 1142-1146

Wassenegger M, Heimes S, Riedel L, Sanger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell **76:** 567-576

Wassenegger M, Krczal G (2006) Nomenclature and functions of RNA-directed RNA polymerases. Trends Plant Sci **11:** 142-151

Waterman MS, Eggert M (1987) A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons. J Mol Biol **197:** 723-728

Waterworth HE, Tousignant ME, Kaper JM (1978) A lethal disease of tomato experimentally induced by rna 5 associated with cucumber mosaic virus isolated from commelina-diffusa from el-salvador. Phytopathology **68:** 561-566

Xie Z, Allen E, Wilken A, Carrington JC (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. Proc Natl Acad Sci U S A **102:** 12984-12989

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol 2: E104

Ye K, Malinina L, Patel DJ (2003) Recognition of small interfering RNA by a viral suppressor of RNA silencing. Nature **426:** 874-878

Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev **17:** 3011-3016

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X (2005) Methylation as a crucial step in plant microRNA biogenesis. Science **307:** 932-935

Yu B, Chapman EJ, Yang Z, Carrington JC, Chen X (2006) Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in Arabidopsis. FEBS Lett **580:** 3117-3120

Zhang L, Handa K, Palukaitis P (1994) Mapping local and systemic symptom determinants of cucumber mosaic cucumovirus in tobacco. J Gen Virol **75 (Pt 11):** 3185-3191

Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science **299:** 716-719

Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC, Jacobsen SE (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol **14:** 1214-1220