Biochemistry and modeling of human Dicer, a key protein involved in RNA interference

Inauguraldissertation

zur
erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultaet
der Universität Basel

von

Haidi Zhang
aus Beijing, China

Basel 2004

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultaet

auf Antrag von

Prof.Dr.Witold Filipowicz, Mitglied des Dissertationskomitees

Basel, den 03.02.2003

Prof.Dr.Marcel Tanner, Dekan

Acknowledgements

There are so many people I want to thanks during the past four and half years of my Ph.D study and without all those people, I could not finish it and have this thesis done!

But above all, I need to thank Dr. Filipowicz, whom I met five years ago in Beijing, believed in me and gave me the chance to come to Switzerland and work at his lab. He showed me how a real scientist should be and helped me so much of all the great works we did together with his clear vision, his great knowledge, and his enthusiasm.

Those years in Basel, I meet so many great people, especially people from the lab. I made friends of my life! Vanda, the angel who worked next to me for two years, Lukasz, the tall Polish guy who worked in front of my bench for three years! I can not describe how fortunate I felt for knowing them and being friends of them. All the things we did together, all the time we spent together, all the discussion we had together – all those great moments as sweet memory will always deeply remain in my mind.

And of course, great thanks go to my family, every one of them, all the love they selflessly give me for all the years.

Summary

Five years ago, an unexpected discovery opened a whole new paradigm of biology – RNA interference (RNAi). From the simple notion that dsRNA, when introduced into various organisms, can specificly inhibit expression of homologous endogenous gene, the RNA interference has expanded into a wide range of gene regulatory pathways of great biological importance. At the same time, RNAi started to be widely used as powerful experimental tool for probing gene function in almost any organism. The research on RNAi is moving forward at high speed, both at the mechanistic level and as a tool. Genetic and biochemical studies in various systems have revealed much information about the mechanism of RNAi. It is now well established that dsRNAs is processed by a nuclease Dicer into short dsRNAs varing in length from 21 to 25 nt, named siRNAs, which in turn are incorporated into the RNA induced silencing complex (RISC) to target mRNA degradation. Identification of siRNAs led to the discovery of a whole new class of regulatory small RNAs of similar size, named microRNAs (miRNAs), which have diverse biological functions. Hundreds of miRNAs were cloned, and their functions are being investigated. The single stranded miRNAs are also processed by Dicer from miRNA precursors and incorporated into a complex similar, if not identical, to RISC. In animals, miRNAs imperfectly base-pair with mRNA leading to translational repression. Dicer, a central protein of the RNAi and miRNA pathways is a focus of the study presented in this thesis. A full length human Dicer cDNA was cloned and protein overexpressed in the baculovirus system and purified. Its processing activity was demonstrated using both dsRNA and pre-miRNAs as substrates. Detailed study of the RNase III-like activity of Dicer, its biochemical properties and a model of its function are described in two experimental chapters of this thesis.

This thesis is divided into three major chapters followed by a short general discussion. Chapter 1 contains a general introduction to RNA interference. It describes a history of RNAi discovery, summarizes what is known about the RNAi mechanism in general, and also about the species-specific differences. The mechanistic aspects of the miRNA pathway are also described. An overview of all important proteins involved in RNAi is presented. Finally, a summary of RNAi as a tool for reverse genetics is provided.

Chapter 2 describes the characterization of the purified recombinant human Dicer protein. In vitro experiments showed that the purified protein cleaves dsRNAs into ~22 nucleotide siRNAs. This was a first direct evidence that Dicer indeed has RNase III-like nuclease activity. Accumulation

of processing intermediates of discrete sizes, and experiments performed with substrates containing modified ends, indicated that Dicer preferentially cleaves dsRNAs at their termini. Binding of the enzyme to the substrate could be uncoupled from the cleavage step by omitting Mg^{2+} or performing the reaction at 4°C. Activity of the recombinant Dicer, and of the endogenous protein present in mammalian cells extracts, was stimulated by limited proteolysis, and the proteolysed enzyme became active at 4°C. Cleavage of dsRNA by purified Dicer and the endogenous enzyme was ATP independent, in contrast to results obtained in *Drosophila* and *C. elegans*. Additional experiments suggested that if ATP participates in the Dicer reaction in mammalian cells, it might be involved in the product release needed for the multiple turnover of the enzyme.

Chapter 3 describes the mutagenesis study of the human Dicer RNase III domains, which revealed that Dicer contains a single compound catalytic center. Both RNase III domains in Dicer contribute to the dsRNA cleavage reaction. The Dicer mutagenesis study was initiated whether a model of dsRNA cleavage originating from an X-ray structural study of the *Aquifex aeolicus* RNase III also applies to Dicer. Mutants containing changes in residues implicated in the catalysis in both Dicer RNase III domains were prepared to study their effect on RNA processing. Our results were in conflict with the bacterial Rnase III model and all speculated Dicer model. We have further mutated the catalytic residues of the *E. coli* RNase III and tested their effect on processing of different substrates. The results are consistent with those obtained with Dicer mutants. More specifically, our results indicate that instead the two catalytic centers proposed previously, both enzymes contain only one catalytic center, generating products with 2-nt 3' overhangs. Together with other data, a new model was proposed according to which Dicer functions as an intra-molecular dimer of its two RNase III domains, assisted by the flanking RNA binding domains, PAZ and dsRBD.

Chapter 1

Introduction

RNA interference (RNAi), a conserved biological response to double-stranded RNA (dsRNA), mediates resistance to both endogenous parasitic and exogenous pathogenic nucleic acid, and regulates the expression of protein-coding genes. Over the past five years, the way in which cells respond to dsRNA by silencing homologous genes has revealed a new regulatory paradigm in biology. We are only beginning to appreciate the mechanistic complexity of this process and its biological ramifications. Meanwhile, RNAi has already begun to revolutionize experimental biology in organisms ranging from unicellular protozoans and fungal to mammals.

