The RNAi revolution. Nature

ARTICLE *in* NATURE · AUGUST 2004

Impact Factor: 41.46 · DOI: 10.1038/430161a · Source: PubMed

2 AUTHORS, INCLUDING:

Carl D Novina
Harvard Medical School
68 PUBLICATIONS **5,039** CITATIONS

The RNAi revolution

Carl D. Novina and Phillip A. Sharp

The term RNAi — short for RNA interference — crops up again and again in biology research these days. This is in part because of its power as a laboratory tool, and in part because it is a widespread natural phenomenon.

Arguably the most important advance in biology in decades has been the discovery that RNA molecules can regulate the expression of genes. For years, RNAs were thought to have just two broad functions in cells. Single-stranded messenger RNAs (mRNAs) are vital intermediaries in gene expression, transmitting information between DNA and protein. Ribosomal and transfer RNAs have structural, catalytic and information-decoding roles in the process of protein synthesis.

This picture became complicated in 1998, however, when Andrew Fire, Craig Mello and colleagues[1] announced their discovery of RNA interference (RNAi) — the silencing of gene expression by double-stranded RNA molecules — in nematode worms. Since then it has become clear that RNAi is an extremely useful experimental tool for learning what genes do. Not only that, it is an evolutionarily ancient method of genome defence in many organisms. Research on the mechanistic side has also proceeded at a staggering pace. Double-stranded RNA was originally found to silence genes by marking out their mRNA intermediaries for destruction. But at least two more mechanisms have been uncovered: blocking transcription (the synthesis of mRNAs from DNA) and inhibiting translation (the synthesis of proteins from mRNAs). And this is surely not the end of the story.

Short interfering RNAs (siRNAs)	A class of double-stranded RNAs of 21–22 nucleotides in length, generated from dsRNAs. siRNAs silence genes by promoting the cleavage of mRNAs with exactly complementary sequences, or recruiting inhibitory proteins to, or directing the modification of, DNAs with exactly complementary sequences.
MicroRNAs (miRNAs)	A class of 19–25-nucleotide, single-stranded RNAs that are encoded in the genomes of most multicellular organisms studied. Some are evolutionarily conserved and are developmentally regulated. They silence certain cellular genes at the stage of protein synthesis.
Tiny non-coding RNAs (tncRNAs)	A newly discovered class of short, 20–22-nucleotide RNAs that are encoded in the genome of *C. elegans*. They are not evolutionarily conserved, but some are developmentally regulated. Their function is still unknown.
Small modulatory RNA (smRNA)	A short, dsRNA, identified earlier this year in mice, that allows the expression of neuron-specific genes only in adult neurons.

Figure 1 **The short story of gene silencing — the main characteristics of short RNAs.**

Discovery

Like many significant advances, the discovery of RNAi was technically simple: injecting double-stranded RNAs (dsRNAs) into the worm *Caenorhabditis elegans* resulted in the silencing of a gene whose sequence was complementary to that of the dsRNAs[1]. All discoveries have precedents, however, and RNAi was no exception. In the late 1980s and early 1990s, plant biologists working with petunias were surprised to find that introducing numerous copies of a gene that codes for deep purple flowers led, not as expected, to an even darker purple hue, but rather to plants with white or patchy flowers[2,3]. Somehow, the introduced genes (the 'transgenes') had silenced both themselves and the plants' own 'purple-flower' genes. Similarly, when plants were infected with an RNA virus that had been genetically engineered to contain fragments of a plant gene, the plant's gene itself became silenced[4].

At the time these observations were puzzling, but they were readily understood when Fire and Mello's work was published. RNA viruses replicate through dsRNA intermediates; multi-copy transgenes can produce low levels of dsRNAs as well. The results hinted that RNAi, mediated through dsRNAs, is an ancient process, which predates the evolutionary divergence of plants and worms. RNAi thus naturally functions, it is thought, to silence viruses and rogue genetic elements that make dsRNA intermediates — types of RNA not usually produced by cells.

Mechanisms

So how does this silencing process work? An early step in determining the mechanism was the observation that plants that were silencing genes in an RNAi-related process (dubbed post-transcriptional gene silencing) produced short RNAs, some 20–25 nucleotides long, that matched the gene being silenced[5]. These RNAs are much shorter than typical mRNAs and ribosomal RNAs, perhaps explaining why this fundamental process remained undiscovered until recently.

