View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Leading Edge
BenchMarks
Cell

Small RNA Regulators of Gene Expression

Joel R. Neilson[1] and Phillip A. Sharp[1,2,*]
[1]Koch Institute for Integrative Cancer Research
[2]Department of Biology
Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
*Correspondence: sharppa@mit.edu
DOI 10.1016/j.cell.2008.09.006

This year, the Lasker Foundation recognizes Victor Ambros, Gary Ruvkun, and David Baulcombe for their pioneering work elucidating the role of short RNA species in the posttranscriptional regulation of eukaryotic gene expression.

Molecular biology is a young science, similar to a new continent full of territories that have not yet been explored. The discovery that small RNAs are central to gene regulation is a prime example of the importance of the unexplored territories of this young science. On September 26th, the 2008 Albert Lasker Basic Medical Research Award will be conferred upon the pioneers that made this discovery: Victor Ambros of the University of Massachusetts Medical School, Gary Ruvkun of Massachusetts General Hospital and Harvard Medical School, and David Baulcombe of the University of Cambridge.

Historically, pioneers have been ignored, with the significance of their discoveries only recognized years later. The wonderful papers published by Ambros and Ruvkun describing the first microRNA and its target (in an obscure journal named *Cell*) were ignored for nearly a decade. Meanwhile, Baulcombe labored in the forgotten (at least by US federal funding) field of plant molecular biology. His discovery that small RNA species are associated with homology-dependent silencing, also known as cosuppression or RNA interference, foreshadowed elucidation of the chemistry and biochemistry of these processes. The work of these talented researchers has fundamentally changed the scientific community's view of posttranscriptional gene regulation, demonstrating the breadth of biological processes in both plant and animal cells where short RNA species play a central role.

Worm Mutants and MicroRNAs

Part of the story begins with the description of two mutants in the nematode *C. elegans* by Chalfie, Horvitz, and Sulston in this journal in 1981. These two mutants, *unc-86* and *lin-4*, altered the normally invariant cell lineage progression in the roundworm *C. elegans*. In particular, various cell lineages in these mutants became "stuck" at a certain point in their development, reiterating patterns of cell division.

As a postdoctoral fellow in the Horvitz laboratory, Victor Ambros became fascinated with this phenotype. Based on the observation that the *lin-4* worm mutants are unable to lay eggs, he hypothesized that other mutations resulting in an inability to lay eggs might have related abnormalities. Examination of mutants with known egg-laying defects revealed that the *lin-14*, *lin-28*, and *lin-29* mutants were associated with abnormal cell lineage progression (Ambros and Horvitz, 1984). These mutations resulted in two types of irregular cell lineage progression—"precocious" events, in which certain developmental processes occurred before they would normally take place in wild-type animals, and "retarded" events, in which processes occurred after they would normally take place in wild-type animals. For this reason, the new mutants were collectively referred to as "heterochronic" mutants.

The most interesting gene uncovered in this study was *lin-14*, various alleles of which had opposite phenotypes of precocious or retarded cell fate progression. Interestingly, the semidominant retarded alleles of *lin-14* phenocopied defects observed in *lin-4* mutants, whereas recessive precocious alleles of *lin-14* resulted in developmental defects opposite of those found in *lin-4* mutants. Using 19 individual mutant and temperature-sensitive alleles of *lin-14*, Ambros shortly thereafter defined the retarded alleles of *lin-14* as gain-of-function alleles and the precocious alleles of *lin-14* as loss-of-function alleles and deduced that the two classes of alleles effected their phenotypes at distinct times during development. The aggregate of this data indicated that *lin-14* function normally decreased at later stages of development, raising two important questions: (1) How was *lin-14* function temporally regulated during development and (2) how did *lin-14* effect its function?

The most straightforward way to address these questions was to define the *lin-14* gene in molecular terms. To this end, Gary Ruvkun, another postdoctoral fellow in the Horvitz lab, led a fruitful collaboration that included Ambros. These studies revealed that the semidominant *lin-14* gain-of-function mutations were either deletions or rearrangements of the 3′ region of the *lin-14* gene. It was clear that the disrupted sequences encoded an element that negatively regulated *lin-14* gene activity, but it was not clear whether this occurred at the DNA, RNA, or protein level. By this point, both Ruvkun and Ambros had established their own groups at MGH/Harvard Medical School and Harvard University, respectively. The two researchers continued their work on the heterochronic worm genes and in 1989 independently published a pair of studies that further clarified understanding of these genes, in particular *lin-14*.