1.1 Discovery of RNA interference

The discovery of RNAi came out of a desire to use antisense approaches to probe gene function in *Caenorhabditis.elegans.* Guo and Kemphues tried to determine the function of *par-1* gene by injecting antisense *par-1* RNA into worms (Guo and Kemphues, 1995). Although the antisense approach resulted in the expected phenotype, a serious paradox was raised by the observation that sense RNA injected as control also had an identical phenotype. The key breakthrough came when Fire and Mello asked whether injection of both the sense and the anitsense strands into worms might give an additive effect. Shockingly, the mixture of sense and antisense strands silenced expression of a target gene roughly tenfold more efficiently than either strand alone. Interpreting this dsRNA-induced effect as a new phenomenon, the authors named the process RNA interference (RNAi)(Fire et al., 1998)

The ability of dsRNA to affect gene expression was already well known in mammals (Hunter et al., 1975). The key difference between this response and RNAi was their respective specificity: the former inhibited gene expression globally via activating a protein kinase, whereas RNAi had a specific effect on gene expression. One of the first indications that RNAi was a novel biological phenomenon was the potency of its effect. Injecting the worm with only a few molecules of dsRNA per cell was sufficient to almost completely silence the expression of a specific gene. Furthermore, the effect seemed to be systemic. Injection of dsRNA into the gut of the worm caused silencing throughout the animal, and also in the F1 progeny.

From this discovery emerged the notion that a number of previously characterized, homology-dependent gene-silencing mechanisms might share a common biological root. Couple of years earlier, Richard Jorgensen had engineered transgenic petunias to alter pigmentation (Jorgensen et al., 1996). However, introducing exogenous transgenes did not deepen the flower colour as

expected. Instead, flowers showed variegated pigmentation with some lacking pigment altogether. This indicated that not only were the transgenes themselves inactive, but also that the added DNA sequence somehow affected expression of the endogenous gene. This phenomenon was called co-suppression (Jorgensen et al., 1996; Que and Jorgensen, 1998). Following this, many similar events of co-suppression were reported (Birchler et al., 1999). Since all cases of co-suppression resulted in the degradation of endogene and transgene RNAs after nuclear transcription had occurred, the phenomenon was also named as posttranscriptional gene silencing (PTGS). Besides those cases mentioned above, homology-driven RNA degradation also occurs during the propagation of viral genomes in infected plants (English et al., 1996). Viruses can be either the source, the target, or both the source and the target of silencing. PTGS mediated by viruses (VIGS) can occur with RNA viruses, which replicate in the cytoplasm, and also with DNA viruses, which replicate in the nucleus. While reports of PTGS in plants were piling up, homology-driven gene silencing phenomena were also observed independently in fungal systems, which in *Neurospora crassawas* were called quelling (Cogoni et al., 1996).

What is clear in retrospect is that all those homology dependent gene-silencing reactions involve generation of dsRNA. In plant systems, dsRNA that is introduced from exogenous sources or that is transcribed from engineered inverted repeats is a potent inducer of gene silencing. Genetic and biochemical studies have now confirmed that RNAi, PTGS and quelling share similar mechanisms, and that the biological pathways underlying dsRNA-induced gene silencing exist in many, if not most, eukaryotic organisms.

1.2 Overview of the mechanism of RNAi

Since the discovery of RNAi, important insights have been gained in elucidating its mechanism. The early understanding was derived from both genetic and biochemical studies. Genetic screens were carried out in the fungus *Neurospora crassa* (Cogoni and Macino, 1997), the nematode *C.elegans* (Tabara et al., 1999), and the plant *Arabidopsis thaliana* (Fagard et al., 2000) to search for mutants defective in quelling, RNAi or PTGS, respectively. Analyses of these mutants led to the identification of proteins involved in gene silencing and also revealed that a number of essential enzymes or factors are common to these processes. Meanwhile, intensive *in vitro* biochemical studies were carried out by using *Drosophila melanogaster* extracts. As the various pieces of the RNAi machinery are being discovered, the mechanism of RNAi is emerging more clearly. A two

steps model is proposed from combination of several *in vivo* and *in vitro* experiments (Fig 1). The first step, referred to as the RNAi initiating step, involves the processing of dsRNA into discrete 21 to 25 nucleotide RNA fragment by a RNA nuclease named Dicer. In the second step, these small RNA fragments join a multinuclease complex, which degrades the homologous single-stranded mRNAs. Data from different systems suggested that this model is the core biochemical mechanism of homology-dependent gene-silencing responses. However, the varied biology of dsRNA-induced silencing – for example, the heritable and systemic nature of silencing in *C. elegans* compared to apparently cell-autonomous, non-heritable silencing in *Drosophila* and mammals – suggested that this core machinery probably has adapted to meet specific biological needs in different organisms.

1.2.1 The initiation step

When Fire and Mello first discovered RNAi, they had originally proposed that some derivative dsRNA would guide the identification of substrates for RNAi. The first clue in the hunt for such 'guide RNAs' came from a study of silencing in plants. Hamilton and Baulcombe sought antisense RNAs that were homologous to genes being targeted by co-suppression. They found a 25-nucleotide RNA that appeared only in plant lines containing a suppressed transgene, and found that similar species appeared during virus-induced gene silencing too (Hamilton and Baulcombe, 1999). More direct evidence about the initiation step came from biochemical studies in *Drosophila*. When dsRNA induced sequence-specific silencing was shown to work in *Drosophila* embryos (Kennerdell and Carthew, 1998), Tuschl et al (1999) tested weather *Drosophila* embryo extracts, previously used to study translational regulation, might be competent for RNAi. Incubation of dsRNA in these cell-free lysates from Drosophila syncytial blastoderm embryo reduced their ability to synthesize luciferase from a synthetic mRNA. This *in vitro* system is able to reproduce many of the features of RNAi (Tuschl et al., 1999). When dsRNA radiolabeled in either the sense or the antisense strand was incubated with this lysate in a standard RNAi reaction, 21- to 23- nucleotide RNAs were generated with high efficiency. Single-stranded labeled RNA of either strand was not converted to 21- to 23-nucleotide products. Such small RNAs were termed as short interfering RNAs (siRNA), and the formation of siRNAs did not require the presence of corresponding mRNAs (Zamore et al., 2000). Soon, siRNAs were identified in all tested systems in which RNAi and related phenomena exist, and it became a signature of this family of silencing pathways.

Biochemical characterization showed that siRNAs are 21–23-nt dsRNA duplexes with 2–3-nt 3' overhangs and 5'-phosphate and 3'-hydroxyl groups (Zamore et al., 2000). This structure is characteristic of an RNase III-like enzymatic cleavage pattern. This finding led to the identification of the highly conserved family of RNase III enzymes which "dice" dsRNA, so the protein was named Dicer. Dicer was first demonstrated to be needed for processing dsRNA into siRNAs in *Drosophila* (Bernstein et al., 2001). Similar experimental studies were also carried out in *C. elegans* (Knight and Bass, 2001), and genetic evidence from *C. elegans* and *Arabidopsis* has also shown that Dicer acts in the RNAi pathway (review in Hannon, 2002). However, in these systems where Dicer was shown to be necessary for the siRNA generation, it was not known whether it was also sufficient. The direct proof of Dicer cleavage activity came from the work presented in this thesis and also reported by others. Purified recombinant human Dicer protein was shown to be responsible for cleavage of dsRNAs in vitro (Provost et al., 2002; Zhang et al., 2002).