At around the same time, the biochemical reactions of RNAi were reconstituted *in vitro* using fruitfly extracts[6,7]. These studies revealed that the long dsRNAs were diced up into short RNAs with a specific structure: two 21-nucleotide strands of RNA in a staggered duplex, with 19 nucleotides of dsRNA and two unpaired nucleotides at the ends[8]. This species was dubbed a short interfering RNA (siRNA; Fig. 1), and synthetic versions were shown to silence genes[9].

The enzyme responsible for cleaving dsRNAs into siRNAs is called Dicer[10]. In the next step in the process, the 'sense' strand of an siRNA — the strand that has exactly the same sequence as a target gene — is removed, leaving the 'antisense' strand, which is complementary to the target gene, to function in gene silencing (Fig. 2). The antisense strand guides a protein complex to the mRNA produced from the target gene, destroying it.

Early studies of transgene silencing in plants indicated that an RNAi-type process also works at the level of transcription (in a process called transcriptional gene silencing; reviewed in ref. 11). Further studies showed that both strands of a transgene are probably transcribed (rather than just one strand, as is the case for typical genes), generating a dsRNA rather than a single-stranded mRNA. The dsRNA is then processed into siRNA, which somehow initiates transcriptional silencing (Box 1). Simply put, this seems to involve the reversible addition of chemical groups to nuclear DNA and the proteins that package it as dictated by the sequence of the dsRNA[11].

As might be expected, there are nuances in the process of RNAi. For instance, in some organisms, gene silencing by RNAi requires the activity of an RNA-dependent RNA polymerase (RdRP), which uses the antisense strand of an siRNA as a primer with which to make more dsRNAs, thereby amplifying the process (Fig. 2). In plants, this amplification enables RNAi-mediated gene silencing to spread throughout non-reproductive (somatic) tissues, by cell-to-cell transfer of dsRNAs — generating widespread resistance

to viral infections. Amplification is not, however, thought to occur in vertebrates or fruitflies, as genes for a RdRP have not been found in these organisms.

Functions

Short interfering RNAs may have a general role in the silencing of transposable elements ('jumping genes'), repetitive genes (including transgenes) and viruses. About half of the sequences in the human genome, for instance, were generated by duplication and insertion of transposons — parasitic elements that create 'junk' DNA. Silencing of these mobile elements is crucial for the stability of the genome, and RNAi-related processes involving the generation of short RNAs are essential to this silencing process in many organisms.

For example, worms with mutations in genes of the RNAi pathway are unable to silence transposons in germline tissues (those that produce eggs and sperm)[12,13]. Similarly, plants with mutations in subsets of RNAi-pathway genes show defects in the silencing of specific mobile genetic elements[14].

The same process could also limit the degree to which a gene can be expressed in certain tissues. For example, fruitflies show 'dosage-dependent' gene silencing in a process related to RNAi[15]. Here, more gene copies result in higher gene expression — but only up to a point, beyond which gene expression is reduced. In this and similar cases, the proteins that package DNA within cells are modified to suppress transcription. Similarly, in fission yeast, siRNAs directly silence the expression of repetitive transposable elements at the level of transcription[16]. Intriguingly, this silencing can spread to nearby regions, repressing gene regions that are adjacent to the transposable elements (Box 1).

MicroRNAs

The existence of siRNAs produced in response to experimentally introduced long dsRNAs and transgenes prompted a search for short RNAs that are actually encoded by genes in the host genome. Researchers embarked on the process of purifying short RNAs (of 19–25 nucleotides) from several animal species, and were surprised to find, not siRNAs, but a new class of short RNAs, dubbed microRNAs (miRNAs; Fig. 1).

MicroRNAs are abundant, single-stranded RNAs, varying from a few thousand to 40,000 molecules per cell[17]. They are encoded in the genome, but, unlike mRNAs, do not represent a step towards the production of a protein. Instead, they regulate the expression of mRNAs. They are thought to be formed from 'hairpin' RNAs — RNAs that are effectively double-stranded by virtue of the fact that they are self-complementary and fold over in the middle. Like siRNAs, miRNAs are produced by the cleavage of dsRNAs by Dicer and related activities. They have been found in all multicellular organisms studied, and their encoding genes seem to make up 0.5–1% of the predicted genes in these organisms — 200–255 in humans, for instance[18]. So far, some 175 of these human miRNA genes have been confirmed by biochemical analysis[19].

