RNA Interference: Biology, Mechanism, and Applications

Neema Agrawal, P. V. N. Dasaradhi, Asif Mohmmed, Pawan Malhotra, Raj K. Bhatnagar, and Sunil K. Mukherjee*

International Center for Genetic Engineering and Biotechnology, New Delhi 110 067, India

INTRODUCTION

RNA silencing is a novel gene regulatory mechanism that limits the transcript level by either suppressing transcription (transcriptional gene silencing [TGS]) or by activating a sequence-specific RNA degradation process (posttranscriptional gene silencing [PTGS]; RNA interference [RNAi]). Although there is a mechanistic connection between TGS and PTGS, TGS is an emerging field while PTGS is undergoing an explosion in its information content. Here, we have limited our discussion to PTGS/RNAi-related phenomena.

Pioneering observations on PTGS/RNAi were reported in plants, but later on RNAi-related events were described in almost all eukaryotic organisms, including protozoa, flies, nematodes, insects, parasites, and mouse and human cell lines, as shown in Table 1. Three phenotypically different but mechanistically similar forms of RNAi, cosuppression or PTGS in
plants, quelling in fungi, and RNAi in the animal kingdom, have been described. More recently, micro-RNA formation, heterochromatinization, etc., have been revealed as other facets of naturally occurring RNAi processes of eukaryotic cells.

During the occurrence of RNAi/PTGS, double-stranded RNA (dsRNA) molecules, which cleave the inducer molecules first (16) and eventually destroy the cellular or viral cognate mRNA molecules (called the target) (17) act as inducers or activators of this process. As a result, the target mRNAs cannot accumulate in the cytosol, although they remain detectable by nuclear run-on assays (73). Jorgensen coined the term cosuppression to describe the loss of mRNAs of both the endo- and the transgene.

### UNRAVELING RNA SILENCING

In order to understand the process of homology-dependent RNA silencing, it would be prudent to overview the process itself and describe its important features. In the later part of this review, the genetics, biochemistry, and potential therapeutic applications of the process will be dealt with.

#### PTGS in Plants

In plants, the RNA silencing story unfolded serendipitously during a search for transgenic petunia flowers that were expected to be more purple. In 1990, R. Jorgensen’s laboratory wanted to upregulate the activity of a gene for chalcone synthase (chsA), an enzyme involved in the production of anthocyanin pigments. Surprisingly, some of the transgenic petunia plants harboring the chsA coding region under the control of a 35S promoter lost both endogene and transgene chalcone synthase activity, and thus many of the flowers were variegated or developed white sectors (163). Jorgensen coined the term cosuppression to describe the loss of mRNAs of both the endo- and the transgene.

### TABLE 1. Eukaryotic organisms exhibiting RNAi-related phenomena

<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Species</th>
<th>Stage tested</th>
<th>Delivery method</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protozoans</td>
<td>Trypanosoma brucei</td>
<td>Procyclic forms</td>
<td>Transfection</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Plasmodium falciparum</td>
<td>Blood stage</td>
<td>Electroporation and soaking</td>
<td>143, 150</td>
</tr>
<tr>
<td></td>
<td>Toxoplasma gondii</td>
<td>Mature forms in fibroblast</td>
<td>Transfection</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Leishmania donovani</td>
<td>Mature form</td>
<td>Transfection and feeding</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tried but not working</td>
<td>183</td>
</tr>
<tr>
<td>Invertebrates</td>
<td>Caenorhabditis elegans</td>
<td>Larval stage and adult stage</td>
<td>Transfection, feeding bacteria</td>
<td>26, 31</td>
</tr>
<tr>
<td></td>
<td>Caenorhabditis briggsae</td>
<td>Adult</td>
<td>Injection</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Brugia malayi (filarial worm)</td>
<td>Adult worm</td>
<td>Soaking</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Schistosoma mansoni</td>
<td>Sporocysts</td>
<td>Soaking</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Hydra</td>
<td>Adult</td>
<td>Delivered by micropipette</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Planaria</td>
<td>Adult</td>
<td>Soaking</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Lymania stagnalis (snail)</td>
<td>Adult</td>
<td>Injection</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Drosophila melanogaster</td>
<td>Cell lines, adult, embryio</td>
<td>Injection for adult and embryonic stages, soaking and transfection for cell lines</td>
<td>96, 114, 155</td>
</tr>
<tr>
<td></td>
<td>Cyclorphan (fly)</td>
<td>Early embryonic stages</td>
<td>Injection</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Milkweed bug</td>
<td>Early embryonic stages</td>
<td>Injection</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Beetle</td>
<td>Early embryonic stages</td>
<td>Injection</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Cockroach</td>
<td>Larval stage</td>
<td>Injection</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>Spodoptera frugipenda</td>
<td>Adult and cell line</td>
<td>Injection and soaking</td>
<td>176, 215</td>
</tr>
<tr>
<td>Vertebrates</td>
<td>Zebra fish</td>
<td>Embryo</td>
<td>Microinjection</td>
<td>224</td>
</tr>
<tr>
<td></td>
<td>Xenopus laevis</td>
<td>Embryo</td>
<td>Injection</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Mice</td>
<td>Prenatal, embryonic stages, and adult</td>
<td>Injection</td>
<td>31, 229</td>
</tr>
<tr>
<td></td>
<td>Humans</td>
<td>Human cell lines</td>
<td>Transfection</td>
<td>42</td>
</tr>
<tr>
<td>Plants</td>
<td>Monocots/dicots</td>
<td>Plant</td>
<td>Particle bombardment with siRNA/transgenics</td>
<td>88</td>
</tr>
<tr>
<td>Fungi</td>
<td>Neurospora crassa</td>
<td>Filamentous fungi</td>
<td>Transfection</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Schizosaccharomyces pombe</td>
<td>Filamentous fungi</td>
<td>Transgene</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>Dictyostelium discoideum</td>
<td>Filamentous fungi</td>
<td>Transgene</td>
<td>147</td>
</tr>
<tr>
<td>Algae</td>
<td>Chlamydomonas reinhardtii</td>
<td></td>
<td>Transfection</td>
<td>231</td>
</tr>
</tbody>
</table>
Around the same time, two other laboratories (105, 217) also reported that introduction of the transcribing-sense transgenes could downregulate the expression of homologous endogenous genes. Subsequently, many similar events of cosuppression were reported in the literature. All cases of cosuppression resulted in the degradation of endogene and transgene RNAs after nuclear transcription had occurred (120). Since posttranscriptional RNA degradation was observed in a wide range of transgenes expressing the plant, bacterial, or viral sequences, it was rechristened posttranscriptional gene silencing (PTGS). PTGS could be initiated not only by sense transgenes but also by antisense transgenes, and biochemical evidence suggests that similar mechanisms might operate in both cases (81). It is worthwhile to point out to that although the cosuppression phenomenon was originally observed in plants, it is not restricted to plants and has also been demonstrated in metazoans and mammals (98).

In keeping with the times, the observed alterations in the PTGS-related phenotypes were attributed to multiple-site integrations, aberrant RNA formations, repeat structures of the transgenes, etc. Later on, it became clear that the expression of the transgene led to the formation of dsRNA, which, in turn, initiated PTGS. For example, in the case of cosuppressed petunia plants, chs4 mRNA formed a partial duplex, since there are regions of self-complementarity located between chs4 3′ coding region and its 3′ untranslated region (154). This was revealed by DNA sequence analysis and experimental detection of in vitro-transcribed, RNase-resistant duplex chs4 RNA. In an independent study, a p35S-ACC (1-aminocyclopropane-1-carboxylate [ACC] oxidase) sense transgene carrying a small inverted repeat in the 5′ untranslated region was introduced into tomato to test the role of dsRNA structure as an inducer of PTGS. Cosuppression of the endogenous acc gene occurred at a higher frequency in these plants than in those harboring only the p35S-ACC sense transgene without the inverted repeat (93).

Reports from several laboratories in the past few years have established that the loss in steady-state accumulation of the target mRNA is almost total if the designed transgene construct of the transgenic plant produces the nuclear transcript in a homology-dependent manner. The silencing spread not only from the gut of the worm to the remainder of the body, but also through the germ line to the wild-type strain, whereas the native all mRNA was highly reduced, indicating that quelling and not the rate of transcription affected the level of mature mRNA in a homology-dependent manner.

The phenomenon of RNAi first came into the limelight following the discovery by Fire et al. (78), who unequivocally demonstrated the biochemical nature of inducers in gene silencing by introducing purified dsRNA directly into the body of Caenorhabditis elegans. The investigators injected dsRNA corresponding to a 742-nucleotide segment of unc22 into either the gonad or body cavity region of an adult nematode. unc22 encodes an abundant but nonessential myofibril protein, and the decrease in unc22 activity is supposed to produce an increasingly severe twitching phenotype. The injected animal showed weak twitching, whereas the progeny individuals were strong twitchers. The investigators showed that similar loss-of-function individuals could also be generated with dsRNAs corresponding to four other nematode genes. The phenotypes produced by interference by various dsRNAs were extremely specific.

This experiment paved the way for easy production of null mutants, and the process of silencing a functional gene by exogenous application of dsRNA was termed RNA interference (RNAi). RNAi in C. elegans was also initiated simply by soaking the worms in a solution containing dsRNAs or feeding the worms Escherichia coli organisms that expressed the dsRNAs (209). This is a very potent method, requiring only catalytic amounts of dsRNA per cell to silence gene expression. The silencing spread not only from the gut of the worm to the remainder of the body, but also through the germ line to several generations. These phenomena of RNAi have also been demonstrated to occur in Drosophila melanogaster and many other invertebrates and vertebrates.

Quelling and RNAi

While reports of PTGS in plants were piling up, homology-dependent gene silencing phenomena were also observed independently in fungal systems. These events were called quelling. Quelling came to light during attempts to boost the production of an orange pigment made by the gene all of the fungus Neurospora crassa (50). An N. crassa strain containing a wild-type all gene (orange phenotype) was transformed with a plasmid containing a 1,500-bp fragment of the coding sequence of the all gene. A few transformants were stably quelled and showed albino phenotypes. In the all-quelled strain, the level of unspliced all mRNA was similar to that of the wild-type strain, whereas the native all mRNA was highly reduced, indicating that quelling and not the rate of transcription affected the level of mature mRNA in a homology-dependent manner.

The loss of function in a transgene was also initiated simply by feeding the worms the dsRNA (209). This is a very potent method, requiring only catalytic amounts of dsRNA per cell to silence gene expression. The silencing spread not only from the gut of the worm to the remainder of the body, but also through the germ line to several generations. These phenomena of RNAi have also been demonstrated to occur in Drosophila melanogaster and many other invertebrates and vertebrates.

Insights from Virus-Infected Plants
(Virus-Induced Gene Silencing)

Besides the processes mentioned above, homology-driven RNA degradation also occurs during the growth of viral genomes in infected plants (73). Viruses can be either the source, the target, or both the source and the target of silencing. PTGS
mediated by viruses can occur with RNA viruses, which replicate in the cytoplasm, and also with DNA viruses, which replicate in the nucleus (71). As early as in the 1920s, it was known that plants could be protected from a severe virus by prior infection with a mild strain of a closely related virus. Although the mechanism of such cross protection in plants remained unknown for a long time, such phenomena could be explained partly in terms of PTGS that could be induced by the mild strain and targeted later against the virulent viral genome. It was also found that transforming plants with virus-derived transgenes gave protection against the challenge viruses even when no transgene protein was produced (132).

Analyses of these virus-resistant plants revealed that the transgenes were highly transcribed in the nucleus, whereas the steady-state level of cytoplasmic mRNA was very low. Further analysis suggested that some of the transgenic mRNA molecules assumed the conformation of dsRNA, which triggered sequence-specific degradation of self and other homologous or foreign cRNA sequences in the cytoplasm. Thus, in the virus-resistant lines, not only the transgene mRNAs but also the mRNA from the homologous endogenous gene and the invading viral RNA (with homology to the transgene) were degraded.

Another form of virus-induced gene silencing is the phenomenon of viral recovery itself. When *Brassica napus* was inoculated with *cauliflower mosaic virus* (a DNA virus), lesions at the site of virus entry were visible 5 to 7 days postinoculation. Symptoms of systemic infections were apparent by 10 to 14 days postinoculation. Symptoms were most prominent at 30 to 40 days postinoculation and declined thereafter (i.e., the plants recovered), with the newly emergent leaves remaining asymptomatic at 50 days postinoculation (5).

Figure 1 diagrammatically illustrates the systemic spread of RNAi in plants. Such recovery occurred by a PTGS-like mechanism because 19S and 35S RNAs encoded by the *cauliflower mosaic virus* (a DNA virus), lesions at the site of virus entry were visible 5 to 7 days postinoculation. Symptoms of systemic infections were apparent by 10 to 14 days postinoculation. Symptoms were most prominent at 30 to 40 days postinoculation and declined thereafter (i.e., the plants recovered), with the newly emergent leaves remaining asymptomatic at 50 days postinoculation (5).

Figure 1 diagrammatically illustrates the systemic spread of RNAi in plants. Such recovery occurred by a PTGS-like mechanism because 19S and 35S RNAs encoded by the *cauliflower mosaic virus* (a DNA virus), lesions at the site of virus entry were visible 5 to 7 days postinoculation. Symptoms of systemic infections were apparent by 10 to 14 days postinoculation. Symptoms were most prominent at 30 to 40 days postinoculation and declined thereafter (i.e., the plants recovered), with the newly emergent leaves remaining asymptomatic at 50 days postinoculation (5).