Biochemical experiments conducted in *Drosophila* embryo lysates and cultured S2 cells showed that the initiation step is ATP dependent. The rate of siRNA formation from dsRNA has been shown to be six times slower in the *Drosophila* extract depleted for ATP by treatment of hexokinase and glucose (Nykanen et al., 2001). Dicer immunoprecipitates from Drosophila as well as S2 cell extracts and from *C. elegans* extract required ATP for the production of siRNAs (Bernstein et al., 2001; Ketting et al., 2001). In addition, a recent paper also demonstrated that purified recombinant *Drosophila* Dicer needed ATP for dsRNA cleavage reaction *in vitro* (Liu et al., 2003). This ATP requirement by Dicer is unique among endonucleases. It is attributed to the presence of an ATP-dependent RNA ‘helicase’ domain at the Dicer’s N-terminus. However, it is still unclear what is a role of ATP during production of siRNAs from dsRNA. Since experiments described in Chapter 2 of this thesis, involving use of mammalian cell extracts and the overexpressed recombinant Dicer, do not show any ATP requirement for the cleavage reaction (Billy et al., 2001; Zhang et al., 2002), differences might exist between the early steps of *Drosophila* and mammalian RNAi pathways.

1.2.2 The effector step

From studies in the *Drosophila* system, a protein-RNA effector nuclease complex was isolated and shown to be responsible for recognizing and destroying the target mRNA. This complex was named RISC (RNA induced silencing complex), and the first identified component of RISC was the

siRNA, which presumably identifies its mRNA through Watson–Crick base-pairing (Hammond et al., 2001a). In *Drosophila*, Zamore and colleagues have shown that RISC is formed in embryo extracts as a precursor complex of 250-Kd; this becomes activated upon addition of ATP to form a 100-Kd complex that can cleave substrate mRNAs (Nykanen et al., 2001). The cleavage of mRNA substrate is apparently endonucleolytic, and occurs only in the region homologous to the siRNA. Experiments with the 5'-end radiolabeled target RNAs performed to map the sites of cleavage triggered by dsRNA demonstrated that each siRNA directs the endonucleolytic cleavage of the target RNA at a nucleotide across from the center of the guide siRNA strand.

RISC purified from *Drosophila* S2 cells is a 500-Kd ribonucleoprotein complex with slightly different characteristics (Hammond et al., 2001b). In embryo extracts, RISC* (the 100-Kd active RISC species) cleaves its substrates endonucleolytically. However, cleavage products are never observed even with the most highly purified RISC preparations from S2 cells, suggesting the presence of an exonuclease in this enzyme complex. Therefore, the complex formed *in vivo* probably contains additional factors that account for observed differences in size and activity. Alternatively, RISC purified from S2 cells may become activated — perhaps changing size and subunit composition — upon incubation with ATP.

The first identified protein component of RISC was AGO2, which co-purified with RISC from S2 cells (Hammond et al., 2001a). AGO2 belongs to Argonaute gene family - a large, evolutionarily conserved gene family found in most eukaryotic genomes. In a later section, this family of proteins will be discussed in detail. Argonaute proteins were linked to RNAi also by genetic studies in *C. elegans* (Grishok et al., 2001; Tabara et al., 1999). Two mutants identified in *C. elegans*, *rde*1 and *rde*4, are required for initiation of silencing in a parental animal. However, neither function was required for systemic silencing in F1 progeny. RDE1 is an Argonaute family protein and RDE4 is a small dsRNA-binding protein. Both proteins can interact with the *C. elegans* Dicer (Tabara et al., 1999). Similarly, in *Neurospora*, mutations in the Argonaute family member qde-2 eliminate quelling, but do not alter accumulation of siRNAs (Catalanotto et al., 2002). Additional protein components of RISC were identified. Two RNA binding proteins, the Vasa intronic gene VIG and dFMR proteins, were found in the RISC complex isolated from Drosophila (Caudy et al., 2002) together with Tudor-SN (tudor staphylococcal nuclease, TSN), a protein containing five staphylococcal/micrococcal nuclease domains and a tudor domain (Caudy et al., 2003). dFMR is a homologue of the human fragile X mental retardation protein (FMRP, further discussion of FMRP will be in a later section). A novel ribonucleoprotein complex from the *Drosophila* lysate was also

isolated, which contains dFMR1, AGO2, a *Drosophlia* homologue of p68 RNA helicase (Dmp68), and two ribosomal proteins, L5 and L11, along with 5S rRNA (Ishizuka et al., 2002). The nuclease which cuts the mRNA in RISC is certainly one of the important yet-to-be-identified components required for the effector step. Finally, the cleaved mRNAs are likely degraded by exoribonucleases. The RISC component TSN was demonstrated to have non-sequence-specific nucleases activity and cleave both RNA and DNA. It was proposed that the degradation of the mRNA processed by RISC is carried out by this enzyme (Caudy et al., 2003).

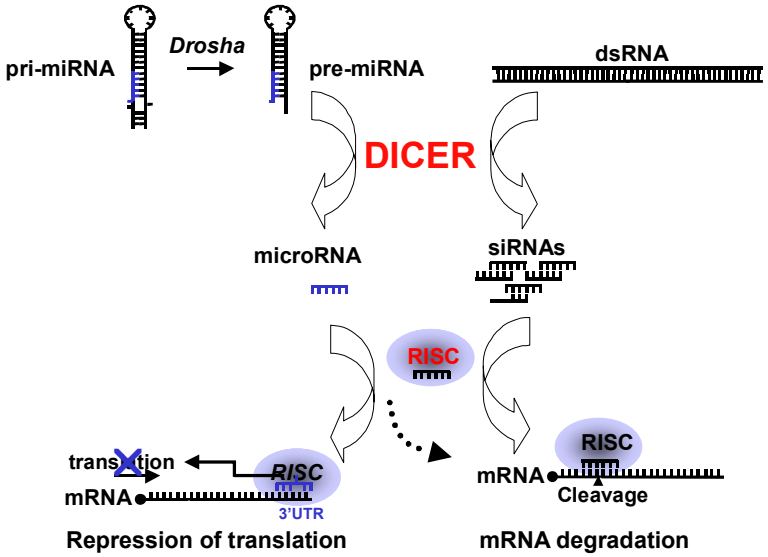

Figure 1. RNAi pathways. Dicer processes dsRNA or pre-miRNA to generate siRNA or miRNA, which will form RISC to target the mRNA via mRNA degradation or translational repression.