On the basis of a model established from studies of nematode development, miRNAs in mature animals are thought to regulate the expression of distinct genes by base-pairing with partially complementary sequences in mRNAs and simply inhibiting their translation into proteins (Fig. 3). They therefore work by a somewhat different mechanism from siRNAs, in that transcription is not impeded and mRNAs are not destroyed. Another crucial difference is that siRNAs generally target the genes or genetic elements from which they originated, whereas miRNAs regulate separate genes — perhaps hundreds or more per miRNA. Furthermore, the degree of translational inhibition by miRNAs is thought to depend on how many of these molecules are bound to the target mRNA. Typically, such an mRNA contains many binding sites at one end (the 3′-untranslated region), and several different miRNAs can target the same 3′ region.

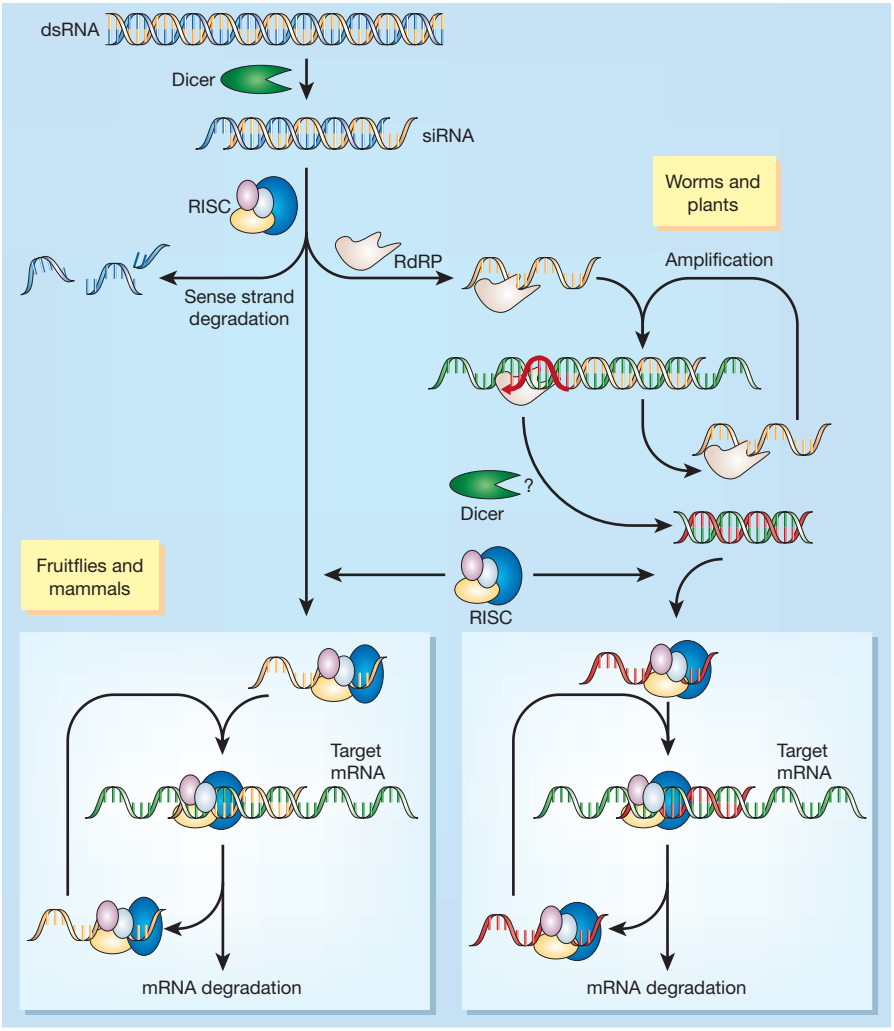


Figure 2 **Short interfering RNAs — post-transcriptional gene silencing. RNAi is triggered when a cell encounters a long double-stranded RNA (dsRNA), which might be produced from an introduced transgene, a viral intruder or a rogue genetic element. An enzyme called Dicer cleaves the long dsRNA into siRNAs. An RNA-induced silencing complex (RISC) then distinguishes between the different strands of the siRNA. The sense strand (blue) is degraded. The antisense strand (yellow) is used to target genes for silencing, and has one of several fates depending upon the organism. In fruitflies and mammals, the antisense strand is incorporated directly into RISC to target a complementary mRNA (green) for destruction. In the absence of siRNAs, the RISC lacks sequence-specific mRNA-binding properties. But when bound to the antisense strand, the now activated RISC can participate in repeated cycles of degradation of specific mRNAs, such that no protein is made — effectively silencing the gene from which the mRNAs are produced. In worms and plants, the antisense strand of the siRNA might first be used in an amplification process. The antisense strand, bound by an RNA-dependent RNA polymerase (RdRP) enzyme, can pair up with a complementary mRNA (green) and act as a start point for the synthesis of a new long dsRNA. Dicer is then required to generate new siRNAs (red), which are specific to different sequences on the same mRNA. Again, the target mRNA is destroyed.**

So what sort of biological processes do miRNAs regulate? In worms, the two founding members of the miRNA class of genes — *lin-4* and *let-7* — determine transitions between larval stages in post-embryonic development[20,21]. In plants, miRNAs also control crucial developmental transitions, and in flies they control cell division and cell death (reviewed in ref. 22). Human miRNAs were identified only a short time ago, and little is known about their function. As miRNAs are evolutionarily conserved, however, there can be little doubt about their importance in human physiology.