Based on the principles of virus-induced gene silencing, vectors designed with the genome sequence of RNA viruses tobacco mosaic virus, potato virus X, and tobacco rattle virus are being widely used to knock down the expression of host genes. The characteristics of many plant genes were revealed by observing the loss-of-function-related phenotypic changes when the recombinant vectors incorporating the concerned host genes were introduced into plants (136). Of these vectors, the TRV-based are more promising because these are capable of inducing-meristematic gene silencing, which has not been possible to achieve with other RNA virus-based vectors. Meristematic gene silencing employing TGMV vectors has also been reported (173). Thus, virus-induced gene silencing-based techniques are extremely useful for studies related to functional genomics in plants.

**IMPORTANT FEATURES OF RNA SILENCING**

Independently of one another, investigations on diverse organisms, labeled variously as PTGS in plants, RNAi in animals, silencing in fungi, and virus-induced gene silencing, have converged on a universal paradigm of gene regulation. The critical common components of the paradigm are that (i) the inducer is the dsRNA, (ii) the target RNA is degraded in a homology-dependent fashion, and, as we will see later, (iii) the degradation...
tive machinery requires a set of proteins which are similar in structure and function across most organisms. In most of these processes, certain invariant features are observed, including the formation of small interfering RNA (siRNA) and the organism-specific systemic transmission of silencing from its site of initiation.

siRNA

The key insight in the process of PTGS was provided from the experiments of Baulecombe and Hamilton (92), who identified the product of RNA degradation as a small RNA species (siRNA) of ~25 nucleotides of both sense and antisense polarity. siRNAs are formed and accumulate as double-stranded RNA molecules of defined chemical structures, as mentioned later. siRNAs were detected first in plants undergoing either cosuppression or virus-induced gene silencing and were not detectable in control plants that were not silenced. siRNAs were subsequently discovered in Drosophila tissue culture cells in which RNAi was induced by introducing >500-nucleotide-long exogenous dsRNA (96), in Drosophila embryo extracts that were carrying out RNAi in vitro (240), and also in Drosophila embryos that were injected with dsRNA (236). Thus, the generation of siRNA (21 to 25 nucleotides) turned out to be the signature of any homology-dependent RNA-silencing event.

The siRNAs resemble breakdown products of an E. coli RNase III-like digestion (13). In particular, each strand of siRNA has 5’-phosphate and 3’-hydroxyl termini and 2- to 3-nucleotide 3’ overhangs. Interestingly, in vitro-synthesized siRNAs can, in turn, induce specific RNA degradation when added exogenously to Drosophila cell extracts (69). Specific inhibition of gene expression by these siRNAs has also been observed in many invertebrate and some vertebrate systems (67). Recently, Schwarz et al. (189) provided direct biochemical evidence that the siRNAs could act as guide RNAs for cognate mRNA degradation.

Amplification and Systemic Transmission

Besides the formation of siRNAs, another intriguing characteristic of homology-dependent gene silencing is that the inducer dsRNA molecules do not act stoichiometrically. It was estimated that only two molecules of dsRNA per cell were able to induce RNAi of an abundantly expressed C. elegans gene such as unc22. In another report, injection of dsRNA into the intestine of a C. elegans hermaphrodite generated RNAi, which could be stably inherited to the F2 generation. These two findings led to the proposal that RNAi signals could be systemic and amplifiable in nature (78). The similar systemic effects of RNAi have also been demonstrated in the planarian Schmidtea mediterranea and the cnidarian Hydra magnipapillata (140).

Similar evidence is also available for plant PTGS. The new tissues growing from a GUS-expressing scion grafted onto a GUS-silenced rootstock show progressive silencing of GUS expression (168). The silencing signal seems to spread by a nonmetabolic, gene-specific diffusible signal, which travels both between cells, through plasmadesmata, and long distances via the phloem (75). In the case of virus-induced gene silencing, the systemic character has also been revealed (185). To account for the gene specificity of a systemic signal, it has been proposed that the signal could be an RNA molecule (228). However, such processes are not universal, as these are not found in flies and mammals.

COMPONENTS OF GENE SILENCING

Both genetic and biochemical approaches have been undertaken to understand the basis of silencing. Genetic screens were carried out in the fungus Neurospora crassa, the alga Chlamydomonas reinhardtii, the nematode Caenorhabditis elegans, and the plant A. thaliana to search for mutants defective in quelling, RNA interference, or PTGS. Analyses of these mutants led to the identification of host-encoded proteins involved in gene silencing and also revealed that a number of essential enzymes or factors are common to these processes. Some of the components identified serve as initiators, while others serve as effectors, amplifiers, and transmitters of the gene silencing process. In the years to come, many other components as well as their interrelations will be revealed. Here, we outline what is known so far.

Dicer

RNase III family members are among the few nucleases that show specificity for dsRNAs (164) and cleave them with 3’ overhangs of 2 to 3 nucleotides and 5’-phosphate and 3’-hydroxyl termini (69). Bernstein et al. (17) identified an RNase III-like enzyme in Drosophila extract which was shown to have the ability to produce fragments of 22 nucleotides, similar to the size produced during RNAi. These authors showed that this enzyme is involved in the initiation of RNAi. Owing to its ability to digest dsRNA into uniformly sized small RNAs (siRNA), this enzyme was named Dicer (DCR). These nucleases are evolutionarily conserved in worms, flies, fungi, plants, and mammals. Dicer has four distinct domains: an amino-terminal helicase domain, dual RNase III motifs, a dsRNA binding domain, and a PAZ domain (a 110-amino-acid domain present in proteins like Piwi, Argo, and Zwilie/Pinhead), which it shares with the RDE1/ODE2/Argonaute family of proteins that has been genetically linked to RNAi by independent studies (34, 203). Cleavage by Dicer is thought to be catalyzed by its tandem RNase III domains. Some DCR proteins, including the one from D. melanogaster, contain an ATP-binding motif along with the DEAD box RNA helicase domain.

The predicted C. elegans Dicer homologue, K12H4.8, was referred as DCR1 because it was demonstrated to be the functional ortholog of the Drosophila Dicer protein (173). The 8,165-bp DCR1 protein has a domain structure similar to that of the Drosophila Dicer protein. dcr1 mutants of C. elegans showed defects in RNAi of germ line-expressed genes but no effect on the RNAi response of somatic genes. These mutants were found to be sterile, suggesting the important role of this gene in germ line development apart from RNAi (119). CAF1 has been identified as a Dicer homologue in A. thaliana, but it is not involved in PTGS activity. The structure of CAF1 shows the presence of the four distinct domains that were identified in the Drosophila Dicer protein (17, 36, 108). Dicer homologues from many different sources have been identified; some
recombinant Dicers have also been examined in vitro, and phylogenetic analysis of the known Dicer-like proteins indicates a common ancestry of these proteins (83).

Complete digestion by RNase III enzyme results in dsRNA fragments of 12 to 15 bp, half the size of siRNAs (235). The RNase III enzyme acts as a dimer and thus digests dsRNA with the help of two compound catalytic centers, whereas each monomer of the Dicer enzyme possesses two catalytic domains, with one of them deviating from the consensus catalytic sequences.

Recently, the crystal structure of the RNase III catalytic domain was solved, and this led to the model for generation of 23- to 28-mer diced siRNA products (20). In this model, the dimeric Dicer folds on the dsRNA substrate to produce four compound catalytic sites so that the two terminal sites having the maximum homology with the consensus RNase III catalytic sequence remain active, while the other two internal sites bearing partial homology lose functional significance. Thus, the diced products appear as the limit digests of the RNase III enzymes and are double the size of the normal 12- to 15-mer fragments. Such a model also predicts that certain changes in Dicer structure might modify the spacing between the two active terminal sites and thus generate siRNAs of variable sizes bearing species-specific imprints (98). Clearly, the crystal structure of Dicer is necessary to authenticate this model.

Guide RNAs and RNA-Induced Silencing Complex

Hammond et al. (96) determined that the endogenous genes of Drosophila S2 cells could be targeted in a sequence-specific manner by transfection with dsRNA, and loss-of-function phenotypes were created in cultured Drosophila cells. The inability of cellular extracts treated with a Ca2+-dependent nuclease (micrococcal nuclease, which can degrade both DNA and RNA) to degrade the cognate mRNAs and the absence of this effect with DNase I treatment showed that RNA was an essential component of the nuclease activity. The sequence-specific nuclease activity observed in the cellular extracts responsible for ablating target mRNAs was termed the RNA-induced silencing complex (RISC) (96).

After partial purification of crude extracts through differential centrifugation and anion exchange chromatography, the nuclease cofractionated with a discrete ~25-nucleotide RNA species. These results suggested that small RNAs were associated with sequence-specific nuclease and served as guides to target specific messages based upon sequence recognition. In another report, the multicomponent RNAi nuclease was purified to homogeneity as a ribonucleoprotein complex of ~500 kDa (97). One of the protein components of this complex was identified as a member of the Argonaute family of proteins and was termed Argonaute2 (AGO2). AGO2 is homologous to RDE1, a protein required for dsRNA-mediated gene silencing in C. elegans. AGO2 is a ~130-kDa protein containing polyglutamine residues, PAZ, and PIWI domains characteristic of members of the Argonaute gene family. The Argonaute family members have been linked both to the gene-silencing phenomenon and to the control of development in diverse species. The first link between Argonaute protein and RNAi was shown by isolation of rde1 mutants of C. elegans in a screen for RNAi-deficient mutants. Argonaute family members have been shown to be involved in RNAi in Neurospora crassa (QDE3) as well as in A. thaliana (AGO1) (75).

Recently, two independent groups identified additional components of the RISC complex. Hammond and group showed the presence of two RNA binding proteins, the Vasa intronic gene and dFMR, in the RISC complex isolated from Drosophila flies (35). Of these, dFMR is a homologue of the human fragile X mental retardation protein. In a parallel study, Siomi and group also isolated a novel ribonucleoprotein complex from the Drosophila lysate that contained dFMR1, AGO2, a Drosophila homologue of p68 RNA helicase (Dmp68), and two ribosomal proteins, L5 and L11, along with 5S rRNA (106). Both of these groups showed not only the presence of these components in the RISC complex, but also interactions among these proteins in vitro. Other components of RISC have not been clearly established yet. Nevertheless, some of the proteins mentioned below could very well constitute the RISC complex.

RNA and DNA Helicases

Aberrant RNA elimination surveillance seems to be common to most eukaryotic organisms. However, a diverse array of proteins specific for each organism seem to carry out such surveillance. Broadly, they fall in the biochemically similar group of RNA-DNA helicases. A mutant strain (mut6) of C. reinhardtii was isolated in which a gene required for silencing a transgene was disrupted (232). This RNAi-resistant mutant also showed an elevated transposition activity. The mut6 gene was cloned and sequenced. The deduced MUT6 protein contains 1,431 amino acids and is a member of the DEAH box RNA helicase family. It also has a glycine-rich region that includes several RGG repeats, resembling an RGG box, a motif implicated in RNA binding and protein-protein interactions. MUT6 also has three putative nuclear localization signals and is predicted to be nuclear by PSORT analysis (161). MUT6 RNA helicase may be involved in degradation of misprocessed aberrant RNAs and thus could be a part of an RNAi-related surveillance system.

In Neurospora crassa, three classes of quelling-defective mutants (qde1, qde2, and qde3) have been isolated (46). The qde3 gene has been cloned, and the sequence encodes a 1,955-amino-acid protein (48). The protein shows homology with several polypeptides belonging to the family of RecQ DNA helicases, which includes the human proteins for Bloom’s syndrome and Werner’s syndrome (238). In addition, QDE3 is believed to be involved in the activation step of gene silencing. The DNA helicase activity of QDE3 may function in the DNA-DNA interaction between introduced transgenes or with a putative endogenous gene required for gene-silencing activation by unwinding the double-stranded DNA. These interactions may induce changes in methylation or chromatin structure, producing an altered state that could result in aberrant RNA production. Thus, QDE3 protein may be more important for the transcriptional part of gene silencing, i.e., TGS.

When the RNAi sensitivity of several existing C. elegans mutants was examined, two mutant strains, mut2 and mut7, that had previously shown elevated levels of transposon mobilization also showed resistance to RNAi. Ketting et al. (116) identified a mutator gene, mut7, in C. elegans and character-
ized it at the molecular level. MUT7 was found to be homologous to proteins with 3′-5′ exonuclease domains, such as Werner’s syndrome protein and E. coli RNase D. It contained all the key catalytic residues for nuclease activity. A model was proposed in which MUT7 was speculated to play a role in repressing transposition by degrading the target mRNA with its exonuclease activity.

smg (suppressor of morphological effects on genitalia) mutants of C. elegans, defective in a process called nonsense-mediated decay, have been isolated (63). Seven smg genes, which are involved in nonsense-mediated decay, have been identified (29, 100). Since this process also involves RNA degradation, the function of these genes, if any, in the RNAi process was examined. Animals mutant for a subset of these genes, smg2, smg5, and smg6, were initially silenced by dsRNA but later showed rapid recovery from the effects of RNAi, unlike the wild-type worms, which remained silenced. Thus, these genes might affect the persistence of RNA interference. On the other hand, smg1, smg3, and smg4 mutant animals behaved like wild-type worms and did not recover from RNAi at all, indicating that these genes are not required for RNAi persistence. The smg5 and smg6 genes have not been cloned, but the smg2 gene shows homology to Saccharomyces cerevisiae upf1, which encodes an ATPase with RNA-binding and helicase activities.

The SMG proteins could unwind dsRNA to provide a template for amplification activity. In this way, the three SMG proteins might facilitate amplification of the silencing signal and cause persistence of the silenced state. Alternatively, SMG proteins could increase the number of dsRNA molecules by promoting endonucleolytic cleavage of existing dsRNA molecules, which has been observed in Drosophila flies. No SMG2 homologues have been identified in plants or fungi. However, a search of the A. thaliana genome sequence database revealed a number of candidates with either helicase and/or RNase domains.