Ruvkun had raised an antibody to the *lin-14* gene product and demonstrated that this protein became localized to the nucleus of nearly all cell lineages known to be affected in *lin-14* mutants. Yet the expression was temporally limited—in wild-type animals lin-14 was present in

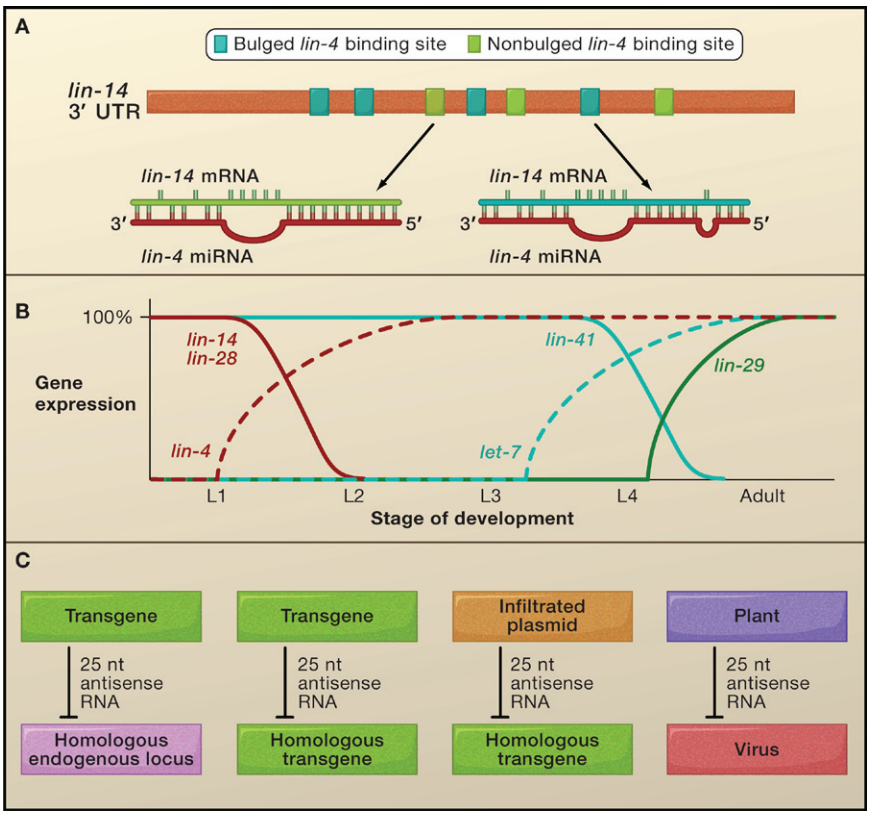

Figure 1. Interactions Revealing the Role of Small RNAs in Gene Regulation

(A) In *C. elegans*, the *lin-4* microRNA is complementary to seven individual binding sites in the 3′ UTR of the *lin-14* gene. The region of uninterrupted pairing between the 5′ end of the microRNA (the heptamer or octamer between bases 1 and 8) and the mRNA is known as the "seed" and is used for microRNA target prediction in animals. Curiously, in the originally defined microRNA/mRNA interaction, the binding sites characterized by "bulged" seed pairing architecture conferred repression on a heterologous UTR, whereas binding sites with uninterrupted seed pairing were less active (Ha et al., 1996). Figure adapted from Wightman et al. (1993).

(B) The stage-specific expression of heterochronic genes and their interactions as defined by Ambros and Ruvkun in adult worms and in worms at different stages of larval development (L1–L4). (Expression of RNA gene products, dotted lines; protein gene products, solid lines; direct interactions are indicated by like color.) Figure adapted from Reinhart et al. (2000).

(C) The four classes of gene silencing in which Hamilton and Baulcombe initially demonstrated the presence of 25 nucleotide (nt) RNA species that were antisense to the silenced sequences (Hamilton and Baulcombe, 1999).

cellular nuclei at the earliest (L1) stage of larval development but absent at the subsequent (L2) stage of development. In contrast, protein expression in *lin-14* gain-of-function mutants continued through all of the stages of larval development and was even observed in adult worms. These results complemented Ambros' genetic work, providing additional evidence that the decrease in expression of lin-14 protein at the L1/L2 transition was critical for normal development.

Two weeks after the appearance of Ruvkun's study, work authored by Ambros was published describing epistasis analysis of the *lin-4*, *lin-14*, *lin-28*, and *lin-29* mutants. By analyzing the worms carrying various combinations of these mutations, Ambros ordered the genes into a pathway where *lin-4* negatively regulated *lin-14* and *lin-28*. In turn, *lin-14* and *lin-28* negatively regulated *lin-29*, which even in the simultaneous absence of the other three genes was necessary and sufficient to drive the transition of *C. elegans* larvae into adults.