It can be difficult to identify the precise mRNA targets of miRNAs; computational approaches to the problem are complicated by the imperfect base-pairing between the two. Several studies have used rules inferred from known examples of miRNA–mRNA pairing, combined with a requirement for evolutionary conservation, to predict possible targets. One such study predicted an average of around five target mRNAs per miRNA[23]. Given that there are some 200 human miRNAs[18], it seems likely that several thousand human mRNAs could be targeted.

Common ground

Not all miRNAs work by binding with imperfect complementarity to mRNAs and preventing their translation. When miRNAs were first cloned from plants, it was recognized that many show perfect or near-perfect complementarity to cellular mRNAs, implying that these mRNAs might instead be targeted for degradation in an siRNA mode of action. This was indeed shown to be the case. Many of the plant mRNAs that are regulated by miRNA-directed cleavage encode gene-transcription factors that are important in plant development (reviewed in ref. 22). Recently, a conserved miRNA has been found in mammals which directs the cleavage of a target mRNA and, as in plants, regulates the expression of a transcription factor that is important for normal development[24]. Therefore, miRNA activities probably indirectly control the expression of vast numbers of genes.

Could siRNAs also work like miRNAs — inhibiting translation rather than targeting mRNAs for destruction? Our research with mammalian cells, using chemically synthesized siRNAs that were partially complementary to numerous target sites in a 'reporter' gene, showed that the siRNAs indeed inhibited the expression of the reporter at the stage of protein synthesis[25]. The major distinction, as noted above, is that siRNAs usually silence a target mRNA efficiently (through degradation) if the mRNA contains a single site that is almost exactly complementary to the short RNA. In contrast, the miRNA mode of action requires that the mRNA contains numerous partially complementary binding sites, which function synergistically. These observations may explain why siRNAs do not often inadvertently and potently silence mRNAs. Statistically, any 21-nucleotide RNA could partially base-pair with many mRNAs through continuous sequences of six to eight base pairs. But if translational silencing requires that many short RNAs bind the mRNA, the miRNA mode could be as specific as the siRNA mode.

Broadening the scope

Over the past couple of years, yet more mechanisms of action of short RNAs have been discovered — and potential new classes of such RNAs have been revealed. For instance, in the ciliate *Tetrahymena thermophila*, certain siRNAs are thought to specify sites in the organism's DNA that allow it to fragment[26]. The exact mechanism is unknown, but the discovery supports the contention that siRNAs can specifically recognize DNA sequences. DNA fragmentation in this process requires two genes whose relatives have been implicated in RNAi and related processes in other species. Ciliates lacking either gene do not accumulate siRNAs, and do not undergo DNA fragmentation.