In a recent report, Tijsterman et al. (208) showed that unlike sense oligomers, single-stranded oligomers of antisense polarity could induce gene silencing in C. elegans. The antisense RNA-induced gene silencing was explained by proposing that RNA synthesis was primed on the mRNA by antisense RNA, resulting in dsRNAs, which act as substrates for Dicer-dependent degradation. Antisense RNAs showed a requirement for the mutator/RNAi genes mut7 and mut14 but acted independently of the RNAi genes rde1 and rde4 of C. elegans. The mut14 gene was cloned by genetic mapping and subsequent candidate gene approach. The MUT14 protein is a member of the family of putative RNA helicases that contain the signature DEAD box motif. These proteins are involved in diverse cellular functions. The helicase activity of MUT14 might thus act to permit de novo RNA synthesis on the target.

Dalmay et al. (54) identified an sde3 locus in A. thaliana plants which is required for the PTGS phenotype. They proposed that SDE3 protein might be involved in the production of dsRNA. SDE3 differs markedly from QDE3/MUT7 and has slight similarity to MUT6 in the helicase motif. Although it is highly similar to Upf1p and SMG2, it is unlikely that SDE3 is the functional homologue of Upf1p and SMG2 because it lacks important motifs (167). Notably, no SDE3 homologue was found in C. elegans, suggesting that SDE3-like proteins are regulators rather than essential cofactors of PTGS and are not used in C. elegans. This is further supported by the observation that sde3 mutant plants exhibit only partial loss of PTGS (55).

The closest homologue of SDE3 as identified by BlastP was a mouse protein encoded by gb110 (91, 159). These SDE3 homologues have RNA helicase motifs that are quite distinct from those of the DEAD, DEAH, and Ski2p types of RNA helicase (134). It has been speculated that SDE3 and SMG2 are multifunctional RNA helicases involved in PTGS.

Translation Initiation Factor

Mutants of C. elegans showing resistance to dsRNA-mediated RNAi were selected by Tabara et al. (203). They genetically mapped seven mutant strains that were placed in four complementation groups. One of the groups, rde1, consisted of three alleles. Gene rde1 is a member of a large family which includes Drosophila homologues (piwi and sting) and Arabidopsis homologues (argonaute and zwille) and rabbit eIF2C.

Translation Initiation Factor

The full-length cDNA sequence for rde1 was determined, and the deduced protein, consisting of 1,020 amino acids, was referred to as RDE1. The RDE1 protein is homologous to the product of the quelling deficiency (qde2) gene in Neurospora crassa (75). The initiation step of RNAi might be affected in the rde1 mutant, as it completely lacks an interference response to several dsRNAs. It does not show any increase in transposon mobilization and or any effect on growth and development.

RNA-Dependent RNA Polymerase

The effects of both RNAi and PTGS are potent and systemic in nature. This has led to a proposed mechanism in which RNA-dependent RNA polymerases (RdRPs) play a role in both triggering and amplifying the silencing effect. Transgenic and virus-infected plants show an accumulation of aberrant transgenic and viral RNAs. The RdRP enzymes might recognize these aberrant RNAs as templates and synthesize antisense RNAs to form dsRNAs that are necessary for quelling and PTGS, respectively. This testifies to the importance of RdRP in gene silencing. Cogni et al. (45) cloned the qde1 gene from N. crassa. It encodes a 158-kDa protein which lacks the typical signal peptide or a transmembrane domain, indicating its intracellular location. Dalmay et al. (54) found that the 113-kDa Arabidopsis RdRP is encoded by sde1. It is a plant homologue of QDE1 in N. crassa and EGO1 in C. elegans, which are required for quelling and RNAi, respectively. The SDE1 protein is required for transgene silencing but not for virus-induced PTGS, suggesting that SDE1 might be required to produce dsRNA, the initiator of PTGS (54).

The dsRNA produced as an intermediate in virus replication by virus-encoded RdRP might induce PTGS itself, and thus SDE1 may not be required for virus-induced PTGS. Plants with the sde mutation grow and develop normally, excluding a role for sde in development or basic cellular function. Two PTGS-controlling genes, sgs2 and sgs3, were identified in A.
RNA interference Worm (*Caenorhabditis elegans*)

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>Organism</th>
<th>Mutation causing defective silencing</th>
<th>Gene function</th>
<th>Developmental defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttranscriptional gene silencing</td>
<td>Plant (<em>Arabidopsis thaliana</em>)</td>
<td>sgs2/sde1, sgs3, sde3, ago1, caf1</td>
<td>RdRP, Unknown function, RecQ helicase, Translation initiation factor</td>
<td>None</td>
</tr>
<tr>
<td>Quelling</td>
<td>Fungus (<em>Neurospora crassa</em>)</td>
<td>qde-1, qde-2, qde-3</td>
<td>RdRP, Translation initiation factor, RecQ DNA helicase</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mut-7, mut-14, smg-2, smg-5, sid-1</td>
<td>Helicase &amp; RNase D, DEAD box RNA helicase, Upf helicase, Unknown function, Transmembrane protein</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mut-6</td>
<td>DEAH box RNA helicase</td>
<td></td>
</tr>
<tr>
<td>Alga</td>
<td>(<em>Chlamydomonas reinhardtii</em>)</td>
<td>mut-6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*thaliana* by another group of workers (160). Later, it was found that sgs2 and sde1 are different descriptions of the same gene. On comparing the protein sequence of all the RdRPs, a conserved block was identified which seems to be crucial for RdRP function in PTGS and RNAi. sgs3 mutants have the same molecular and phenotypic characteristics as sgs2 mutants, but the SGS3 protein shows no significant similarity with any known putative proteins.

In *C. elegans*, EGO1, a protein required for RNAi, was found to be similar to tomato RdRP and the QDE1 protein of *Neurospora crassa* (197), as mentioned earlier. For a number of germ line-expressed genes, ego1 mutants were resistant to RNA interference. The ego1 transcript is found predominantly in the germ line. ego1 is thus yet another example of a gene encoding an RdRP-related protein with an essential developmental function. RdRP is speculated to play a role in the amplification of the dsRNA signal, allowing its spread throughout the organism (50, 77, 168, 221). The RdRP is also perhaps responsible for sustaining PTGS at the maintenance level even in the absence of the dsRNA that initiates the RNAi effect.

In spite of its omnipresence in different kinds of eukaryotic cells, RdRP homologues are not coded by either the *C. elegans* germ line or human genome. Though the systemic characteristics of RNAi have not been revealed yet in either flies or humans, the amplification of siRNAs may be an essential step of RNAi even in these systems. Hence, it is important to know how these steps of RNAi are biochemically carried out in the absence of RdRP activity.

**Transmembrane Protein (Channel or Receptor)**

The systemic spread of gene silencing from one tissue to another has been well established in *C. elegans* and plants. To investigate the mechanism of systemic RNAi, Winston et al. (231) constructed and used a special transgenic strain of *C. elegans* (HC57). They identified a systemic RNA interference-deficient (*sid*) locus required to transmit the effects of gene silencing between cells with green fluorescent protein (GFP) as a marker protein. Of the 106 *sid* mutants belonging to three complementation groups (*sid1, sid2, and sid3*), they isolated and characterized *sid1* mutants. The *sid1* mutants had no readily detectable mutant phenotype other than failure to show systemic RNAi. Interestingly, these mutants also failed to transmit the effect of RNAi to the progeny.

The SID1 polypeptide is predicted to be a 776-amino-acid membrane protein consisting of a signal peptide and 11 putative transmembrane domains. Based on the structure of SID1, it was suggested that it might act as a channel for the import or export of a systemic RNAi signal or might be necessary for endocytosis of the systemic RNAi signal, perhaps functioning as a receptor. No homologue of *sid1* was detected in *D. melanogaster*, which may be consistent with the apparent lack of systemic RNAi in the organism (80, 174). However, the presence of SID homologues in humans and mice might hint at the systemic characteristics of RNAi in mammals.

**Genetic Mutations with Unknown Function**

The three other complementation groups identified by Tabara et al. (203) in *C. elegans* are *rde2* and *rde3*, with one allele each, and *rde4*, with two alleles. rde4 mutants behaved like the *rde1* strain in not showing any increase in transposon mobilization and no effect on growth and development. The product of *rde2* remains to be identified. *mut2, rde2*, and *rde3* exhibited high-level transposition similar to *mut7*. This suggests a possible biological role of RNAi in transposon silencing (203).

Mello and colleagues (87) have proposed that *rde1* and *rde4* respond to dsRNA by producing a secondary extragenic agent that is used by the downstream genes *rde2* and *mut7* to target...
specific mRNAs for PTGS. According to this view, rde1 and rde4 act as initiators of RNAi whereas rde2 and mut7 are effectors. Various components of gene silencing have been listed in Table 2.

MECHANISM OF RNA INTERFERENCE

As the various pieces of the RNAi machinery are being discovered, the mechanism of RNAi is emerging more clearly. In the last few years, important insights have been gained in elucidating the mechanism of RNAi. A combination of results obtained from several in vivo and in vitro experiments have gelled into a two-step mechanistic model for RNAi/PTGS. The first step, referred to as the RNAi initiating step, involves binding of the RNA nucleases to a large dsRNA and its cleavage into discrete ~21- to ~25-nucleotide RNA fragments (siRNA). In the second step, these siRNAs join a multi-nuclease complex, RISC, which degrades the homologous single-stranded mRNAs. At present, little is known about the RNAi intermediates, RNA-protein complexes, and mechanisms of formation of different complexes during RNAi. In addition to several missing links in the process of RNAi, the molecular basis of its systemic spread is also largely unknown.

Processing of dsRNA into siRNAs

Studies of PTGS in plants provided the first evidence that small RNA molecules are important intermediates of the RNAi process. Hamilton and Baulcombe (92), while studying transgene-induced PTGS in five tomato lines transformed with a tomato 1-aminocyclopropane-1-carboxyl oxidase (ACO), found accumulation of aco small RNAs of 25 nucleotides. More direct evidence about the generation of siRNAs in RNAi came from an in vitro cell-free system obtained from a Drosophila syncytial blastoderm embryo by Tuschi et al. (212). These authors were able to reproduce many of the features of RNAi in this system. When dsRNAs radiolabeled within either the sense or the antisense strand were incubated with S2 cell lysate in a standard RNAi reaction, 21- to 23-nucleotide RNAs were generated with high efficiency. Single-stranded 32P-labeled RNA of either the sense or antisense strand was added to cell-free systems, and these Dicer protein immunoprecipitates were able to produce RNAs of about 22 nucleotides from the dsRNA substrate. The direct correspondence in size of these RNAs with those generated from dsRNA by cell extract suggested a role of this protein in dsRNA degradation. The role of Dicer in RNAi was further confirmed by the fact that the introduction of Dicer dsRNA into Drosophila cells diminished the ability of the transfected cells to carry out RNAi in vitro. Similar experimental studies were carried out with C. elegans extract, and an ortholog of Dicer named DCRI was identified.

A number of in vivo and in vitro experimental studies have shown that the production of 21- to 23-nucleotide RNAs from dsRNA requires ATP. The rate of 21- to 23-nucleotide RNA formation from corresponding dsRNAs has been shown to be six times slower in the Drosophila extract depleted for ATP by treatment with hexokinase and glucose (165). Bernstein et al. (17) and Ketting et al. (115) showed that the Dicer immunoprecipitates from D. melanogaster as well as S2 cell extracts and DCR1 immunoprecipitates from C. elegans extract required ATP for the production of 22-nucleotide RNAs (17, 115). Recently, Nykanen et al. (165) reduced ATP levels in Drosophila extract by 5,000-fold with a sensitive ATP depletion strategy and showed considerable reduction in the rate of siRNA production in the Drosophila cell extract. These experiments suggest that ATP controls the rate of siRNA formation. However, it is still unclear whether ATP is absolutely rate limiting for the production of siRNAs from dsRNA.

The RNase activity and dsRNA binding of 218-kDa recombinant human Dicer have also been examined in vitro (175). The enzyme generated siRNA products from dsRNA quite efficiently in the presence of Mg2+ and the absence of ATP. The RNase activity was sensitive to ionic interactions, whereas the dsRNA binding was quite effective in presence of high salt and did not require Mg2+ at all. The dsRNA binding domain is located at the C terminus of Dicer, which is separable from the helicase and PAZ motifs. Human Dicer expressed in mammalian cells colocalized with calreticulin, a resident protein of
the endoplasmic reticulum. In other systems, Dicer has also been found to complex with various other proteins (35, 106). Hence, it is possible that the Dicer RNase activity functions as a complex of proteins in vivo.

Amplification of siRNAs

One of the many intriguing features of RNA interference is the apparently catalytic nature of the phenomenon. A few molecules of dsRNA are sufficient to degrade a continuously transcribed target mRNA for a long period of time. Although the conversion of long dsRNA into many small siRNAs results in some degree of amplification, it is not sufficient to bring about such continuous mRNA degradation. Since mutations in genes encoding RNA-dependent RNA polymerase (RdRP) affect RNAi, it was proposed that this type of polymerase might replicate siRNAs as epigenetic agents, permitting their spread throughout plants and between generations in C. elegans. Recent studies by Lipardi et al. (135) and Sijen et al. (193) provided convincing biochemical and genetic evidence that RdRP indeed plays a critical role in amplifying RNAi effects.

Lipardi et al. (135), while investigating the dsRNA-dependent degradation of target mRNA in a Drosophila embryo cell extract system, showed the generation of full-length cognate dsRNAs from labeled siRNAs at early time points. Both single-stranded RNAs (equivalent to target mRNA) and dsRNAs served as templates for copying by RdRP. New full-length dsRNAs were formed rapidly and cleaved. They also showed a strict requirement for the 3'-hydroxyl group and 5'-phosphate group on siRNAs for primer extension in the RdRP-mediated reaction (135).