Although it was clear that the *lin-4* gene turned the *lin-14* and *lin-28* genes "off" in larval stages of development, the epistatic analysis could not unambiguously resolve whether *lin-4* regulated both of the genes separately or one through the other. Even so, given the similarities of *lin-14* gain-of-function and *lin-4* loss-of-function phenotypes, both investigators speculated that it was likely that *lin-4* was involved (perhaps directly) with the negative regulation of *lin-14* through the element deleted in *lin-14* gain-of-function mutants.

Roughly two years later, Ruvkun published back-to-back studies shedding additional light on the nature of *lin-14* regulation. Extensive molecular characterization of the genetic locus encoding *lin-14* revealed that the *lin-14* gain-of-function mutations localized to the 3′ untranslated region (UTR) of transcripts encoded by the *lin-14* gene, indicating that these mutations did not affect protein-coding sequence. Given that *lin-14* mRNAs in these mutants were expressed at normal levels, the mutations likely disrupted a negative regulatory element acting at the RNA level.

By analyzing the expression of the lin-14 protein in heterochronic mutant backgrounds, the Ruvkun group demonstrated that *lin-14* was antagonistically regulated by *lin-28* and *lin-4*. Specifically, whereas lin-14 protein expression was not reduced following the L1 stage in *lin-4* loss-of-function mutants, lin-14 protein expression prematurely disappeared during the L1 stage in a *lin-28* mutant background. These results clarified previous ordering of the pathway by Ambros and implied that the product of the *lin-4* gene (or something regulated by this product) directly bound the region of the *lin-14* 3′ UTR deleted in *lin-14* gain-of-function alleles. The remaining questions were quite clear: What was the molecular identity of *lin-4*, and did *lin-4* regulate *lin-14* directly or indirectly? Two years later, Ambros and Ruvkun published back-to back studies answering these questions.

The Ambros study described the molecular cloning of the *lin-4* gene (Lee et al., 1993). Astoundingly, the multi-kilobase genomic lesion producing the *lin-4* phenotype could be complemented by only 693 base pairs of genomic sequence. Corresponding regions from *C. briggsae*, *C. remenei*, and *C. vulgaris* also rescued the *lin-4* phenotype. However, whereas detailed analysis revealed two blocks of DNA sequence conservation among the four species of nematode, Ambros and his team were unable to identify strong

candidate protein-coding sequences or canonical start or stop codons within these regions. Indeed, lesions designed to disrupt potential coding sequences had no effect on *lin-4* rescue.

Undeterred, the authors moved forward. Northern analysis revealed that the *lin-4* locus produced two detectable RNA species—a minor 61 nucleotide product and a major ~21 nucleotide product. The two products were related, with the smaller product corresponding to the 5′-most 21 nucleotides of the larger. To demonstrate that these two species indeed effected *lin-4* gene function, the Ambros lab screened over 20,000 chromosomes to identify a second *lin-4* mutation, designated *lin-4(ma161)*. This mutant allele was a cytosine to thymidine transition altering position 5 of the two *lin-4* gene products. Therefore, one or both of these products must be involved in effecting *lin-4* gene function.

In the accompanying study, Ruvkun and coworkers followed up on their previous observation that *lin-14* gain-of-function mutations were localized to the 3′ UTR of *lin-14*. Immunoblot and RNase protection experiments from worms at different stages of development unequivocally demonstrated that the *lin-14* gene was posttranscriptionally regulated in a stage-specific manner (Figure 1; Wightman et al., 1993). In wild-type worms, although lin-14 protein was essentially absent following the first stage of larval development, the *lin-14* message was stably expressed in all larval and even adult stages. Using reporter transgenes in both *C. elegans* and *C. briggsae*, Ruvkun's team demonstrated that the 3′ UTR of *lin-14* was sufficient to confer the observed posttranscriptional regulation, and that this regulation was disrupted in two *lin-4* mutants—both the classical *lin-4(e912)* allele and the newly isolated *lin-4(ma161)* allele. Both groups noted that the small RNAs derived from the *lin-4* locus were complementary to seven partially repetitive sequences in the *lin-14* 3′ UTR that were deleted in whole or in part in *lin-14* gain-of-function alleles. Ambros and Ruvkun concluded that the short RNAs from the *lin-4* locus directly regulated the *lin-14* gene product through base-pairing to the repetitive sequences in the 3′ UTR of the *lin-14* message. Thus, the first microRNA, and the first microRNA target, had been identified.