1.2.3 Amplification and spreading of silencing

One of the most intriguing aspects of RNAi in *C. elegans* is its ability to spread throughout the organism, even when triggered by a few molecules of dsRNA (Fire et al., 1998). Similar systemic silencing phenomena have been observed in plants, in which silencing could spead the whole plant or even be transferred to a naive grafted scion (Palauqui et al., 1997), but were not found in flies and mammals.

For systemic silencing to occur, it requires firstly a system to pass a signal from cell to cell, and an additional strategy for amplifying the signal. A phenomenon termed 'transitive RNAi' has provided some useful clues. Transitive RNAi refers to the movement of the silencing signal along a particular gene. In *C. elegans*, targeting the 3' portion of a transcript results in the production of siRNAs homologous to the targeted region and suppression of the gene. In addition, siRNAs

complementary to regions of the transcript positioned upstream from the targeted area also accumulate. If these siRNAs are complementary to other mRNAs, these mRNAs are also targeted (Sijen et al., 2001a). In various systems, a class of proteins, similar in sequence to a tomato RNA dependent RNA polymerase (RdRP), was identified as required for RNAi (Cogoni and Macino, 1999; Mourrain et al., 2000a; Schiebel et al., 1998). Involvement of this class of proteins in RNAi might explain how the transitive RNAi works. Genetic studies led to a simple model for transitive RNAi in which siRNAs might prime the synthesis of additional dsRNA by RdRPs. In the later section, RdRPs in various organisms will be discussed in more detail. As discussed in a previous section, the fact that RDE-1 and RDE-4 are required only for the initiation of RNAi in parental *C. elegans,* adds an additional layer of complexity to the model. Perhaps exogenous dsRNAs are recognized initially in manner that is distinct from recognition of secondary dsRNA, which may be produced by RdRPs. For example, the proposed function of RDE-4 in delivering dsRNA to Dicer could be substituted for secondary dsRNAs by another hypothetical protein. Alternatively, Dicer could exist in a stable complex with an RdRP, making dsRNA delivery unnecessary. The requirement for RdRP homologue RRF-1/RDE-9 throughout the *C. elegans* soma — and the similar requirement in plants — also suggests that most RNAi in these systems is driven by secondary siRNAs produced through the action of RdRPs (Simmer et al., 2002).

However, other possibilities also exist. In plants, transitive RNAi spreads in both 3'→5' and 5'→3' directions (Vaistij et al., 2002), which is inconsistent with the simple notion of siRNAs priming dsRNA synthesis. Instead, one can imagine that genomic loci may serve as a reservoir for silencing. In some systems, it is known that exposure to dsRNA can produce alterations in chromatin structure, which could lead to the production of 'aberrant' mRNAs that are substrates for conversion to dsRNA by RdRPs. This model would permit bi-directional spread, as such an expansion of altered chromatin structure is an established phenomenon. Moreover, a similar model could explain co-suppression that is occasionally triggered by single-copy, dispersed transgenes. Finally, this model would be consistent with transitive effects that have been observed for both transcriptional and post-transcriptional silencing in *Drosophila*, which operate in the absence of any homology in the transcribed RNA, and thus differ from 'transitive RNAi' in *C. elegans* (Pal-Bhadra et al., 1999; Pal-Bhadra et al., 2002). However, support for a genome-based amplification model remains elusive, as does the nature of the 'aberrant' RNAs that trigger siRNA formation. Although these models suggest mechanisms for the cell-autonomous amplification of the silencing signal, the character of the signal that transmits systemic silencing in plants and animals is unknown. Two

candidates are siRNAs themselves and long dsRNAs, perhaps formed via the RdRP-dependent amplification.

In plants, there are two types of transmission. The first is short-range, cell-to-cell transmission. Plant cells are intimately connected through cytoplasmic bridges known as plasmodesmata. Movement of RNA and proteins via these cell–cell junctions is well known, and it is likely that either long dsRNA or siRNAs could be passed through these connections. Furthermore, the silencing signal must also be passed over a longer range through the plant vasculature (Voinnet et al., 1998). In this regard, studies of a viral silencing inhibitor, Hc-Pro, have provided evidence against siRNAs being critical for systemic silencing in plants. Hc-Pro expression in a silenced rootstock relieves silencing and inhibits siRNA production, but a systemic signal can still be passed from this rootstock to an engrafted scion lacking Hc-Pro expression (Mallory et al., 2002).

A protein was identified in *C. elegans* that is required for systemic silencing (Winston et al., 2002). The *sid-1* gene encodes a transmembrane protein that may act as a channel for importing of the silencing signal. Expression of *sid-1* is largely lacking from neuronal cells, perhaps explaining initial observations that *C. elegans* neurons were resistant to systemic RNAi. In a later section, this protein will be discussed in detail.