Sijen et al. (193) further revealed the role of RdRP activity in RNAi. In an RNAi reaction, they observed the formation of new siRNAs species corresponding to target mRNAs but different from trigger dsRNAs. They named these new siRNAs secondary siRNAs. With a primary trigger dsRNA specific for the lacZ region of the target mRNA that encoded a GFP-LacZ fusion protein, these authors demonstrated the degradation of a separate GFP mRNA target. This kind of RNAi induced by secondary siRNAs was named transitive RNAi. These authors demonstrated the requirement for the rrf1 gene, a C. elegans gene with sequence homology to RdRP. These models were summarized as RNAs with secondary siRNAs and transitive RNAi (193).

Amplification of siRNAs might occur at various stages of the RNAi reaction and has been documented in plants, C. elegans, N. crassa, and Dictyostelium discoideum but not in flies and mammals (66). Though the RdRP activity is present in Drosophila embryo extract, as mentioned earlier, it is surprising that the fly genome does not code for RdRP. Additionally, numerous experiments also suggest that RdRP is not required for RNAi in D. melanogaster (98).

Degradation of mRNA

In the effector step of RNAi, the double-stranded siRNAs produced in the first step are believed to bind an RNAi-specific protein complex to form a RISC. This complex might undergo activation in the presence of ATP so that the antisense component of the unwound siRNA becomes exposed and allows the RISC to perform the downstream RNAi reaction. Zamore and colleagues (240) demonstrated that a ~250-kDa precursor RISC, found in Drosophila embryo extract, was converted into a ~100-kDa complex upon being activated by ATP. This activated complex cleaved the substrate. The size and constitution of the precursor as well as the activated RISC might vary depending on the choice of system (98). The antisense siRNAs in the activated RISC pair with cognate mRNAs, and the complex cuts this mRNA approximately in the middle of the duplex region.

A few independent studies demonstrated the importance of the RISC complex in this part of RNAi reactions. The mRNA-cleaving RNA-protein complexes have also been referred to as siRNP (small interfering ribonucleoprotein particles). It is widely believed that this nuclease is probably different from Dicer, judging from the substrate requirements and the nature of the end products. Since the target cleavage site has been mapped to 11 or 12 nucleotides downstream of 5' end of the guide siRNA, a conformational rearrangement or a change in the composition of an siRNAP ahead of the cleavage of target mRNA is postulated. Finally, the cleaved mRNAs are perhaps degraded by exoribonucleases (96).

A part of cleaved fragments of mRNA at the end of step 2 might also be converted to the duplex forms by the RdRP-like activity. These forms might have siRNA-like functions and eventually enter the pool of the amplification reaction. Thus, it is likely that amplification of the RNAi reaction takes place at both step 1 and step 2 of RNAi. In another model, it has been proposed that siRNAs do not act as primers for the RdRP-like enzymes, but instead assemble along the length of the target RNA and are then ligated together by an RNA ligase to generate cRNA. The cRNA and target RNA hybrid would then be acted upon by the DCR protein. All these models were summarized by Schwarz et al. (189). Most of the steps involved in the mechanism of RNAi have been illustrated schematically in Fig. 2.

RNA SILENCING FOR GENOME INTEGRITY AND DEFENSE

Considerable evidence indicates that PTGS has evolved as a protective mechanism against parasitic DNA sequences such as transposons and the RNA sequences of plant viruses. DNA methylation and transcriptional gene silencing (TGS) are mainly responsible for keeping the transposition frequency at a minimum. However, PTGS also provides additional protection against the genomic instability caused by transposons. Mutations in the C. elegans mut-7 gene increase the transposition frequency in the germ line and downregulate RNAi as well (58), implicating RNAi in the control of transposons. Recently, Djikeng et al. (61) cloned and sequenced the siRNA products of an RNA interference event occurring in Trypanosoma brucei. By sequencing over 1,300 siRNA-like fragments, they observed abundant 24- to 26-nucleotide fragments homologous to the ubiquitous retrotransposon INGI and the site-specific retropon SLACS. Thus, they convincingly demonstrated that RNAi is involved in silencing the retropon transcript.

In plants, PTGS has been widely linked with RNA virus resistance mechanisms (219, 227). Plant RNA viruses are, in fact, both inducers and targets for PTGS and gene-silencing-
defective mutants of plants show increased sensitivity to viral infections (160). The direct role of dsRNA in inhibiting viral infection has recently been demonstrated by Tenllado and Diaz-Ruiz (207). They showed that dsRNAs derived from viral replicase sequences could interfere with virus infection in a sequence-specific manner by directly delivering the dsRNAs to leaf cells either by mechanical coinoculation with the virus or via an Agrobacterium-mediated transient-expression approach. Successful interference with the infection of plants by representative viruses belonging to the tobamovirus, potyvirus, and alfamovirus genera has been demonstrated. These results support the view that a dsRNA intermediate in virus replication acts as an efficient initiator of PTGS in natural virus infections.

The clinching support for the notion that PTGS has evolved as an antiviral mechanism has come from reports that plant viruses encode proteins that are suppressors of PTGS (8, 25, 222). These suppressors have evolved to save the viral RNA genomes from the PTGS degradative machinery of host plants. Different types of viral suppressors have been identified through the use of a variety of silencing suppression assays. Suppressors HC-PRO, P1, and AC2 are one type (encoded by potyviruses, rice yellow mottle sobemovirus, and geminiviruses of subgroup III, respectively) that is able to activate GFP expression in all tissues of previously silenced GFP-expressing plants (222). HC-PRO reduces target mRNA degradation and is thus responsible for reduced accumulation of siRNAs (137, 145). The second type of suppressors include movement proteins, i.e., p25 of potato virus X, which are involved in curbing the systemic aspect of transgene-induced RNA silencing (220). The third type includes cytomegalovirus 2b protein, which is

---

**FIG. 2.** Two-step model for the mechanism of gene silencing induced by double-stranded RNA. In step I, dsRNA is cleaved by the Dicer enzyme to produce siRNAs. A putative kinase seems to maintain 5’ phosphorylation at this step. The siRNAs have also been proposed to be responsible for nuclear DNA methylation (●) and systemic spread of silencing. Amplification might occur due to the presence of RdRP (▲). In step II, the siRNAs generated in step I bind to the nuclease complex (RISC). A helicase present in the complex might activate RISC by unwinding the siRNAs. The antisense component of siRNA in the RISC guides the complex towards the cognate mRNA (—), resulting in endonucleolytic cleavage (↓) of the mRNA. RdDM, RNA-dependent DNA methylation.
involved in systemic signal-mediated RNA silencing (60). The cytomegalovirus 2b protein is nucleus localized and also inhibits salicylic acid-mediated virus resistance (141). Other types of viral suppressors with undefined biochemical activities are also known (128). These findings not only provide the strongest support that PTGS functions as a natural, antiviral defense mechanism, but also offer valuable tools for dissecting the biochemical pathways of PTGS (128).

The PTGS degradative machinery can both detect and inactivate repetitive DNA sequences, suggesting it controls the expansion of repetitive elements, including endogenous genes (18). Although RNAi occurs in mammals and mammalian cell cultures, its role in animal virus protection is not clear. In mammals, dsRNA induces RNAi as well as interferon-mediated nonspecific RNA degradation and other nonspecific responses leading to blockage in protein synthesis and cell death (2). Thus, mammals seem to have evolved multiple mechanisms to detect and target dsRNA and to fight viruses. These various mechanisms may have different specificities or can function in distinct tissues or during development (210). A few other roles of RNAi in development and genome maintenance will be discussed in later sections.

**MECHANISTIC DIFFERENCES AMONG THE BIOSYNTHETIC PATHWAYS OF siRNA**

Although the functional parallelism of gene silencing is quite apparent in plants and animals, a few unique attributes separate the pathways in these groups. For example, systemic spreading of the RNAi reaction from the site of initiation is known to occur in plants and worms (74, 79), but not in flies or mammals. The noteworthy distinct molecules that have been identified to cause differences at the pre-Dicer, Dicer, and post-Dicer stages of gene silencing pathways are mentioned below.

**Pre-Dicer stage.** Plant proteins such as SGS2 (RdRP), SGS3 (coiled protein), AGO1 (responsible for plant development), and HEN1 (enhancer of floral hua1 mutation) are required for
PTGS activities induced by the sense transgenes. But if the transgenes are in the form of hairpins expressing the panhandle dsRNA, the absence of or defects in the above-mentioned proteins do not play any role in altering the PTGS/cosuppression function. Hence, those proteins supposedly play a role upstream of dicing of dsRNA and may be involved in the formation and stabilization of dsRNA (22).

Homologues of SGS3 are unknown beyond the plant world. Even though HEN1 analogs are known in bacteria, yeasts, and animals, their roles in sense PTGS have not yet been identified. Likewise, SGS2 homologues are known in C. elegans, N. crassa, and Dictyostelium discoideum, but their roles at the pre-dicing stage have not been established yet in those systems. The equivalents of SGS2 in other animal systems are nonexistent both structurally and functionally (205). The role of worm AGO1 protein, i.e., RDE1, is also unique, as described earlier. AGO1 homologues are present in all eukaryotes, but they mostly function as a component (AGO2) of the animal RISC complex (32). The plant HEN1 protein is believed to be nuclear because of the presence of the nuclear localization signal at its N-terminal region (40). Since HEN1 is essential for plant PTGS (cosuppression), which is supposedly a cytoplasmic activity, the exploration of its subcellular distribution is of utmost importance. Boutet et al. (22) speculated that HEN1 could be a dsRNA stabilizing protein, and since many such proteins are known in the animal kingdom, it would be of interest to find animal analogs of the plant HEN1 protein. In fungus as well as the animal system, sense transgene-induced PTGS phenomena are known, but the machinery operative at the pre-dicing stage is still elusive.

The roles of plant SGS2, SGS3, AGO1, and HEN1 proteins may be limited at the stage of production of dsRNA from the transcript of sense transgenes, but no mechanism has been established regarding the presentation of the dsRNA to Dicer for the generation of siRNA. However, such a mechanism has been reported in C. elegans. The RDE4 and RDE1 (AGO1)
proteins of *C. elegans* were reported as initiators of RNAi and speculated to have no mechanistic role in the downstream processes of RNAi (87, 203). Unlike the *Arabidopsis* AGO1 and HEN1 proteins, RDE4 and RDE1 proteins are required for RNAi even when the dsRNAs are produced intracellularly in transgenic worms (203), but the defects in RDE4 and RDE1 are of no consequence if exogenous siRNAs or short antisense RNAs drive the RNAi reaction (208). RDE4 binds tightly to dsRNA (during the RNAi reaction) by virtue of its two RNA-binding domains and is always found in a tight complex with RDE1 protein even in absence of the RNAi reaction. During RNAi, RDE4 is found in a complex with RDE1, Dicer (DCR1), and a conserved DEXH-box RNA helicase (DRH1/DRH2). Based on these observations and other genetic evidence, Tabara and coworkers postulated that RDE4 and RDE1 functioned together to detect and retain foreign dsRNA and present the dsRNA to DCR1 for processing into siRNAs (202). Analogs of the RDE4 and DRH proteins are found in many eukaryotes, including plants and humans, but their roles have not been defined yet.

**Dicer stage.** The plant Dicer responsible for biosynthesis of plant siRNA is not known yet, whereas the Dicers of *C. elegans*, *D. melanogaster*, and humans as effectors for siRNA have been well characterized. The *A. thaliana* and rice genomes both encode at least seven RNase III-like proteins, of which at least four are putative homologues of Dicer, conveniently called DCLs (i.e., DCL1, DCL2, DCL3, and DCL4). The genetic evidence rules out that the *Arabidopsis* DCL1 (or CAF1) could be competent for siRNA formation (76). The roles of other DCL proteins are still to be revealed.

Interestingly, both in vivo and in vitro data suggest that the end products of plant dicing activities are different from those of the animal Dicers. When uniformly-32P-labeled dsRNA was incubated with wheat germ extract, Zamore et al. (205) found that the dsRNA was chopped into siRNAs of two discrete size classes, one ~21 nucleotides and the other 24 to 25 nucleotides long, whereas *D. melanogaster* and human Dicers generated only the 21-nucleotide siRNAs. Two similar size classes were also produced with cauliflower extract and were found independently in the set of 423 endogenous small RNAs cloned from *A. thaliana*. Thus, in plants, dicing activity leads to the generation of two distinct classes of siRNAs.

With specific synthetic siRNAs that supposedly bind tightly to and inhibit Dicer as competitors, Zamore et al. concluded that a different Dicer-like enzyme was responsible for the generation of each class of siRNA. These two distinct classes of siRNAs were reported first in vivo from transgenic plants bearing the silenced GFP sense transgenes (94). With an array of plant virus-encoded suppressors of gene silencing, Baulcombe et al. proposed that the 21-mer siRNAs controlled localized PTGS via mRNA degradation and the 24-mer siRNAs triggered systemic silencing and methylation of the homologous DNA. It remains to be seen whether this kind of dual dicing activity reflects any novel pathway intrinsic to plant RNAi. Interestingly, the longer (~25-mer) siRNAs have also been detected in the natural RNAi biology of *Trypanosoma brucei* (61).