Another class of short, 20–22-nucleotide RNAs in *C. elegans* was described last year[27] (Fig. 1). These tiny non-coding RNAs (tncRNAs) do not have highly complementary sequences nearby in the genome, and so may not be processed from hairpins, although most tncRNAs seems to be produced by Dicer. Unlike many miRNAs, the sequences of tncRNAs are not commonly conserved among related species. Although their functions are unknown, many are complementary to mRNAs and might target them for degradation.

Finally, earlier this year, a short, non-protein-coding, 20-nucleotide dsRNA, called a small modulatory RNA (smRNA), was isolated from neural stem cells in the hippocampus of adult mice[28]. This dsRNA has the same sequence as a 21–23-nucleotide stretch of DNA called the neuron-restrictive silencer element. This is evolutionarily conserved in vertebrates and is bound by a regulatory protein

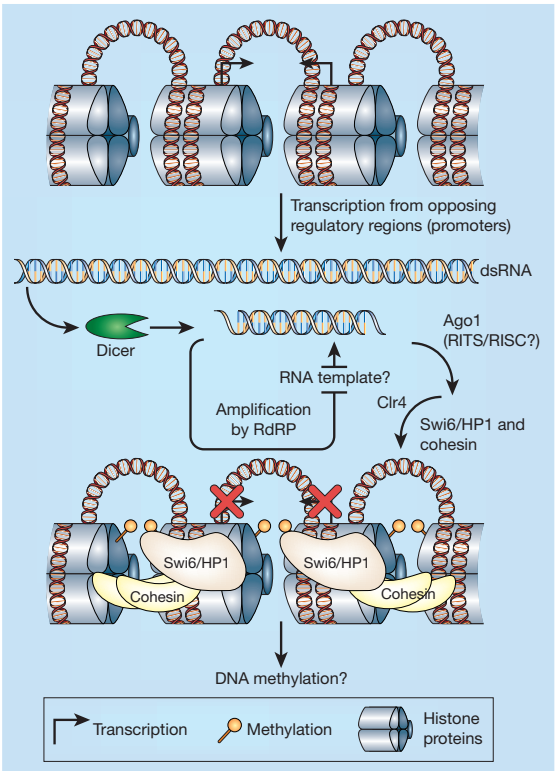
Box 1 Transcriptional gene silencing

One mechanism of action of siRNAs involves silencing gene transcription. This occurs through its effects on chromatin, the compact form of DNA, in which DNA is wrapped around complexes of histone proteins.

Long dsRNAs of similar sequence to genomic DNA might be produced in flies and mammals by symmetrical transcription from opposing promoters; this often occurs when many copies of transgenes or other repeated sequences are present, and in plants and worms through the activity of an RNA-dependent RNA polymerase (RdRP; see Fig. 1).

In plants, worms and flies, siRNAs produced from these dsRNAs can direct the formation of silenced DNA regions. As described in Fig. 1, the sense strand of the siRNA is degraded; the antisense strand is used to recruit proteins that inhibit transcription. The antisense strand identifies the correct position along the DNA sequence, by base-pairing either with DNA or with newly forming mRNA transcripts.

The details are best understood in fission yeast. Here, the Ago1 protein acts in the RNA-induced silencing complex (RISC), or in an RNA-induced initiation of transcriptional gene silencing (RITS) complex[30], to distinguish the antisense strand from the sense strand. This strand helps to recruit an enzyme that adds methyl groups to histone protein H3, forming heterochromatin, the silenced form of chromatin.



The methylated histones promote the recruitment of proteins that can spread the silencing state[16]. In the fully repressed state of chromatin, the DNA may also be methylated by processes requiring siRNA, although this has only been seen in plants.