1.2.4 RNAi in Mammalian cells

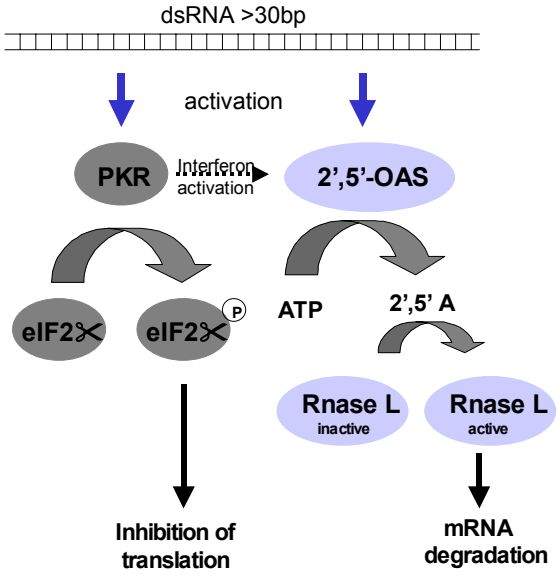

Figure 2. Nonspecific response to dsRNA in mammalian somatic cells

It has been known for years that the exposure of mammalian cells to dsRNA, regardless of its sequence, triggers a global repression of protein synthesis. In most mammalian somatic cells, dsRNA activates protein kinase PKR, which catalyzes phosphorylation of target molecules, such as the translation initiation factor eIF2α, what in turn inhibits translation. PKR is also involved in the regulation of NF-kB, which has a key role in interferon induction. Interferon and dsRNA also activate 2',5'-oligoadenylate (2',5'-OAS) synthetase, leading to the production of 2',5'-oligoadenylates with 5'-terminal triphosphate residues. This subsequently induces activation of general RNA degradation protein RNase L (for review, see Barber, 2001). PKR and 2',5'-OAS are essential for the apoptotic response to dsRNA that has been demonstrated in knock-out mice (Der et al., 1997; Zhou et al., 1997) (Figure 2). This response to dsRNA caused many investigators to doubt that RNAi will function in mammalian systems, following first reports of RNAi in *C. elegans* and *Drosophlia*. However, since long dsRNA do not cause nonspecific effects when injected to oocytes and early embryos, RNAi might operate at these early stages of development. The initial reports about RNAi in mouse indeed showed RNAi in early embryos and oocytes and they demonstrated that it can be efficiently used as a gene knock down tool. (Svoboda et al., 2000b; Wianny and Zernicka-Goetz, 2000). In addition, many important proteins involved in RNAi are present in mammals, especially the dsRNA processing enzyme Dicer. This implied that RNAi might exist in somatic mammalian cells besides the apoptotic response of dsRNA. A deeper understanding of the RNAi mechanism in somatic cells was allowed by bypassing the apoptotic response to dsRNA. siRNAs are too short to trigger the PKR/2',5'OAS pathway in somatic cells but can be used to induce RNAi (Elbashir et al., 2001). Although the RNAi mechanism in mammals has not yet been analyzed in great detail, there is some experimental evidence that the RNAi pathway in mammals is conserved and in principle very similar to that of *Drosophila* and *C. elegans*. Gene silencing by siRNA in mammals, like in other systems, is achieved via sequence-specific mRNA degradation. Biochemical studies were carried using HeLa cell S100 extracts, and a mammalian RISC was isolated from it (Martinez et al., 2002) (mammalian RISC will be discussed in a later section in detail). The human dicer family member is capable of generating siRNA from dsRNA substrates, and Dicer activity has been detected in several cell lines including embryonic carcinoma cells, embryonic stem cells, CHO-K1, mouse embryonic filbroblasts and HeLa cells (Billy et al., 2001; Yang et al., 2001), as well as mouse oocytes and preimplantation embryos. In fact, siRNA can induce an RNAi effect in mammalian cell lines as efficiently as in oocytes and early embryos (Elbashir et al., 2001; Harborth et al., 2001). Although the RNAi response in mammalian

systems appears not as robust as in *C. elegans* or *Drosophila*, it can be efficiently used as a tool to analyze gene function and inactivation method.

1.3 Biological functions of RNAi

The ability of dsRNA to induce gene silencing is a response that has been conserved throughout eukaryotic evolution. This indicates that RNAi might be biologically important. An often proposed function of RNAi is a generalized defence mechanism against unwanted nucleic acids, either in the form of viruses or in the form of parasitic DNA sequences in the genome. Considerable evidence shows RNAi as a protective mechanism against parasitic DNA sequences such as transposons and the RNA sequences of plant viruses. Many genomes contain highly repetitive sequences, including transposons, which normally reside in heterochromatin. Derepression of transposons could also disrupt the heterochromatic state and provide homologous sequences for recombination between non-homologous regions of chromosomes. In this way, transposon activation could result in large scale destabilizaion of the genome. DNA methylation and transcriptional gene silencing are mainly responsible for keeping the transposition frequency at a minimum (for review, see Martienssen and Colot, 2001). However, PTGS/RNAi 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 down-regulate RNAi as well, implicating RNAi in the control of transposons (Ketting et al., 1999). In *Trypanosoma brucei*, siRNA products of an RNA interference event were cloned and sequenced. By sequencing over 1300 siRNA-like fragments, abundant 24-26 nt fragments homologous to the ubiquitous retrotransposon INGI and the site-specific retroposon SLACS were observed (Djikeng et al., 2001).

In plants, PTGS has been widely linked to RNA virus resistance mechanisms. In fact, plant RNA viruses are both inducers and targets for PTGS and gene-silencing-defective mutants of plants show increased sensitivity to viral infection. Evidence also support the view that a dsRNA intermediate in virus replication acts as an efficient initiator of PTGS in natural virus infections (reviewed in Baulcombe, 1999). The decisive support for PTGS as an anti-virual mechanism has come from reports that plant viruses encode proteins that are suppressors of PTGS (Anandalakshmi et al., 1998; Brigneti et al., 1998; Voinnet et al., 2000). These suppressors have evolved to protect viral RNA genomes from the degradative PTGS 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 prepresent a group that is able to activate GFP expression in all tissues of previously silenced GFP-expressing plants. HC-PRO reduces target mRNA degradation and is thus responsible for reduced accumulation of siRNAs (Llave et al., 2000; Mallory et al., 2002). The second type of suppressors includes movement proteins, e.g. p25 of potato virus X, which is involved in curbing the systemic aspect of transgene-induced RNA silencing (Voinnet et al., 2000). The third type includes cytomegalovirus 2b protein, which is involved in systemic signal-mediated RNA silencing (Ding et al., 1996). The cytomegalovirus 2b protein is nucleus localized and also inhibits salicylic acid-mediated virus resistance (Lucy et al., 2000). 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. 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 non-specific RNA degradation and other non-specific responses leading to the blockage of protein synthesis and cell death. Thus, mammals seem to have evolved multiple mechanisms to detect and target dsRNA and to fight viruses.

1.3.1 Gene regulation by miRNA

Apart from the function of RNAi as a defence mechanism, an additional role for RNAi pathways in the normal regulation of endogenous protein-coding genes was suggested through the analysis of animal and plant RNAi mutants. Defects in *C. elegans* RNAi genes *ego1* and *dcl-1* cause specific developmental errors (Grishok et al., 2001; Knight and Bass, 2001) and the Dicer homologue in *Arabidopsis*, CAF1, is required for embryo development (Jacobsen et al., 1999). Mutations in the *Argonaute-1* gene of *Arabidopsis* also cause pleiotropic developmental abnormalities and mutations of this family of gene in *Drosophila* impact normal development. In particular, mutations in the *Drosophila Argonaute-1* have drastic effects on neuronal develoment (Kataoka et al., 2001) and *piwi* mutants have defects in both germline stem-cell proliferation and maintenance (Cox et al., 1998). This genetic evidence illustrates the role of RNAi machinery as a controller of development-related genes. The real excitement came when the link between the two previously known small regulatory RNAs, lin-4 and let-7, and RNAi machinery was discovered. This fueled an already intensive search for endogenously encoded small RNAs that might function via the RNAi pathway. Soon, hundreds of similar small RNAs, named as microRNA (miRNA) were

cloned in various organisms. An evolutionarily conserved system of the RNA-based gene regulation is emerging as a new paradigm for control of gene expression during development.