**Post-Dicer stage.** RISC has been isolated from *D. melanogaster*, *C. elegans*, and humans, and only some of its components have been characterized biochemically and genetically. Both mammalian and *Drosophila* RISC contain AGO2 proteins, whereas the GEMIN3 (a DEAD box helicase) and GEMIN4 proteins are found only in mammalian RISC (103). Similarly, dFXR, a homologue of the human fragile X mental retardation protein, is found only in *Drosophila* RISC (35). However, there is no report on the isolation of RISC in plants. Hence, mechanistically little is known about postdicing activity, especially in plants. A worthwhile question to address is whether there is any anchoring site for the occurrence of RNAi in the cytoplasm. Recently, it was reported that *E. coli* RNase III binds to the 70S ribosome and is functionally modified after binding (6). It is widely believed that the RISC associates with eukaryotic ribosomes (96). Hence, the exploration of ribosome association of the RNAi activities, especially of dicing and postdicing leading to mRNA degradation, might shed light on RNAi mechanisms in the future. The various affinities of ribosome-binding complexes might also reveal interesting system-specific features.

RdRP-dependent siRNA amplification and systemic spreading from the site of origin is another area where many system-specific variations have been noticed. RdRP homologues are not present in many organisms, so the mechanisms by which sense transgene-mediated PTGS are effected in those organisms remain a mystery (98). In other systems where RdRP is present, the biochemical steps and details of siRNA amplification may not necessarily be the same.

In *C. elegans*, RRF1 (a putative homologue of RdRP), along with other proteins, is required for RNAi even when the trigger dsRNA is expressed directly from the hairpin transgene in the nuclei of somatic tissues, whereas SGS2 (*Arabidopsis* RdRP) is dispensable for PTGS activity if induced directly by hairpin sense transgenes in *A. thaliana*. This suggests that the RdRP-mediated putative amplification steps of worms are different from those of plants (37). In plants, the SGS2-dependent spreading of silencing occurs from the region homologous to the trigger dsRNA into both the adjacent nonhomologous 5' and 3' regions of a target transgene (214). In contrast, spreading occurs only in the 5' region in worms and fungi, which is consistent with the primer-dependent 5'-3' copying activity of RdRP. Hence, in plants, the spread of silencing requires other activities (such as chromatin modification) in addition to that of RdRP (37).

In worms, tissue-specific variations of RdRP-dependent RNAi have also been reported, but not in plants or other systems. EGO1 is essential for RNAi in the germ line of *C. elegans*, whereas another RdRP homologue, RRF1, is required for silencing in soma (193, 197). Another intriguing observation is that the loss of function of RRF3 (third putative RdRP of worms) is responsible for the enhancement of sensitivity to RNAi in several tissues of *C. elegans*. Here, RRF3 acts as a negative regulator of RNAi, a fact difficult to reconcile with the postulated activity of RdRP (195). For systemic transmission of gene silencing, the membrane-bound SID1 protein of *C. elegans* and the plasmodesmal connections of plants are implicated, but in both cases, the molecular nature of the moving signal has not been ascertained yet. An association between Dicer and the RdRP has been suspected in the case of *Dictostelium discoideum* and *C. elegans*, but conclusive evidence is still lacking (37).
siRNA: SYNTHESIS, DELIVERY, AND GENE KNOCKDOWN

The natural RNAi biology of eukaryotic cells offers a protection mechanism against foreign nucleic acids; however, only in the recent past has the exploitation of its mechanistic details sparked a revolution in the investigation of cellular gene functions. Transcriptional regulation with the dsRNA technology provides an easy means to identify cellular characteristics in response to both internal and external cues. However, the application of RNAi in higher eukaryotes, particularly mammalian cells, has been hampered by the presence of a number of dsRNA-triggered pathways that mediate nonspecific suppression of gene expression (152). These nonspecific responses to dsRNA are not triggered by dsRNAs shorter than 30 bp, including the siRNA duplexes. Moreover, studies in C. elegans and D. melanogaster have clearly demonstrated that synthetic siRNAs can produce effects similar to those of the long dsRNAs (69, 236). Based on these experimental analyses, siRNAs are now being optimized for systematic exploration of the function of genes in a variety of organisms.

Prior to the siRNA era, approaches such as gene targeting by homologous recombination, ribozymes, and antisense technologies were commonly used to determine gene functions. All such approaches have their limitations, and none can be applied universally (201). The dawn of siRNA-directed knockdown approaches facilitated studies of gene function in a rapid and inexpensive way. This siRNA technology has the potential to decipher the function of virtually any gene that is expressed in a cell type- or pathway-specific manner. In the span of only a few years, large-scale functional analysis of almost all the ~19,000 genes of C. elegans has been carried out with the siRNA-directed knockdown approach. A fairly detailed account of this technology has recently been reviewed by Dykxhoorn et al. (66).

Here, some of the salient aspects of the technology are summarized. In brief, the application of siRNA for gene silencing involves a careful consideration of the following variables: (i) selecting the siRNA sequence in the target gene; (ii) synthesis of siRNAs or construction of plasmids bearing DNA sequence encoding for siRNAs; (iii) optimizing transfection of the siRNAs or the plasmids expressing siRNAs in the target cells; and (iv) monitoring the efficacy of gene silencing.

Selection and Generation of siRNA

Several siRNAs synthesized against different regions of the same target mRNA show different silencing efficiencies (101). A number of groups have analyzed several parameters for optimizing siRNA-induced gene silencing, and these include the length, secondary structure, sugar backbone, and sequence specificity of the siRNA duplex. The efficacy of these parameters has been tested on several occasions for induction of RNAi in D. melanogaster and human cells (69, 189). No consensus on choosing the siRNA sequence has evolved. A line of thinking seems to suggest the following. The sequence should be selected in the region 50 to 100 bp downstream of the start codon. The 5' or 3' untranslated regions and regions near the start codon should be avoided, assuming that untranslated region-binding proteins and translation initiation complexes may interfere with the binding of siRNP or RISC endonuclease complex. The GC content of the siRNAs should be kept between 30 and 70%. The computer programs developed by Lin (Jack Lin's siRNA sequence finder; www.1c.sunysb.edu/stu/shi/kln/rna1.html) and by Ambion (www.ambion.com) offer helpful guidelines to select potential siRNA sequences and determine whether these selected sequences match mRNA sequences other than those of intended target.

Based on different experimental approaches, a few guidelines have been laid for the synthesis of siRNAs. A general rule is that the sequence of one strand should be AA(N19)TT, where N is any nucleotide, i.e., these siRNAs should have a 2-nucleotide 3' overhang of uridine residues. The siRNAs should be 21 nucleotides long. The siRNAs should have 5'-phosphate and 3'-hydroxyl group for efficiency. Compared to antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. The 21-nucleotide siRNAs can be chemically synthesized with appropriately protected ribonucleoside phosphoramidites and a conventional synthesizer and thus are widely available commercially. However, the use of chemically synthesized siRNA in RNAi has been restricted because of the high synthesis cost. Due to the paucity of information on the selection of siRNAs and their structures, these general guidelines are suggestive and do not guarantee the silencing effect. To overcome the siRNA selection ambiguity, Yang et al. (235) incubated dsRNA with the E. coli RNase III enzyme to generate a random array of siRNAs. The introduction of such a reaction soup resulted in the silencing of the target gene.

The exorbitant cost of synthesizing siRNAs and their lack of amplification in mammalian cells have compelled investigators to explore alternative strategies to generate a continuous supply of a battery of siRNAs. Several groups have devised strategies to synthesize short RNAs in vitro (64) or by introducing plasmids with the ability to make de novo siRNAs inside the cell (235, 239). DNA-based plasmid vectors have been designed by cloning siRNA templates downstream of an RNA polymerase III transcription unit, which normally encodes the small nuclear RNA U6 or human RNome H1.

Two approaches have been developed for expressing siRNAs. In the first, sense and antisense strands constituting the siRNA duplex are transcribed by individual promoters (64), and in the second, siRNAs are expressed as fold-back stem-loop structures that give rise to siRNAs with a small loop. A stretch of four to five thymines is added at the end to the siRNA template that acts as a transcription termination signal. Many of these plasmid-based vectors, such as pSilencer 1.0 (Ambion) and pSuper (DNA Engine), are now commercially available. These vectors provide advantages over chemically synthesized siRNAs, but use of these plasmid vectors also remains limited due to numerous disadvantages, including the transient nature of siRNA expression and low as well as variable transfection efficiency.

To circumvent these problems, virus-based high-efficiency siRNA delivery systems are also being developed. A retrovirus-based system developed by Devree and Silver (59) is cited here as an example. The U6 promoter along with the siRNA-generating hairpin construct was cloned upstream of the 3' long terminal repeat of the commercially available pMSCV-puro vector. The in vitro-packaged recombinant virus was allowed to
transfect HeLa cells with high efficiency in the presence of puromycin selection, and a dramatic downregulation of the target gene product was observed. A downregulation of this extent was not possible with the plasmid-based delivery system. Such virus-based vectors or their improved variants hold the promise to efficiently detect the function of any gene in virtually any cell type, provided that the production of recombinant virus is not a limitation.

Most of the siRNA expression vectors produced to date use RNA polymerase III regulatory units, which do not allow tissue-specific siRNA expression. However, Shiagawa and Ishii reported a polymerase II promoter-based plasmid encoding a dsRNA expression system that could eventually express siRNA in a tissue-specific manner (192). In their novel scheme, a pDECAP vector was used, which expressed long dsRNAs corresponding to the ski gene (encoding a transcriptional repressor) in the form of a hairpin. The engineered hairpin RNA expressed from a cytomegalovirus promoter lacked the 7-methylguanosine cap structure at its 5’ end and a poly(A) tail at its 3’ end. The transcript of such a design did not exit the nucleus to reach the cytoplasm and thus prevent the interferon pathway-mediated nonspecific antiviral response. The double-stranded transcript was diced in the nucleus, and the siRNAs were subsequently released into the cytoplasm to mediate the gene-specific silencing. The silencing was specific, since the level of a related protein, SNO, remained unaffected.

The same vector was also used to create ski knockdown mice, the phenotype of which was similar to that of ski knockout embryos, which exhibited defects in neural tube and eye formation. Later generations of such vectors may use more tissue-specific cis-acting elements in the employed promoter to stringently knock down gene functions in the animal system. It is pertinent to highlight here that because plants do not elicit an interferon-mediated antiviral response, the dsRNA/siRNA delivery system need not be as complex as the pDECAP system.

Transfection of siRNA and Detection of Gene Silencing

An attempt to understand a gene’s function in diverse organisms necessitates optimization of protocols for efficient delivery of siRNAs into cells. A number of transfection reagents are being employed for transfecting siRNA into different cell lines. Lipofectamine 2000 and Oligofectamine (Invitrogen) are being routinely used for siRNA delivery. A few newer transfection reagents such as TransIT-TKO (Mirus) and Ambion’s Sport Amine and SiPort, have also been used successfully in cultured cell lines. Electroporation has been used to transfect siRNAs in cell lines as well as in parasites such as Trypanosoma brucei and Plasmodium falciparum (150, 213). In adult mice, naked siRNAs have been delivered by hydrodynamic transfection methods to combat hepatitis C virus infection in the intact liver (151). The transfecting siRNAs have been used successfully for studying the role of proteins in DNA damage response and cell cycle control, general cell cycle metabolism, signaling, the cytoskeleton and its rearrangement during mitosis, membrane trafficking, transcription, and DNA methylation (211). These molecules have also been used to differentiate between housekeeping and other genes (112).

The preferred way to detect specific gene knockdown by RNAi is to study the depletion of the target protein by immunofluorescence and Western blotting with the specific antibody. In addition, the knockdown phenotype and Northern blot analysis can also be used to detect the effects of siRNA. If the gene is essential, cellular growth is delayed or arrested, and [3H]thymidine uptake can also be used to assign the function of a particular gene (70).

siRNA Introduction into Plants

siRNAs have been delivered into tobacco plants by biolistic pressure to cause silencing of GFP expression. Silencing occasionally was detected as early as a day after bombardment, and it continued to potentiate up to 3 to 4 days postbombardment. Systemic spread of silencing occurred 2 weeks later to manifest in the vascular tissues of the nonbombarded leaves that were closest to the bombarded ones. After a month or so, the loss of GFP expression was seen in nonvascular tissues as well. RNA blot hybridization with systemic leaves indicated that the biolistically delivered siRNAs induced the de novo formation of siRNAs, which accumulated to cause systemic silencing (118).

MICRO-RNA

Since the RNAi machinery is present constitutively within eukaryotic cells, it is important to explore and understand the metabolic advantages that are accorded by RNAi-related proteins during the intrinsic normal growth of cells and development of organisms. The natural RNAi machinery not only keeps the mobile transposable elements from disrupting the integrity of genomes, as was suggested by analyses in lower plants, A. thaliana, C. elegans, D. melanogaster, and animals (9, 94, 138, 203, 232), but also participates in organism development. Genetic defects in C. elegans RNAi genes ego1 and dicer cause known, specific developmental errors (87, 119, 197). Similarly, the Argonaute family of genes of A. thaliana (especially the ZWILLE proteins) is also responsible for plant architecture and meristem development (32), and the Dicer homologue of A. thaliana, CAF1, is required for embryo development (83). Thus, genetic evidence illustrates the role of the RNAi machinery as a controller of development-related genes. The mechanistic details of these developmental processes are beginning to emerge.

In 1991, Ambros and coworkers first isolated a lin4 mutant of C. elegans which was arrested at the first larval stage (127). Later on, the let7 mutation was isolated in the same system, which was responsible for development through the fourth larval stage. Both lin4 and let7 encode short 22-nucleotide mature RNAs and were called short temporal RNA because they control the temporal development program of C. elegans. The mature lin4 RNA defines (negatively regulates) the mRNA expression of the lin14 and lin28 heterochronic genes with the antisense-mediated repression mechanism of translation initiation and thus specifies the fate of cells during the first three larval stages. Recent studies have revealed that the short temporal RNAs are actually members of a group of tiny RNAs (21 to 28 nucleotides) called the micro-RNAs, isolated members of which could easily run to a few hundreds. Some of the components of the RNAi machinery have also been clearly

Downloaded from http://mmbr.asm.org/ on January 22, 2021 by guest
established as the effector proteins for the maturation of micro-RNAs.