Ruvkun's group essentially confirmed the model of direct interaction between *lin-4* and *lin-14* in a subsequent study. They demonstrated that point mutations within the predicted *lin-4*-binding sites of the *lin-14* 3′ UTR abolished stage-specific regulation of reporter transgenes in vivo and that transfer of predicted *lin-4* sites with a bulged architecture could confer stage-specific regulation upon a heterologous UTR (Ha et al., 1996). Shortly thereafter, Ambros's group defined *lin-28* as a second target of *lin-4* and generated and characterized a *lin-4* gain-of-function allele. A separate study demonstrated that *lin-4* functioned by blocking lin-14 protein synthesis after the initiation of translation of the *lin-14* message (Olsen and Ambros, 1999). Yet, the importance of these observations in relation to established paradigms of gene regulation remained unclear. There was no evidence for conservation of the *lin-4* gene product beyond *Caenorhabditis*, and even within this genus *lin-4* remained the sole example of a small RNA that worked in this manner.

This all changed in early 2000 when Ruvkun's group, in a collaborative study with the laboratories of Horvitz and Ann Rougvie, described isolation and characterization of the *let-7* gene (Figure 1; Reinhart et al., 2000). Like *lin-14*, *let-7* encoded a small RNA that negatively regulated the expression of a second gene (*lin-41*) through conserved elements in the 3′ UTR of that gene. Unlike *lin-14*, *let-7* was widely conserved across animal species (Pasquinelli et al., 2000). Twenty short months after the description of *let-7*, the groups of Ambros, David Bartel, and Thomas Tuschl simultaneously published studies illustrating that hundreds of genes like *lin-4* and *let-7* existed, and that this entire class of genes was conserved from worms to humans (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). MicroRNAs were once again introduced to the scientific world, and this time, the scientific world took notice.

Short RNAs and Gene Silencing in Plants

The finding that hundreds of genes encoding short regulatory RNAs exist in the animal genome was stunning. However, the importance of RNA species in regulating gene expression had been foreshadowed by studies in the plant kingdom.

In the early 1990s, David Baulcombe and others were working on engineering virus resistance in plants. The common strategy at the time was to actively interfere with viral replication by competing away necessary viral or host factors through transgenic overexpression of viral proteins or structured nucleic acid motifs. However, it soon became clear that viral resistance could be conferred by transgenes even if the transgenes did not encode protein sequence. Infection of a transgenic plant with a virus having sequence homology resulted in a reduction of steady-state levels of viral and transgenic RNA, even though transcription from the transgene remained unchanged. These observations led to a model in which viral resistance was mediated by a posttranscriptional mechanism that acted to degrade viral RNA. Similarities between resistance to viruses and homology-dependent gene silencing in transgenic plants suggested that the two processes were related.

Baulcombe's group strengthened the relationship between homology-dependent gene expression and viral resistance in a series of studies in plants published in the mid-1990s. The first study compared viral resistance across several transgenic lines, noting that transgenes conferring viral resistance produced low steady-state levels of transgenic RNA and were able to suppress in *trans* the accumulation of homologous RNA from other transgene loci (Mueller et al., 1995). The viral resistance conferred by these transgenes was characterized by an exquisite specificity: transgenic lines resistant to certain strains of potato virus X were not resistant to highly related strains of the virus.

A second study demonstrated that silencing of nonviral transgenes prevented accumulation of viruses engineered to contain sequences homologous to these transgenes (English et al., 1996). This revealed that sequence identity, rather than sequence origin, dictated the interactions between silenced transgenic loci and plant RNA viruses. Additional work using plants that had "recovered" from a strain of tomato black ring nepovirus infection

revealed that these plants were resistant to infection following secondary inoculation with the virus. Although the recovered plants remained sensitive to other types of viruses, if a noncoding region of the RNA from the virus in the primary inoculation was included in the heterologous virus used for the secondary inoculation, the plants were resistant to the second virus. The parallels between this virally induced protection and transgene-induced gene silencing led Baulcombe to conclude that the same RNA-based mechanism was likely to be active in both phenomena.