1.3.1.1 stRNAs in C.elegans

More than a decade ago, two interesting genes, *lin-4* and *let-7* were discovered in *C. elegans* through genetic screens for mutants that lacked the ability to control the timing of specific cell fate switches during development (Lee et al., 1993). When these genes were cloned, they were found to encode two unrelated 21-22 nucleotide RNAs. Since northern blot analysis has shown that these genes are temporally regulated, for example let-7 product starts to accumulate in the L3 stage and then increases to high levels in subsequent L4 and adult stages, they were originally referred to as small temporal RNA (stRNA). Both these RNAs are believed to act by base pairing with mRNA 3' untranslated regions (UTRs) of one or more target genes in the developmental timing pathway. By mechanisms that are not fully understood, this interaction leads to the translation repression of the target genes.

Let-7-like RNAs can be found in diverse animals, including humans, demonstrating that these tiny RNAs are not peculiar to worms. Because they are different from any non-coding RNAs described previously, in their size and activity on mRNAs, lin-4 and let-7 represent a new class of RNAs. stRNAs are 21-22 nucleotides long single stranded RNA, which are processed from longer stem-loop precusor around 70 bp. Although siRNAs are double stranded, the size similarity of siRNAs and stRNAs led the proposition that stRNAs might also be processed by Dicer from its precursor. Indeed, Dicer can generate mature stRNA from synthetic precursors in vitro (Hutvagner et al., 2001; Ketting et al., 2001) and Dicer null worms accumulate both lin-4 and let-7 precursors with a corresponding loss of mature stRNAs (Ketting et al., 2001; Knight and Bass, 2001). Similarly, in HeLa cells, depletion of Dicer by RNAi causes a buildup of pre-let-7 and a loss of mature let-7 (Hutvagner et al., 2001). These results linked the stRNA and RNAi pathway by the central enzyme Dicer. Furthermore, two members of the Argonaute gene family in worms, *ALG-1* and *ALG-2*, are essential for stRNA-mediated regulation. This finding strengthens the hypothesis that stRNAs might act through the RNAi pathway on their regulatory targets (Grishok et al., 2001).

1.3.1.2 Discovery of new gene regulators - miRNAs

Soon after the link between stRNA and RNAi machinery was found, three groups uncovered nearly 100 new similar genes encoding short (21-24 nucleotides) non-coding RNAs, now termed microRNAs (miRNAs) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). One lab came upon them unexpectedly while searching for the endogenous products of the RNAi reaction. The other two groups each suspected that *C. elegans* will have more than the two stRNAs already known to control development. Some of the new miRNA encoding genes are conserved in worms, flies and humans. Because of their sequence diversity, regulated expression, and resemblance to lin-4/let-7, these miRNAs are likely to regulate the expression of protein-encoding genes. More of such RNAs from animals and also plants have been subsequently identified through computational and cloning approaches. The vast majority of the miRNA genes were found at intronic regions or between genes. Occasionally they occur in clusters, some so spaced by close as to suggest that the tandemly arranged miRNAs are processed from a single transcript to allow coordinate regulation.

After these three reports, a fourth group described about 40 more miRNAs, a few of which were identical to those reported in the earlier articles (Mourelatos et al., 2002). These new miRNAs were identified from immunoprecipitations of Gemin3 and Gemin4, two core components of the Survival of Motor Neurons (SMN) complex. The SMN complex has important roles in the assembly/restructuring and function of diverse ribonucleoprotein (RNP) complexes, including spliceosomal small nuclear RNPs (snRNPs) (Fischer et al., 1997; Meister et al., 2001; Pellizzoni et al., 1998), small nucleolar RNPs (snoRNPs) (Jones et al., 2001), heterogeneous nuclear RNPs (hnRNPs) (Mourelatos et al., 2001) and transcriptosomes (Pellizzoni et al., 2001). The link between the SMN complex and siRNA/miRNA-mediated silencing became apparent when it was discovered that components of the SMN complex associate also with miRNAs and a member of the Argonaute family of proteins. These studies add yet another link between RNAi and the endogenous miRNAs.

1.3.1.2.1 Stepwise biogenesis of miRNA

By analogy with stRNAs, it was believed that miRNAs are processed from a stem-loop precursor by Dicer, but it turns out that the miRNA biogenesis is more complex than originally thought. Recently, studies in human cell lines showed that miRNAs are transcribed as long primary transcripts (pri-miRNAs) whose maturation occurs through sequential processing events (Lee et al., 2003; Lee et al., 2002b) (Fig.1). First, in the nucleus, the pri-miRNAs are processed by Drosha,

a member of the RNase III family of enzymes, into a stem-loop of about 70 nucleotides named pre-miRNAs. The pre-miRNAs are exported to the cytoplasm by Exportin-5 (Exp5), which binds correctly processed pre-miRNAs directly and specifically (Lund et al., 2004; Yi et al., 2003). Secondly, in the cytoplasm, Dicer processes the pre-miRNAs into mature miRNAs. Since both RNase III enzymes are found in *C. elegans, Drosophila*, mice and humans, the stepwise processing of miRNAs is likely to be conserved, at least in animals. Stepwise processing might be beneficial in terms of efficiency and accuracy of processing. In addition, stepwise processing and compartmentalization might allow a fine regulation of miRNA biogenesis at multiple steps.

1.3.1.2.2. Structure of miRNAs

Although miRNAs are derived from dsRNA hairpin precursors, typically only a single strand of the precursor stem, corresponding to mature miRNA accumulates in the cell. The pre-miRNAs appear to be processed as an approximately 70-nucleotide precursor hairpins containing a 4-15 nucleotide loop. Sometimes the 21-23 nucleotide miRNA forms a perfect duplex within the hairpin, but more often, multiple bulges disrupt the perfect 21-23 nucleotide duplex. Similar to other RNase III enzymes, Dicer can process complex hairpin structures that can contain multiple mismatches in the helical stem (Hutvagner et al., 2001). Little is known about the structural determinants necessary for processing of miRNAs into approximately 21-nucleotide RNAs. Studies performed on RNase III enzymes in other organisms have shown that dsRNA cleavage relies on antideterminants in the double-stranded stem (bacterial RNase III), or sequence determinants in the terminal loop of the stem-loop RNA structure (yeast RNase III) (Chanfreau et al., 2000; Wu et al., 2001; Zhang et al., 1997). However, comparison of many miRNAs that are likely to be processed by Dicer did not so far reveal any obvious features that might guide Dicer recognition or processing.