Identification and Biogenesis

A range of biochemical techniques have been applied to clone the 21- to 28-nucleotide RNAs that are present during the normal cellular development of many organisms, for exploring the abundance and complexity of micro-RNA. Micro-RNAs have been found to be abundant and phylogenetically extensive in plants, flies, worms, and humans. In D. melanogaster, C. elegans, plants, and humans, more than 600 micro-RNAs have been identified (123, 125, 126, 137). Bioinformatic analyses of the complete genome sequences have been extremely useful for identification studies. The genome sequences of a variety of organisms revealed the authenticity of these micro-RNAs, the nature of the precursor RNAs, the genomic locations of micro-RNA genes, and the evolutionarily conserved character of some of these micro-RNAs. With the RNA folding program mfold (148) and Northern analyses of micro-RNA, it has been universally inferred that most micro-RNAs arise from the imperfectly annealed 70-nucleotide hairpin precursor RNA whose expression is often developmentally regulated. These micro-RNAs are thus predicted to be processed from multiple bulged and partially duplex precursors, like the short temporal RNA precursors (186).

The identification of micro-RNAs is the first major hurdle in micro-RNA-related research. The first step in computational identification of micro-RNAs from genome sequences is identification of sequences forming hairpin loops (stem-loop sequences). For this purpose, software such as smloop (85) and RNA fold (130) is used. These are Blast-like software packages which identify short complementary sequences within a specified distance on the genome. The hairpin sequences obtained by this analysis are then evaluated as candidate micro-RNAs based on different criteria, such as GC content and minimum free energy, and by passing through different filters, such as short-repeat filters and structure quality filters (85).

Another important criterion that has been used for the identification of candidate micro-RNAs is the correspondence of a hairpin of one species with that of another species. Two hairpins are said to be in correspondence if a short sequence (>19 nucleotides) in the stem of one hairpin is also present in the stem of another hairpin, although the two hairpins may have otherwise variable sequences. If a hairpin from one species has correspondence with a hairpin from one or more other species, this strengthens its status as a candidate to be a micro-RNA. The homology of hairpins with known micro-RNAs is also considered a useful criterion to select candidate micro-RNAs.

D. P. Bartel’s laboratory has developed a computational procedure, called MiRscan, to identify micro-RNAs based on their homology to known micro-RNAs with respect to the characteristic features in the stem region. MiRscan evaluates the stem-loops by passing a 21-nucleotide window along the stem region and assigning a likelihood score to each window that measures how well its attributes resemble those of the previously experimentally identified and validated micro-RNAs (129, 130). The candidate micro-RNAs identified by these procedures are experimentally validated by Northern blot assay of total small RNAs with the stem region of the candidate as a probe or by a more sensitive PCR assay of the amplified small RNA library (85, 130). Detection of a 21- to 24-nucleotide band in these assays validates a candidate micro-RNA, whereas a ~70-nucleotide band is detected in Dicer-deficient mutants, further confirming that the micro-RNAs arise from a ~70-nucleotide precursor.

An analysis of micro-RNA expression in cell lines and tissues suggests cell- or tissue-specific expression. For example, micro-RNA 1 (miR1) is specifically expressed in human heart tissues and stage-specifically in mouse embryogenesis (126). A. thaliana small RNA 39 is detected exclusively in inflorescence tissues and downregulates the expression of a Scarecrow-like transcription factors (137, 139). Considering the diverse functions in which micro-RNAs have been implicated, micro-RNAs have also been named variously, i.e., micro-RNAs which mediate spatial development are referred to as sdRNAs, while cell cycle micro-RNAs are referred to as cRNAs, etc. The regulated expression patterns of these micro-RNAs are suggestive of their functions in developmental control. However, many micro-RNAs are uniformly expressed, suggesting their role in general gene regulation (186). Downregulation of micro-RNAs leads to serious developmental defects, as evidenced by isolation of various micro-RNA mutants. Recent reports reveal that mir15 and mir16 are located in human chromosome 13q14, a region which gets deleted in more than half of B-cell chronic lymphocytic leukemias. Detailed deletion and expression analyses point out that these two micro-RNAs are located within a 30-kb region of loss in chronic lymphocytic leukemias, and both genes are deleted or downregulated in a majority (~68%) of chronic lymphocytic leukemia cases (30).

A majority of micro-RNAs occur in relatively short (~70-nucleotide) and single stem-loop precursor structures. However, in both animals and plants, some micro-RNAs are arranged in clusters. The genes in the tandem clusters are coexpressed, for example, in the germ line and early embryos of C. elegans and D. melanogaster (123, 125) and in the inflorescence tissues of A. thaliana (137). A set of seven highly related C. elegans micro-RNA genes that are coexpressed are so tightly clustered within 1-kb region that they are predicted to form a precursor from which all the seven mature micro-RNAs are processed (186). Similarly, several micro-RNAs originate from each of the five chromosomes of A. thaliana containing clusters of two to four micro-RNAs spaced irregularly within the intergenic region. Interestingly, three of the clusters contain micro-RNA sequences of both sense and antisense polarities, a scenario not found in the animal system yet. Such variations in the precursor structures of micro-RNAs may point towards distinct mechanisms of biosynthesis of micro-RNAs, although all micro-RNAs originate by transcription events that are independent of adjacent conventional genes.

Apoptosis-Related Micro-RNA

The proliferation of tissues and organs of any organism requires careful coordination between cell proliferation and cellular death. The proliferation processes of a cell include active inhibition of the apoptotic process. Recently, two micro-RNA genes, bantam and mir14, that suppress cell death by inhibiting the translation of apoptotic messages have been isolated from D. melanogaster. Expression of the bantam 21-nu-
cleotide micro-RNA is temporally and spatially regulated in response to patterning cues. The proapoptosis gene hid has been identified as a target for regulation by bantam micro-RNA (24). bantam deletion mutants grow poorly and die as early pupae, whereas mir14 mutants are viable but stress sensitive and cursed with a reduced life span. The mir14 suppresses death induced by expression of Rpr, Hid, Grim, or the mir14 early pupae, whereas bantam RNA (24).

The proapoptosis gene cleotide micro-RNA is temporally and spatially regulated in animals as well as in plants (86, 96, 104, 115, 240). Human recombinant Dicer can process pre-let7 RNA to mature let7 quite efficiently in vitro (175). Recent work by D. P. Bartel's group (181) has also shown that caf1 (dicer homologue) mutants of A. thaliana fail to process micro-RNAs. The genetic and biochemical data point toward interaction between Dicer and the Argonaute group of proteins in C. elegans and D. melanogaster for processing the micro-RNAs (86, 95). The similar interaction is possibly also present in plants between Dicer on one hand and PNH (zwille/pinhead) on the other to generate plant micro-RNAs (83). Additionally, both forms of small RNA, micro-RNAs and siRNAs, were found integrally associated with riboprotein complexes containing a member of the PIWI/PAZ domain family, siRNAs in the RISC and micro-RNAs in the microribonucleoprotein complexes (96).

In fact, recent evidence suggests that at least for some micro-RNAs, the microribonucleoprotein and the RISC complex could be the same entity (103, 137, 139, 182). Though the same or similar DCR and subsequent ribonucleocores are required to process mature forms of the micro-RNAs, in some cases, such as C. elegans lin4 and let7, the ~22-nucleotide form is processed from the 5' part of the stem, and in other cases, such as miR1 and miR58, maturation results from the 3' part of the precursors. Thus, there is a gene specificity of micro-RNA processing and/or stabilization (126). Since the biosynthetic pathways of micro-RNAs and siRNAs are somewhat similar, the viral suppressors that inhibit siRNA formation are also expected to interfere in the biogenesis of micro-RNAs. A detailed understanding of this suppression process may unravel the hitherto unknown molecular basis of virus-induced development-related diseases in eukaryotes, especially in plants.

An RNA-silencing suppressor, PI/HC-PRO of turnip mosaic virus, induces a number of developmental defects in the vegetative and reproductive organs of A. thaliana. Many of these defects are reminiscent of observed defects in Dicer-like mutants of A. thaliana. The PI/HC-PRO suppressor interferes with the formation of miR171, and as a result the downstream target mRNAs accumulate instead of being cleaved, causing developmental errors (113). Thus, it is interesting that the counterdefensive strategy of the viruses has evolved not only to protect the viral RNA genome from the host degradative machinery but also to subvert the cellular development program in favor of the virus.

However, it is important to mark the distinctions among the pathways leading to the formation as well as the activities of siRNAs and micro-RNAs. Although over 600 micro-RNAs from various organisms have been identified (33), only about 3% of them are fully complementary to the target mRNA sequences. All known micro-RNAs are derived in vivo from dsRNA precursors which are imperfectly annealed. Since the biosynthesis and activities of the micro-RNAs do not require perfect complementarity, noncanonical pathways of RNAi may be involved for the micro-RNAs because the usual RNAi calls for extensive complementarity of the dsRNA. It is only because of this characteristic mismatch between the sequences of micro-RNA and cognate mRNA that the in silico identification of the target mRNA is so difficult (182). The imperfect nature of annealing between the two partners is viewed as the prime cause for translational repression of the target mRNA (172).

Second, the mature micro-RNAs are always found in the single-stranded conformation in nature for some unknown reason, whereas siRNAs are double-stranded when detected. Third, unlike siRNAs, micro-RNAs enter riboprotein complexes with differing PPD proteins (PAZ and Piwi domains), depending on the specificity of the micro-RNA or its precursor with the cognate PPD proteins (86). The sequence or structure of a micro-RNA or its precursor might ensure that it functions as a translational repressor and not as a trigger of RNAi. It is widely speculated that the siRNAs and micro-RNAs are distinguished following their biosyntheses, and these two are then allowed to form related but distinct ribonucleoprotein complexes that target downstream substrates for degradation or translation repression, respectively. This hypothesis is based on the observation that siRNAs or exogenously supplied hairpin RNAs containing even a single mismatch with their substrate fail to repress the target mRNAs and do not simply shift their regulatory mode to translation inhibition (98).

Fourth, a viral suppressor of RNA silencing, the HC-PRO protein of potato virus Y, has been found to differentially regulate the accumulation of siRNAs and micro-RNAs in tobacco (144). The HC-PRO protein prevents accumulation of siRNAs of the silenced genes and thus releases silencing in a universal manner, but the same protein helps accumulation of all micro-RNAs tested, namely, mir167, mir164, and mir156 of tobacco, in vivo. This result indicates that the dicing complexes for siRNA and micro-RNA may not be exactly similar in biochemical features, and as a result the biochemical functions of the complexes are different in response to this particular HC-PRO protein. Lastly, not all RNAi pathway mutants are developmentally aberrant, whereas micro-RNA pathway mutants are expected to be defective in organism architecture and development. For example, an RNA-dependent RNA polymerase-defective mutant of A. thaliana, the sgs2/sde1 mutant, shows defects only in cosuppression phenomena (a form of RNAi) but is perfectly normal in phenotypic development (54, 160). This observation raises the question of whether the...
RdRP-dependent amplification step is required at all for micro-RNAs.

**Functional Classifications**

A number of micro-RNAs, including let7, are conserved across all organisms throughout evolution. About 12% of the micro-RNAs identified so far in animal systems are conserved at least among nematodes, *D. melanogaster*, and humans. Interestingly, a majority of the micro-RNAs are speculated to control development-related genes. However, the mechanisms of such control are not quite established yet. Micro-RNAs probably employ a variety of mechanisms to downregulate target genes. Micro-RNAs such as *C. elegans lin4* and *let7* have been shown to imperfectly anneal to the 3’ untranslated region of the target mRNA.

A vast majority of micro-RNAs probably belongs to this category. Due to imperfect complementarity, some micro-RNAs may also anneal to a host of different target mRNAs either simultaneously or in a temporally controlled manner. On the other hand, there are micro-RNAs, located mostly in the *A. thaliana* intergenic region, which have perfect or nearly perfect complementarity to the target mRNAs. Such micro-RNAs might trigger site-specific cleavage of the mRNA after being incorporated into a functional RISC-like complex. In such a situation, micro-RNAs act like siRNAs. The *A. thaliana* inflorescence-specific small RNA 39 cleaves the middle part of mRNA of the three scarecrow gene family members in a similar fashion (137).

On the basis of nearly perfect complementarity with the micro-RNAs, numerous *A. thaliana* mRNA targets have been predicted, and these targets have also been phylogenetically conserved in rice. Fifteen cleavage-type targets were validated recently by in vitro or in vivo micro-RNA-guided cleavage assays. The majority of these predicted mRNA targets encode members of large family of transcription factors, including Phavlota (PHV), Phabulosa (PHB), cup-shaped cotyledon 1 (CUC1), CUC2, etc. These transcription factors are required for meristem identity, cell division, organ separation, and organ polarity (33). On the other hand, *mir172* likely acts in cell fate specifications as a translational repressor of APETAL2 in *Arabidopsis* flower development (39).

There are other varieties of micro-RNA which also interact with target mRNAs affecting the posttranscriptional steps, such as RNA splicing, mRNA localization, and RNA turnover. In *D. melanogaster*, many micro-RNAs are known to be complementary to the 3’ untranslated region sequence motifs, which are responsible for mediating negative posttranscriptional regulation. These sequence motifs include the K box (CUGGAUA), the B-rd box (AGCUUUA), and the recently found GY box (UGUCUUC). All micro-RNAs showing complementarity to these motifs are expressed either broadly throughout development or in the narrow window of embryogenesis of *D. melanogaster* (124). It is possible to have even a fourth class of micro-RNAs, which may serve as guides for modification of chromosomal DNA and control the epigenetic processes of nuclear genomes (225).