Shortly thereafter, Baulcombe's group used *Agrobacterium* infiltration of transgenic tobacco plants expressing green fluorescent protein (GFP) to demonstrate the existence of a systemic signal that could mediate sequence-specific gene silencing of transgenic loci in plants (Voinnet and Baulcombe, 1997). Systemic silencing of GFP could also be induced by localized bombardment with DNA-coated gold particles and could spread into nontransgenic tissue following grafting to rootstocks in which silencing had been initiated (Voinnet et al., 1998). The kinetics and pattern of spread of the silencing signal indicated that it was able to spread both from cell to cell via plasmodesmata and systemically through the phloem; the messenger was assumed to be a nucleic acid.

Still, although most evidence pointed to antisense RNA as the mediator of the various forms of posttranscriptional gene silencing—from cosuppression to antiviral activity—no one had identified RNA that was antisense to the sequences being targeted. In a 1996 review, Baulcombe speculated that methods used to look for antisense RNA associated with posttranscriptional gene silencing might have missed the "small or heterodisperse" RNAs that would be the likely products of plant-encoded RNA-dependent RNA polymerases. Working with Andrew Hamilton in 1999, he directly demonstrated that this was the case.

Examination of three transgenic tomato lines exhibiting transgene-induced posttranscriptional silencing (cosuppression) of an endogenous gene revealed sense and antisense ~25 nucleotide RNA species corresponding to the transgene (Figure 1; Hamilton and Baulcombe, 1999). Strikingly, these short RNA species were absent in other tomato plants with the same transgene that did not exhibit cosuppression. They next demonstrated that antisense RNA species were present in transgenic tobacco lines undergoing posttranscriptional gene silencing where the transgenes did not have any homology to endogenous sequence. Again, the RNA species were absent in a tobacco line where the transgenic locus was not silenced. The short antisense RNA species associated with silenced genes could be observed in two other models of gene silencing: during systemic posttranscriptional silencing after *Agrobacterium* infiltration and in plants infected with potato virus X. It was clear that the similarities among the different types of posttranscriptional gene silencing that had been noted again and again could be linked by a common denominator: short RNA species complementary to the element that was being silenced.

Future Perspectives

The creative studies of Victor Ambros, Gary Ruvkun, and David Baulcombe revealed the importance of small RNAs in the regulation of genes across diverse organisms. The biochemical analysis of RNA interference led to the discovery of small-interfering RNAs (siRNAs), which have become a ubiquitous tool for silencing genes in mammalian cells and hold promise as therapeutic agents to treat disease. MicroRNAs are thought to regulate over half of all mammalian genes, and alterations in their activities are associated with cancer, inflammation, neuronal development, and chronic heart disease. In plants, these RNAs regulate diverse developmental processes by triggering the generation of other *trans*-acting small RNAs. It is impossible to understand the biology of multicellular organisms without appreciation of the roles that small RNAs play.

It is exciting to contemplate that there may be many other types of small RNAs generated by different processes and with different functions. For example, piRNAs are a new class of small RNAs that are prevalent in germ tissue. These may control the expression of repetitive sequences in the germline, at least partially through epigenetic mechanisms. Although many longer noncoding RNAs are known to exist, in most cases their specific functions are yet to be identified. Examples include the relationship of the *Xist* RNA to inactivation of the X chromosome and of *U19* RNA to imprinting of the *Igf* locus. There are thousands of such RNAs produced in vertebrate systems. The breakthrough research of Ambros, Ruvkun, and Baulcombe underscores the notion that our understanding of the biology of noncoding RNAs has only just begun.

ACKNOWLEDGMENTS

We thank H.R. Horvitz and J.L. Goldstein for helpful comments on the manuscript and are grateful to M. Lindstrom for assistance with figures.

REFERENCES

Ambros, V., and Horvitz, H.R. (1984). Science *226*, 409–416.

English, J.J., Mueller, E., and Baulcombe, D.C. (1996). Plant Cell *8*, 179–188.

Ha, I., Wightman, B., and Ruvkun, G. (1996). Genes Dev. *10*, 3041–3050.

Hamilton, A.J., and Baulcombe, D.C. (1999). Science *286*, 950–952.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Science *294*, 853–858.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). Science *294*, 858–862.

Lee, R.C., and Ambros, V. (2001). Science *294*, 862–864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). Cell *75*, 843–854.

Mueller, E., Gibert, J., Davenport, G., Brigneti, G., and Baulcombe, D.C. (1995). Plant J. *7*, 1001–1013.

Olsen, P.H., and Ambros, V. (1999). Dev. Biol. *216*, 671–680.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., et al. (2000). Nature *408*, 86–89.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). Nature *403*, 901–906.

Voinnet, O., and Baulcombe, D.C. (1997). Nature *389*, 553.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D.C. (1998). Cell *95*, 177–187.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Cell *75*, 855–862.