1.3.1.3 Targets of miRNAs

Significantly, when the miRNA is base-paired to the mRNA, it does not exhibit perfect complementarity: this situation is in contrast to siRNA-mediated degradation (Hutvagner and Zamore, 2002). In the existing models for the base-paring, typically 50-85% of the miRNA residues are base-paired to the mRNA 3' UTR. Because the complimentarity between miRNAs and target

mRNAs is not perfect, it is difficult to predict candidate targets using bioinformatics searches. However, computational methods were developed by various groups to predict miRNA targets in *Drosophila* and mammals. Several hundreds targets were predicted and some of them were also confirmed experimentally. In both *Drosophila* and mammals, predicted target genes were enriched in transcription factor but also encompassed broad range of other functional gene classes (Enright et al., 2003; Lewis et al., 2003; Stark et al., 2003). In *Drosophila*, two 3' UTR sequence motifs, the K box (cUGUGAUa) and the Brd box (AGCUUUA), have been implicated as miRNA target. They are partially complementary to several reported miRNAs and, in addition, these sequences are known to mediate negative posttranscriptional regulation (Lai et al., 1998; Lai and Posakony, 1997). It is interesting to note that among the miRNAs and their known cognate targets, only the 5'-most stretch of miRNA nucleotides is usually perfectly paired to the target sequence..

1.3.1.3.1 miRNAs in animals

Estimates place the total number of distinct *C. elegans* and vertebrate miRNA genes at about 150 and 250, respectively (Ambros et al., 2003; Lim et al., 2003a; Lim et al., 2003b), indicating that miRNAs are a major class of regulatory molecule in animals. About 30% of the *C. elegans* miRNAs are close in sequence to insect and/or vertebrate miRNA, suggesting that a large fraction of miRNAs could play evolutionarily conserved developmental or physiological roles.

In the case for the miRNAs lin-4 and let-7 (Lee et al., 1993; Reinhart et al., 2000), up-regulation of lin-4 RNA in the second larval stage represses the expression of LIN-14 and LIN-28, two key regulators of early larval developmental transitions in *C. elegans* (Figure 3). The role for lin-4 and let-7 as temporal regulators of development in other animals is supported by the phylogenetic conservation of their temporal patterns of up-regulation. In *Drosophila*, let-7 and the lin-4 homolog mir-125 are up-regulated in concert at the onset of metamorphosis (Bashirullah et al., 2003; Lagos-Quintana et al., 2002; Sempere et al., 2003). In *Drosophila* and some vertebrates, the let-7 and mir-125/lin-4 genes are closely linked and therefore may be co-regulated. Perhaps the distinct roles for lin-4 and let-7 at different developmental stages in *C. elegans* represent an adaptation of a more widely conserved collaboration between these two miRNAs. Like worm LIN-28 protein, vertebrate LIN-28 homologues are also down-regulated during development. Moreover, the mouse and human lin-28 3' UTRs contain predicted lin-4 complementary sites, suggesting that the lin-4/lin28 regulatory relationship may also be conserved (Moss and Tang, 2003).

Recent screens for *Drosophila* mutants that exhibit growth defects identified the *bantam* locus, which encodes a miRNA that functions to repress apoptosis and promote cell proliferation in the developing fly (Brennecke et al., 2003). Bantam miRNA seems to be expressed broadly, and represses the translation of the mRNA for Hid, a key activator of programmed cell death. Bantam is related to mir-80-82 of *C. elegans*, suggesting that the mir-80 family might control developmental cell death and/or cell proliferation in the worm (Figure 3). In a screen for Drosophila genes that oppose the cell death activator Reaper, mutations in the mir-14 miRNA gene were identified that also affected aspects of fat metabolism (Xu et al., 2003). Given the importance of these cellular processes in animals, and the conservation of a large proportion of miRNAs across species boundaries, miRNAs are likely to have broad significance in a wide range of developmental processes in animals.

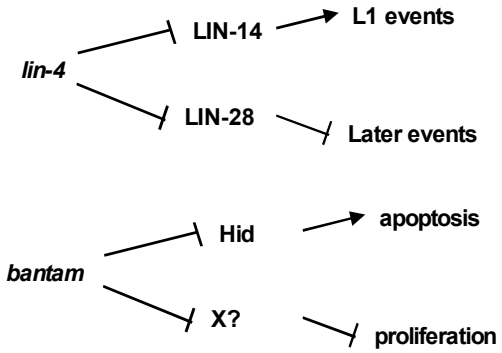

Figure 3. Proposed developmental roles for *lin-4* and *bantam* in *C. elegans* and *Drosophlia*, respectively. *Lin-4* miRNA expression at the end of the worm L1 larval stage results in down-regulation of LIN-14 and LIN-28 protein synthesis, controlling the transition from L1 to later developmental events. *Bantam* miRNA in cells of the fly larva acts through the repression of Hid, and probably other targets, to control apoptosis and cell proliferation, respectively.

1.3.1.3.2 miRNAs in Plants

The search for miRNA was also extended to plants. There are hundreds of plant miRNAs identified by various groups (Llave et al., 2002; Park et al., 2002; Reinhart and Bartel, 2002). Detailed analysis of a subset of these small single-stranded RNAs showed that they have all of the hallmarks of miRNAs: they range in size from 20 to 24 nt and are derived from the stem region of endogenously encoded stem-loop structures, generally larger and more complex than the pre-

miRNAs observed in animals. Moreover, their production requires the activity of the the *Arabidopsis* Dicer homolog, DCL1. Candidate target genes had also been identified in plants. In contrast to *C. elegans*, the plant mRNAs generally contained a single target site, often located in the open reading frame, with a high degree, in not perfect, complementarity to the miRNA. This led to the proposal that plant miRNAs mediate an siRNA-like target cleavage, rather than translational inhibition (Tang et al., 2003).

Now, miRNA-mediated transcript cleavage in plants has been demonstrated. The Carrington group examined the *Arabidopsis* mirRNA29 having perfect complementarity to the targets from the SCAECROW family of transcription factors and also five other miRNAs (Kasschau et al., 2003). They detected mRNA cleavage products and determined that the 5' end of these fragments corresponded to the site of miRNA:mRNA complementarity. Their results also showed that miRNA-directed target cleavage can tolerate some mismatches. The Zamore group also found an endogenous miRNA that cleaved the *PHAVOLUTA (PHV)* and *PHABULOSA (PHB)* transcripts (Tang et al., 2003). Interestingly, the miRNA did not cleave *PHV* contaning a single point mutation, and the same sequence change is responsible for a dominant *phv* mutation in *Arabidopsis*.