**Genetic Diversity in Species-Specific Biosynthesis of Micro-RNA**

The siRNA and micro-RNA pathways closely parallel each other. It has been mentioned earlier that the biosynthesis of siRNAs have interesting system-specific features. Hence, system-specific features of micro-RNAs would also be of no surprise. Here, we illustrate some of those features. In *C. elegans*, *D. melanogaster*, and other animals, the Dicer proteins responsible for siRNA formation are also involved in the biosynthesis of micro-RNAs; but in *A. thaliana*, DCL1 (or CAF1) is responsible for micro-RNA but not for siRNA formation (76). Interestingly, the DCL1 mRNA is predicted to be a micro-RNA target, indicating that the micro-RNA-related apparatus in plants is regulated by a negative feedback loop (233). In plants, HEN1 is required for both siRNA and micro-RNA formation (22). Such a role for HEN1 orthologues in other systems is not known yet. Both CAF1 and HEN1 have nuclear localization sequence signals, raising the question of whether plant micro-RNAs are made intranuclearily.

The functions of plant micro-RNAs may be different from those of their animal counterparts in some events. The animal micro-RNAs act as translational repressors, whereas some plant micro-RNAs act on the target mRNA posttranscriptionally, like siRNAs (139). In animals, the majority of the AGO family members tightly regulate the biosynthesis of micro-RNAs (32), whereas in plants, especially *A. thaliana*, only one member of the 10 constituents of the AGO family, Zwielle, alone contributes to the synthesis of micro-RNA. Surprisingly, though *A. thaliana ago1* mutants show a strong hypermorphic phenotype, AGO1 protein is not responsible for plant micro-RNA formation (22). However, AGO1 is required for initiation of PTGS, whereas Zwielle is not. Recently, Vauchert et al. also isolated a few *ago1* alleles of *A. thaliana* which were hypomorphic in nature (157).

The biogenesis of some plant micro-RNAs seems to be different from that of their animal counterparts. Most of the *Arabidopsis* micro-RNAs belong to group1, as their precursor forms are detected poorly or not at all. Despite the absence or greatly reduced abundance of the mature micro-RNAs, accumulations of pre-micro-RNA are never detected in Dicer-defective *cafl* mutants (171). However, the pre-micro-RNAs accumulate to a higher level than *C. elegans* and metazoans in which Dicer activity is abolished or reduced (86). Very few *Arabidopsis* micro-RNAs belong to group II, including the micro-RNAs *mir176*, *-177*, *-178*, and *-179*, the pre-micro-RNA transcripts of which are, however, detectable. The levels of these precursor transcripts do not change in either the *cafl* or *hen1* mutant background. Such facts indicate that even within the same plant, the biosynthesis pathways of micro-RNAs might vary depending on the particular micro-RNA. The tissue specificity of micro-RNAs is well known. Hence, micro-RNAs specific to tissues that are unique either to animals (e.g., brain) or plants (roots, for example) might exemplify variant pathways of biosynthesis of micro-RNAs.

The discovery of micro-RNAs has been branded one of the top discoveries in developmental molecular biology. The survey of micro-RNAs is still at a subsaturated stage. The future will witness the discovery of hundreds of new micro-RNAs and
their corresponding mRNA targets, and the mysteries of developmental pathways from embryogenesis to adulthood will be unfolded.

**SMALL-RNA-MEDIATED EFFECTS ON CHROMOSOMAL DNA**

The siRNAs work not only at the posttranscriptional stage but also leave their indelible marks on the genomes to repress the gene transcription activity or selectively remove portions of the genomes, especially of protozoans. These stunning discoveries have been reported only in the span of the last 2 years, the detailed mechanisms of which are still to be revealed and have been reviewed in two recent articles (57, 109). In the present review, we describe these effects briefly with special emphasis on plant systems, since the genetics and biochemistry of some of these processes are better illustrated in plants.

Broadly speaking, the siRNAs bring about three different biochemical end products with the chromatin DNA: DNA methylation, as revealed mostly in plant systems; heterochromatin formation; and programmed elimination of DNA. DNA methylation had been reckoned a major source of transcriptional gene silencing (TGS), and mechanistically TGS had been viewed very distinctively from PTGS in the past. But recent developments have caused a blurring in the identity between these two pathways (218), and some of these developments will be highlighted below. The discoveries of such epigenetic changes have ignited a revolution not only in the field of gene regulation but also in gene maintenance and gene evolution.

**RNA-Dependent DNA Methylation**

A role for RNA in guiding de novo cytosine methylation of homologous DNA sequences was first discovered in viriod-infected plants and subsequently also in nonpathogenic plant systems (194). When the dsRNA degradation mediated PTGS occurs in plants, the genomic DNA regions homologous to dsRNA are often found methylated at almost all the sensitive cytosine residues. This process is generally referred to as RNA-dependent-DNA methylation and the corresponding part of the genome, especially the promoter region might remain transcriptionally silent. The initiator of RNA-dependent DNA methylation/TGS could be either the transgene-derived dsRNA or the consequent siRNA (110, 111, 214). Depending on the sequence information of the dsRNA, RNA-dependent DNA methylation was found to occur at the open reading frame and/or the promoter region of the genome (10, 149). If methylation occurred only at the open reading frame, TGS did not result. However, RNA-dependent DNA methylation at the promoter sequences induced TGS, which, unlike PTGS, was stable and heritable (98). RNA-dependent DNA methylation within the host genes has also been found to occur preponderantly during virus-induced gene silencing, a type of RNAi that is generally initiated by plant virus vectors carrying portions of host genes, as has been described earlier (214).

It was demonstrated that the movement of transposons was controlled by transcriptional suppression (TGS) and that methylation also played a role in this suppression, depending on the nature of the transposon (226). In animals and lower plants, siRNAs corresponding to the transposable elements were discovered and cloned earlier (9, 232), and in *A. thaliana* and *Nicotiana* species, the siRNAs corresponding to retroelements have recently been discovered (94). These siRNAs are perhaps responsible for the methylation of the homologous DNA.

There are also conflicting data in the literature concerning the cause-and-effect relationship between PTGS and DNA methylation. In some examples, there is no correlation between PTGS and DNA methylation (153). In other events, as mentioned earlier, the correlation is strong (137). Llave et al. (137) showed that a viral protein, HC-PRO, that suppresses PTGS/RNAi, when introduced into GUS-silenced tobacco, inhibited the maintenance of small RNAs and caused a concomitant decrease in methylation of the GUS sequence in the plant genome. This study suggested that DNA methylation of the silenced gene could be directly correlated with PTGS. However, in a contrasting study carried out by Mette et al. (153), HC-PRO was found to increase the methylation of a target promoter DNA when gene silencing was induced by the promoter dsRNA. The later study also revealed that the amount of promoter siRNA was elevated fivefold in the presence of HC-PRO. Taken together, both of these studies indicate that the level of target DNA methylation is directly related to the amount of siRNA present in the cell, and thus the apparent differences between these observations can be resolved. In other words, the availability of siRNA may determine the level of RNA-directed DNA methylation. In the events of RNA-dependent DNA methylation, the chromodomain containing DNA methyltransferase acts either alone or in combination with other proteins, such as piwi-containing proteins, to form complexes with the siRNAs and cause sequence-specific RNA-dependent DNA methylation, finally resulting in TGS (10).

Evidence of cross talk between PTGS and TGS has been obtained from the mutational analysis of *A. thaliana* and *D. melanogaster*. Two types of *A. thaliana* mutants, *ddm1* (deficient in DNA methylation) and *met1* (methyl transferase), were isolated from a screen of mutations causing a reduction in global methylation of the genome. The locus *ddm1* encodes an SNF2/SW12-like chromatin-modeling protein, whereas MET1 is a major DNA methyltransferase. Both of these mutants exhibit marked reduction in PTGS activity, as measured by the accumulation of transgene transcripts (10, 218). Although the patterns of reduction are different with these mutants, these studies highlight the strong correlation between PTGS and TGS.

In *D. melanogaster*, polycomb protein-dependent TGS is also affected by mutations in PIWI, a family of proteins required for RNAi (169). Other evidence includes the *argonaute* gene of *A. thaliana*, which controls both locus-specific siRNA accumulation and DNA methylation (241); the *Arabidopsis sde4* locus, which is of unknown biochemical function but is responsible for (retroelement TS SINE-specific) siRNA formation (94); and the *Arabidopsis rts1* (RNA-mediated transcription silencing) mutation, which causes a 50% reduction in target promoter DNA methylation (10). However, not all TGS mutations affect the PTGS pathways and vice versa, suggesting that the two pathways diverge at some point (218).

RNA-dependent DNA methylation has been reported only in plants until now. Aufsatz et al. (10) have also shown that
asymmetric non-CpG methylation is mostly affected by RNA-dependent DNA methylation, but the existence of non-CpG methylation in mammals has always been a contentious issue. Mammalian DNA is methylated mostly at symmetric CpG or CNG sites by various forms of DNA methyltransferases. However, using a dual-labeling nearest-neighbor technique and the bisulfite genomic sequencing methods, Ramsahoye et al. (177) found that the genomes of embryonic stem cells but not that of somatic tissues harbored non-CpG methylation, which accounted for 15 to 20% of total cytosine methylation. This methylation is perhaps caused by the methylase Dnmt3a, which is highly expressed in embryonic stem cells but poorly expressed in somatic tissues (179). Other studies have also revealed that in D. melanogaster and mammals, non-CpG methylation is an early embryonic event (10), and this methylation can be catalyzed by Dnmt2, which is primarily active at the initial stages of development (142).

In the above-mentioned studies, however, no connection between non-CpG methylation and any homologous RNA has been shown. Hence, if RNA-dependent DNA methylation occurs at all in animals, it might be limited to the early developmental stages when the effector proteins may be found in abundance. In contrast, RNA-dependent DNA methylation is observed throughout plant development, implying the continuous availability of the appropriate plant DNA methyltransferases. This feature also explains the ease of RNA-dependent DNA methylation detection in plants (10).

**Heterochromatin Formation**

Even for organisms in which RNA-dependent DNA methylation is supposedly absent, there is growing evidence that RNAi processes cause chromatin modifications leading to TGS. This evidence reveals that the connections between TGS and PTGS are strong across all layers of eukaryotic life. For example, in C. elegans, in which DNA methylation has not been detected, some PTGS mutations, namely mut7 and re62, derepress transgenes which are affected by polycomb-dependent TGS (203). The polycomb group of proteins are known to keep the chromatin in the closed or compact conformation. Conversely, it has also been found recently that the polycomb proteins MES3, MES4, and MES6 are required for RNAi, at least under some experimental conditions (65, 121).

Generally, in eukaryotic systems, histone modifications make the chromatin structure inert to transcription by heterochromatin formation, which is modulated greatly by the RNAi processes, as recent discoveries have revealed. In almost all organisms heterochromatin formation requires that histone H3 of the chromatin be deacetylated and then methylated at lysine 9. The SET domain of a special group of histone methyltransferases carries out this function. This methylated lysine is subsequently bound by a heterochromatin binding protein, HP1. The binding of the chromodomain containing HP1 to Met H3-K9 is highly specific and of very high affinity (12). This binding may be followed by multimerization of HP1 and complex formation with other chromatin-remodeling proteins. As a result of this multicomplex formation, the chromatin becomes condensed and locked in a transcriptionally repressed heterochromatic state.

Once formed, the heterochromatin spreads a large distance due to cooperative protein-protein interactions of chromatin-remodeling factors, the components of which have not been fully identified yet. However, these structures are generally initiated at places containing repeated DNA sequences, for example, centromere, telomere, mating locus, and elsewhere in the genome containing repetitive DNA in the fission yeast Schizosaccharomyces pombe (7). These repeats are responsible for producing dsRNAs, which are processed by the RNAi machinery. D. P. Bartel’s group has discovered abundant species of centromeric repeat-specific siRNAs from S. pombe (180). Volpe et al. (223) demonstrated that these siRNAs are blocked and instead, large noncoding RNAs (1.4 to 2.4 kb) homologous to the centromere repeats accumulate in dcr1, ago1, and rdrp1 mutants of S. pombe. These mutant cells also do not show the heterochromatin-mediated silencing of a ura4+ gene inserted into the outer and inner repeats that flank the central core of the centromeres. A corresponding reduction in Met-H3 K9 is also observed in the outer repeats of these mutant cells (223). This loss in gene silencing is phenotype-specific in cells lacking the histone methyltransferase (clr4) or the HP1-like (Swi6) activity.

That the DNA repeats are central to the RNAi-like processing of dsRNA and concomitant heterochromatin formation was clearly established by the findings of Hall and colleagues (89), who inserted a 3.6-kb centromere H repeat, normally present at the silent mating type domain, in a euchromatic position (ura3 locus). The introduction of this repeat was sufficient to turn on the silencing of a linked reporter gene and induce H3-K9 methylation and recruitment of HP1-like factors (Swi6) (89). The link between the RNAi machinery and heterochromatin formation has also been established by a recent finding in A. thaliana. From a large screen of mutants, Zilberman et al. (241) found that the ago4 gene is responsible for the RNAi-related silencing of the A. thaliana superman gene, which is implicated in flower formation. The ago4-1 mutation reactivates the silent superman allele and decreases non-CpG as well as H3-K9 methylation. Significantly, the same mutation also blocks DNA methylation and the accumulation of siRNA corresponding to the retroelement at SN1 (241).