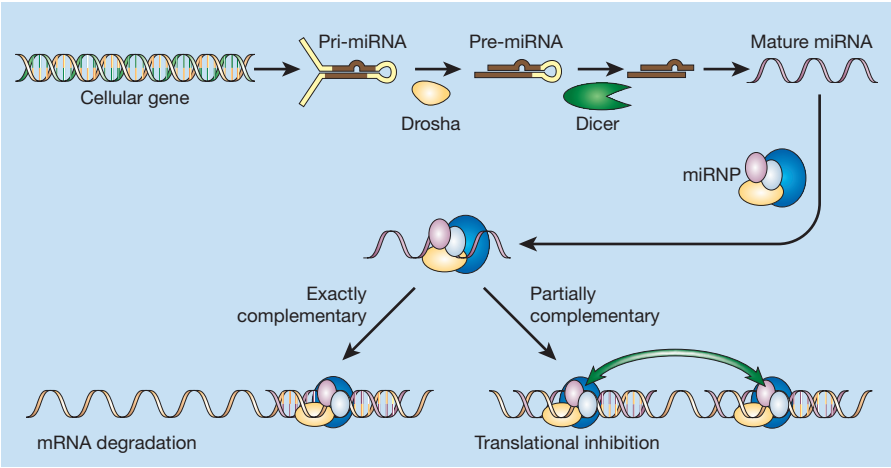


Figure 3 **MicroRNAs — translational silencing. miRNAs are encoded in the genomes of all multicellular organisms studied so far. They can be expressed either from an intergenic cluster, or from single genetic regions. Both types fold into long hairpin RNAs called primary microRNAs (pri-miRNAs), which show imperfect internal sequence complementarity (the bulge represents a region where there is no complementarity). A nuclear enzyme, Drosha, cleaves pri-miRNAs into smaller, roughly 70-nucleotide hairpin RNAs termed precursor miRNAs (pre-miRNAs)[31]. These are exported from the nucleus to the cytoplasm, where another enzyme, Dicer, cleaves them into mature, single-stranded miRNAs, 21–22 nucleotides long. A mature miRNA is then assembled into a ribonucleoprotein (miRNP) complex. This complex most commonly binds to the 3'-untranslated sequences of particular mRNAs through partially complementary sequences, and prevents the mRNAs from being translated into protein. Base-pairing with the 7–8 nucleotides near the 5' terminus of the miRNA is essential here. Numerous miRNP complexes bind to the 3'-untranslated sequences of mRNAs and cooperate for effective translational silencing through interactions (green arrow) that are still unclear. In a few cases, the miRNA is exactly or nearly exactly complementary to a site in an mRNA, and this results in cleavage and degradation similar to that observed with siRNAs (see Fig. 1). This has led to the proposal that siRNAs may function by being recognized by the miRNP complex. There is genetic evidence in plants and worms, however, that different members of an extended family of proteins are required for the miRNA and siRNA pathways — so in these organisms the two pathways may be distinct.**

that helps to prevent neuron-specific genes from being expressed in non-neuronal cells. In neuronal cells, where the neuronal genes are expressed, the smRNA interacts with the regulatory protein, and promotes gene expression and neuronal development. It is unclear exactly how smRNA acts, but it will be interesting to see whether it is the founding member of a new class of short RNAs.

Into the lab and clinic

RNAi promises to be one of the most useful laboratory tools yet. Classical genetic analysis of gene function involves identifying an organism with an abnormal set of physical and behavioural characteristics, and then isolating the underlying mutant genes. Typically, this 'forward genetic' approach is limited to rapidly reproducing organisms. Genetic experiments in mammals have been limited to 'reverse genetic' approaches, which begin with knocking out a specific gene to identify its function. But many of these approaches are laborious and expensive. In contrast, silencing genes with RNAi makes it possible to analyse many gene networks in mammals, and offers a higher-throughput approach to the analysis of model organisms such as *C. elegans*.

In *C. elegans*, the experiments are very simple: inject worms with dsRNA, soak them in a solution of dsRNA or feed them bacteria expressing dsRNA, and the target gene is silenced! There are some 19,000 genes in worms, and libraries of 12,000 different dsRNAs have been used to screen this animal for genes involved in complex phenomena such as obesity and ageing. Similarly, purified long dsRNAs have been used to define the roles of specific genes in cholesterol metabolism and heart formation in fruitflies.

The use of chemically synthesized siRNAs has also helped to define gene function in vertebrate cells[9], and has even been extended to human cells *in vitro*. Human cells are usually killed by long dsRNAs. Short RNA segments (of 21–22 base pairs) do not activate this killing process, but still possess enough sequence complexity to specify the silencing of a particular gene[9]. Libraries of short RNAs, or of DNA vectors encoding short RNAs, have been generated to target some 8,000 of the roughly 35,000 human genes to help determine their functions. Such screens have already identified gene activities in diseases such as cancer. Short RNAs might even be used to treat such diseases. A major obstacle to therapeutic gene silencing is the 'delivery problem' — the necessity of introducing short dsRNAs into specific organs. Emerging evidence suggests, however, that this obstacle might be surmountable.