The above-mentioned facts are put together in a model (Fig. 3) showing a link between siRNA and heterochromatin formation. In the wild-type scenario, one strand of the centromeric region is constitutively expressed, whereas the complementary strand, which is subjected to heterochromatic repression, is occasionally transcribed (57). Such transcription will lead to the formation of dsRNA, which will be processed by the RNAi machinery. This processing might even be a nuclear step, since a component of this machinery, the RdRP, was found to be physically bound to the outer repeats of the centromeric region in a chromatin immunoprecipitation assay (223). The siRNA thus formed might enter a complex containing the histone methyltransferase enzyme. This complex could be a nuclear equivalent of the RISC complex (Nu.RISC of Fig. 3) lacking nuclease activity (98). Such a complex would be guided to the appropriate DNA region following the DNA-RNA base pairing rules, and the histone H3-K9 of the region might be methylated to eventually generate the heterochromatin structure. Since RdRP is found locally, the spread of the heterochromatic structure may be associated with the extension of the 3’ end of the siRNA primer. It has also been shown in N. crassa and A.
FIG. 3. Role of the RNAi process in heterochromatinization of nuclear DNA. The methylated (H3-K9) histone and many chromatin components are involved in the cross talk in several epigenetic regulatory pathways within the nucleus. The centromeric region (black oval) of the chromatin (thick purple line) might be responsible for the production of dsRNA transcripts (continuous red line and broken blue line). The transcript coming out of the DNA strand subjected to heterochromatinization is represented by broken blue lines. The siRNA binding the chromodomain (CHR), methyltransferase (MTase), and the SET domain of the methyltransferase are indicated.
thaliana that H3-K9 methylation directs DNA methylation (107, 204). The methylated DNA could be complexed further with the methyl-binding proteins. Following these binding events, the chromatin structure will be extremely compact and condensed and would remain transcriptionally inert.

DNA Elimination

The most dramatic effect of siRNA-mediated heterochromatin formation followed by chromosomal DNA elimination and rearrangement has been recorded in the ciliated protozoan Tetrahymena pyriformis (156, 206). Among unicellular organisms, T. pyriformis is unique because of its nuclear dimorphism. The two nuclei, the micronucleus and macronucleus, serve different functions. The polyploid macronucleus is the transcription center of the cell during vegetative growth, whereas the diploid and transcriptionally inert micronucleus acts as the germ line nucleus. During conjugation, the micronucleus gives rise to the macronucleus, and this transition is accompanied by two interesting and peculiar recombinant events. First, approximately 6,000 internal eliminated sequences of five pairs of micronuclear chromosomes, accounting for about 15% of genomic micro-DNA, are removed. Second, the remaining parts of these chromosomes are broken into 200 to 300 minichromosomes concomitant with the deletion of <50 nucleotide breakage eliminated sequences. The mechanisms of removal of internal eliminated sequences and breakage eliminated sequences remained elusive for a long time but were recently unveiled, courtesy of the awareness of the siRNA world.

Mochizuki et al. (156) showed that in wild-type cells of Tetrahymena pyriformis, siRNAs of about ~26 to ~31 nucleotides were produced which hybridized to micronuclear genomic DNA and not the macronuclear DNA, indicating that these siRNAs could be internal eliminated sequence/breakage eliminated sequence-specific and are referred to as scan RNAs. These scan RNAs were not made in twi1 mutants, and the production of internal eliminated sequence/breakage eliminated sequence elements was also impaired by the twi1 mutation. twi1 produces a Piwi-related protein during the sexual cycle and can transmit the RNA-encoded information from the micronucleus to the old macronucleus and finally to the new macronucleus to mark the sequences to be eliminated (57). The defect in accumulation of the scan RNAs in the twi1 mutant was similar to the case of another mutant, pdd1 (156). In a related report, Taverna et al. (206) showed that the protein PDD1 was the effector protein for DNA excision and that PDD1 along with Met H3-K9 was associated preferentially with the internal eliminated sequence/breakage eliminated sequence elements in the new macronucleus that developed from the micronucleus during the sexual cycle. PDD1 contains two chromodomains and an additional RNA-binding domain (3).

The above data and the model presented in the earlier section lead to a straightforward and interesting scheme for programmed DNA degradation in Tetrahymena.pyriformis. The bidirectional transcription that occurs across the internal eliminated sequence repeats (38) may form the dsRNA, which would give rise to the scan RNAs following the action of RNAi-related Dicing complexes that perhaps also include the Twi1 and PDD proteins. These scan RNAs eventually may be associated with the nuclear equivalents of RISC factor in the new macronucleus to provide heterochromatic sites at the internal eliminated sequence/breakage eliminated sequence regions. The chromodomain containing PDD proteins may remain bound to the scan RNA and thus guide to destroying the cognate DNA. As an extension of this work, Yao et al. found that a similar RNAi process recognized and deleted a foreign neomycin resistance gene of bacterial origin which was integrated in a Tetrahymena chromosome (237). These two studies together strongly suggest an siRNA- (or scan-RNA)-based mechanism that controls genome-wide DNA arrangements and provides genomic surveillance against invading foreign DNAs.

Thus, the Tetrahymena pyriformis as well as S. pombe data show how dramatic the epigenetic consequences of the genome could be following the formation of siRNA molecules in cells. Discovery of the link between the RNAi processes and the epigenetic chromat modification as well as chromosome behavior is probably the most fascinating and novel face of regulation of gene silencing mechanism. The RNAi machinery is reported to control many explosive features of cellular biology, namely stem cell maintenance (53), cell fate determination (21), nonrandom chromosome segregation (188), etc. A recent report established that the fission yeast RNAi-related genes ago1, dcr1, and rdp1 also control the fidelity of chromosome segregation during mitosis and meiosis. As discussed earlier, these gene products are required to maintain centromeric silencing. The report also demonstrated that the chromosome missegregation of the RNAi mutants occurred due to the loss of centromeric cohesion, suggesting a clear link between centromeric silencing and cohesion. This report broadly hinted that the regulation of chromosomal dynamics could be largely traced to the natural RNAi biology of the eukaryotic cells (90).

It is not difficult to imagine that we might witness RNAi-related unifying signals in diverse chromosome behaviors, namely X-chromosome inactivation, satellite-repeat contraction and expansion, hybrid dysgenesis in D. melanogaster, chromatin diminution in ascarid nematodes, nuclear dominance in plants, and so on in the not so distant future (57).

APPLICATIONS OF RNAi

Besides being an area of intense, upfront basic research, the RNAi process holds the key to future technological applications. Genome sequencing projects generate a wealth of information. However, the ultimate goal of such projects is to accelerate the identification of the biological function of genes. The functions of genes can be analyzed with an appropriate assay, by examining the phenotype of organisms that contain mutations in the gene, or on the basis of knowledge gained from the study of related genes in other organisms. However, a significant fraction of genes identified by the sequencing projects are new and cannot be rapidly assigned functions by these conventional methods.

RNAi technology is proving to be useful to analyze quickly the functions of a number of genes in a wide variety of organisms. RNAi has been adapted with high-throughput screening formats in C. elegans, for which the recombination-based gene knockout technique has not been established. Chromosomes I
and III of *C. elegans* have been screened by RNAi to identify the genes involved in cell division and embryonic development (82, 84). Recently, a large-scale functional analysis of ~19,427 predicted genes of *C. elegans* was carried out with RNA interference. This study identified mutant phenotypes for 1,722 genes (112). Similarly, in *D. melanogaster*, RNAi technology has been successfully applied to identify genes with essential roles in biochemical signaling cascades, embryonic development, and other basic cellular process (44). In plants, gene knockdown-related functional studies are being carried out efficiently when transgenes are present in the form of hairpin (or RNAi) constructs. Plant endotoxins could also be removed if the toxin biosynthesis genes are targeted with the RNAi constructs. Recently, the theobromine synthase of the coffee plant was knocked down with the hairpin construct of the transgene, leading to the production of decaffeinated coffee plants (166). Virus-induced gene silencing has also been proven to be a successful approach for plant genetics (15).

Given the fact that RNAi is easy to apply, whole-genome screens by RNAi may become a common method of choice in the near future. RNAi may facilitate drug screening and development by identifying genes that can confer drug resistance or genes whose mutant phenotypes are ameliorated by drug treatment, providing information about the modes of action of novel compounds. Although RNAi is unlikely to replace the existing knockout technology, it may have a tremendous impact for those organisms that are not amenable to the knock-out strategy. It may also be a method of choice to study the simultaneous functions of a number of analogous genes in organisms in which redundancy exists with respect to a particular function, because many of these genes can be silenced simultaneously.

Given the gene-specific features of RNAi, it is conceivable that this method will play an important role in therapeutic applications. Since siRNAs direct cellular RNAi biology, these are potential therapeutic reagents because of their power to downregulate the expression pattern of mutant genes in diseased cells. However, central to this hypothesis is the assumption that the effect of exogenous siRNA applications will remain gene specific and show no nonspecific side effects relating to mismatched off-target hybridization, protein binding to nucleic acids, etc. Though it was demonstrated that mismatches of more than even one nucleotide within the 19- to 20-mer siRNAs effectively disrupted proper degradation of the target mRNA (68), the gene specificity of siRNAs needs to be confirmed on a genome-wide scale.

Recently, Chi et al. (41) reported that the GFP siRNA-induced gene silencing of transient or stably expressed GFP mRNA was highly specific in the human embryonic kidney (HEK) 293 cell background. The specific silencing did not produce secondary changes in global gene expression, as detected by the DNA microarray experiment. They also failed to detect the presence of transitive RNAi in experimentally engineered human cell lines (41). In their own experiments, Semizarov et al. (190) reached a similar conclusion while using siRNAs corresponding to akt1, rb1, and pkl1 in the human non-small cell lung carcinoma cell line H1299. These experiments prove that siRNAs could be used as highly specific tools for targeted gene knockdown and can be used in high-throughput approaches and drug target validation. This exquisite sequence-specific effect of siRNAs has also been exploited in silencing the mutant allele of the diseased gene while not affecting the wild-type allele of the healthy version of the same gene (158).

siRNAs have been shown to inhibit infection by human immunodeficiency virus, poliovirus, and hepatitis C virus in cultured cell lines (152). Bitko and Barik (19) successfully used siRNAs to silence genes expressed from respiratory syncytial virus, an RNA virus that causes severe respiratory disease in neonates and infants. siRNA treatment has also been shown to reduce the expression of the BCR-ABL oncoprotein in leukemia and lymphoma cell lines, leading to apoptosis in these cells (230). With respect to future medical applications, siRNA-based therapy seems to have a great potential to combat carcinomas, myeloma, and cancer caused by overexpression of an oncoprotein or generation of an oncoprotein by chromosomal translocation and point mutations (211).

Recently, the therapeutic potential of the siRNA technique has been demonstrated in vivo in mouse models. McCaffrey et al. (151) and Song et al. (199) demonstrated effective targeting of a sequence from hepatitis C virus and the fas gene by RNA interference in mouse liver (199). An epiallelic series of p53 hypomorphs created by RNAi have been shown to produce distinct tumor phenotypes in mice in vivo, suggesting that RNAi can stably suppress gene expression (99). Song et al. (199) have shown that treatment with fas siRNA abrogated hepatocyte necrosis and inflammatory infiltration and protected mice from liver fibrosis and fulminant hepatitis. Rubinson et al. (184) showed highly specific, stable, and functional silencing of gene expression in transgenic mice with the lentivirus system for the delivery of siRNAs.

Although the delivery of siRNAs to a proper site remains problematic for gene therapy, chemical modifications of siRNAs such as changing the lipophilicity of the molecules or the methods previously developed for the application of antisense oligonucleotides or nuclease-resistant ribozymes might help the entry and stability of siRNAs within the transfected cells or tissues. The absence of specific micro-RNAs has been demonstrated in carcinoma cells, implying that cancer development could be arrested by introduction of the missing micro-RNAs. The micro-RNAs could be supplied in the form of siRNAs, since the function of micro-RNAs can be mimicked by the exogenous siRNA (62). However, independent of its biomedical applications, RNAi appears to be a forthcoming method for functional genomics.

**CONCLUDING REMARKS**

In the footsteps of the discovery of the double-helical structure of DNA, some outstanding discoveries have been recorded, but few of them really match the explosive content and implication of dsRNA-mediated gene silencing. This homology-dependent silencing has established a novel paradigm with far-reaching consequences in the field of transcription regulation. The regulatory mechanism offers cellular protection against parasitic nucleic acid sequences, carries out epigenetic as well as genetic alterations on the one hand, and governs organisms architecture and development on the other. Capitalizing on the basic principles of silencing, large-scale functional genomics have come into play in diverse organisms.
Studies conducted at the laboratory level have revealed the tremendous power of siRNAs as therapeutics and have demonstrated the potential of micro-RNAs to reverse cellular developmental aberrations.

The new paradigm has a lot more to offer than it has delivered already. The stepwise detailed mechanism of RNAi and its related processes is waiting to be explored. The rationale for many unexplained genetic findings of RNAi in worms, plants, and other organisms will be revealed in the wake of further mechanistic discoveries. The cytoplasmic location of RNAi is evident, but the evidence of nuclear connections of RNAi and related events are also too many. Surprisingly, there are some components of RNAi, GEMIN3 and GEMIN4 of humans, which partition in both the nuclear and cytoplasmic compartments. Hence, clarification of the subcellular locations of the RNAi processes is required. Hopefully, the detailed biochemical framework of RNAi would provide such clarifications.

As we gain more insight into the mechanisms, more effective methods for analysis of gene functions may evolve. We may learn more about geriatrics, nervous diseases, genetic imprints, nuclear dominance in plants, and so on and thus might wield control over such processes in the future. Meanwhile, the knockdown technology might improve vastly with better-designed plasmid- or virus-based vectors for delivery of siRNAs to the appropriate tissues at the appropriate time. Such technology is bound to give a new shape to therapeutic gene silencing as well. The science and technology of RNAi has given us a cultural ocean of virtually bottomless depth.

REFERENCES


166. Pal-Bhadra, M., U. Bhadra, and J. A. Birchler. 2002. Ribonuclease activity and RNA binding of recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant recombinant reco