Speculations

What biological phenomena might arise from RNAi-based gene regulation? In theory, any 21–22-nucleotide sequence in the genome could be converted into a potential regulatory factor by transcription into dsRNA. Consider the following series of events: first, a mutation generates an inverted duplication of a genomic sequence; second, transcription produces a hairpin RNA; and third, Dicer produces an siRNA or miRNA. These events would create a factor that could regulate another gene at the level of mRNA degradation, translation or, possibly, transcription. This series of events seems more likely than that required to create new regulatory possibilities with sequence-specific RNA- or DNA-binding proteins. Thus, we might expect more evolutionary variation in RNAs that control gene regulation than is observed in regulatory proteins.

RNAi-related processes could also be important for surviving periods of stress. For example, if an organism's environment changes radically, extensive alterations might occur in the transcription of both strands of genomic DNA[29]. Large segments of the genome could thereby be 'sampled' for short RNA regulators that are advantageous for survival. At present, such events are all conjecture — made possible by the discovery of the many roles of short RNAs in gene regulation and biology. ■

Carl D. Novina and Phillip A. Sharp are at the Center for Cancer Research, and Phillip A. Sharp is in the Department of Biology and the McGovern Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mails: sharppa@mit.edu; cdnovina@mit.edu

1. Fire, A. *et al. Nature* **391,** 806–811 (1998).
2. Napoli, C., Lemieux, C. & Jorgensen, R. *Plant Cell* **2,** 279–289 (1990).
3. van der Krol, A. R. *et al. Plant Cell* **2,** 291–299 (1990).
4. Wassenegger, M. *et al. Cell* **76,** 567–576 (1994).
5. Hamilton, A. J. & Baulcombe, D. C. *Science* **286,** 950–952 (1999).
6. Tuschl, T. *et al. Genes Dev.* **13,** 3191–3197 (1999).
7. Zamore, P. D. *et al. Cell* **101,** 25–33 (2000).
8. Elbashir, S. M. *et al. Genes Dev.* **15,** 188–200 (2001).
9. Elbashir, S. M. *et al. Nature* **411,** 494–498 (2001).
10. Bernstein, E. *et al. Nature* **409,** 363–366 (2001). See also www.nature.com/focus/rnai/animations/animation/animation.htm
11. Matzke, M. *et al. Biochim. Biophys. Acta* **1677,** 129–141 (2004).
12. Ketting, R. F. *et al. Cell* **99,** 133–141 (1999).
13. Tabara, H. *et al. Cell* **99,** 123–132 (1999).
14. Xie, Z. *et al. PLoS Biol.* **2,** E104 (2004).
15. Pal-Bhadra, M. *et al. Science* **303,** 669–672 (2004).
16. Schramke, V. & Allshire, R. *Science* **301,** 1069–1074 (2003).
17. Lim, L. P. *et al. Genes Dev.* **17,** 991–1008 (2003).
18. Lim, L. P. *et al. Science* **299,** 1540 (2003).
19. Griffiths-Jones, S. *Nucleic Acids Res.* **32,** D109–D111 (2004).
20. Lee, R. C., Feinbaum, R. L. & Ambros, V. *Cell* **75,** 843–854 (1993).
21. Reinhart, B. J. *et al. Nature* **403,** 901–906 (2000).
22. Bartel, D. P. *Cell* **116,** 281–297 (2004).
23. Lewis, B. P. *et al. Cell* **115,** 787–798 (2003).
24. Yekta, S. *et al. Science* **304,** 594–596 (2004).
25. Doench, J. G. *et al. Genes Dev.* **17,** 438–442 (2003).
26. Mochizuki, K. *et al. Cell* **110,** 689–699 (2002).
27. Ambros, V. *et al. Curr. Biol.* **13,** 807–818 (2003).
28. Kuwabara, T. *et al. Cell* **116,** 779–793 (2004).
29. Causton, H. C. *et al. Mol. Biol. Cell* **12,** 323–337 (2001).
30. Verdel, A. *et al. Science* **303,** 672–676 (2004).
31. Lee, Y. *et al. Nature* **425,** 415–419 (2003).