Demonstration of the role of miRNAs in plants can also come through another route – analysis of genes required for miRNA formation or activity. Null alleles of *dcl1* result in arrest before the heart stage of embryogenesis and cause overproliferation of cells in the suspensor. A hypomorphic allele that removes the carboxy-terminal dsRNA-binding domain of Dicer causes narrow, occasionally filamentous, leaves and floral organs, and a loss of determinacy in the central region of the floral meristem (Schauer et al., 2002)). The *Arabidopsis AGO1* gene is required for, among other events, establishing polarity in lateral organs. *Ago1* mutant plants have pointed, unexpanded cotyledons, narrow rosette leaves, radicalized cauline leaves, narrow sepals and petals, and unfused carpels; similar to *dcl1* mutants, they lack axillary meristems. Hypomorphic *ago1* alleles have serrate leaves and delayed flowering (Bohmert et al., 1998a).

1.3.1.4 Shared pathway of siRNAs and miRNAs

Hela cell S100 extracts which recapitulate siRNA-directed target cleavage in vitro were used to dissect mammalian RNAi pathway. Using these extracts, several components of the mammalian RISC have been identified, including the Argonaute homologs eIF2C1 and eIF2C2 , the putative RNA helicase Gemin3, and Gemin4, a protein of unknown function (Hutvagner and Zamore, 2002;

Martinez et al., 2002). Like the RISC detected in *Drosophila* embryo lysates, the mammalian RISC mediates cleavage of the target RNA near the center of the anitsense siRNA strand. The mammalian RISC is similar to the miRNP complex identified by Dreyfuss and colleagues and mentioned earlier (Mourelatos et al., 2002).

The demonstration that an endogenous miRNA was a component of a functional RISC led to the proposal that only a single complex exists to mediate the diverse functions of small RNAs (Hutvagner and Zamore, 2002) (Fig.1). This model proposes that small RNA functional diversity is achieved not by multiple effector complexes, but by a single complex that can carry out at least two types of posttranscriptional regulation: target cleavage and translational repression. Which type of regulation occurs is proposed to be determined solely by the degree of complementarity between the small RNA guide and its target. Additional support from this model comes from studies which show that siRNA can mediate translational control when cleavage is blocked by a lack of complemenarity with the target RNA at the center of the siRNA guide (Doench et al., 2003).

Recently, an additional complex similar in composition to that of Drosophila RISC has been identified in mammalian cells in which RNAi has been initiated by transfection with siRNAs (Caudy et al., 2003). This complex includes siRNAs and miRNAs, an Argonaute family member (eIF2C1/hAgo-1), the mammalian homolog of VIG, the Fragile X mental retardation protein, and the mammalian homolog of the RISC-associated micrococcal nuclease family member, named TSN. This complex is present in low amounts in naive cells, but it can be induced to assemble by transfection with siRNAs. Furthermore, it is related in composition to a similar complex that cofrationates with siRNAs and participates in miRNA-mediated repression in *C. elegans* (Caudy et al., 2003). In RNP complexes purified from adult worms and eggs, miRNAs were present in the complexes that also contained the *C. elegans* homologs of VIG and the nuclease TSN. In addition, depletion of VIG and TSN by RNAi prevents the proper down-regulation of a lin-41 3'UTR reporter gene that normally occurs at the L4 to adult transition. This lin-41 downregulation is dependent on a proper function of let-7, suggesting that VIG and TSN proteins not only have a role in mRNA degradation as found in *Drosophlia*, but also are important for the function of the let-7 miRNA pathway. As an additional example of miRNAs associated with proteins originally identified in degradation complexes, let-7 is present in immunoprecipitates from human cells using antibodies against PAI-RBPI, the human VIG homolog and p-100, the TSN homolog. These immunoprecipitates also contain FMRP, suggesting that complexes present in human cells are similar to those identified in fly cells. The precise relationship of this complex to the Gem3/Gem4-

containing particle previously characterized in HeLa cells remains unclear. However, it is possible that these are the same or closely related complexes that differ only in their association with accessory factors.

1.3.2 Other biological pathways involving small RNAs

In addition to all biological functions of RNAi already discussed above, stunning new discoveries of more biological pathways involving small RNAs have been reported in the span of the last two years. It starts to be clear that the small RNAs work not only at the posttranscriptional stage but also leave their marks on the genomes to repress the gene transcription activity or selectively remove portions of the genomes, especially of protozoans. Broadly speaking, the siRNAs have different biochemical effects on the chromatin: DNA methylation, as revealed mostly in plant systems; heterochromatin formation; and programmed elimination of DNA. 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.

1.3.2.1 RNA-dependent DNA methylation

A role for RNA in guiding de novo cytosine methylation of homologous DNA sequences was first discovered in viroid infected plants and subsequently also in non-infected plants systems(Sijen et al., 2001b). 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 and the corresponding part of the genome, especially the promoter region might become transcriptionally silent. The initiator of the RNA-dependent DNA methylation transcription gene silencing (TGS) could be either the transgene-derived dsRNA or the consequent siRNA (Jones et al., 2001; Jones et al., 1999; Vaistij et al., 2002). Depending on the sequence information of the dsRNA, RNA-dependent DNA methylation was found to occur at the open reading frame or the promoter region of the genome (Aufsatz et al., 2002; Matzke et al., 2001). 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 (Hannon, 2002). In the events of RNA-dependent DNA methylation, the chromodomain containing DNA methylases act either alone or in combination with other proteins, such as piwi domain-containing

proteins, to form complexes with the siRNAs and cause sequence-specific RNA-dependent DNA methylation, finally resulting in TGS (Aufsatz et al., 2002).

Evidence of cross talk between PTGS and TGS has been obtained from the mutational analysis of *Arabidopsis* and *Drosophila*. Two types of Arabidopsis mutants, *ddm1* and *met1* 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 (Vaucheret and Fagard, 2001). Although the patterns of reduction are different with these mutants, their studies highlight the strong correlation between PTGS and TGS. In *Drosophila*, polycomb protein-dependent TGS is also affected by mutations in PIWI, family of proteins required for RNAi (Pal-Bhadra et al., 2002). Other evidence includes the *Argonaute4* gene of *Arabidopsis*, which controls both locus-specific siRNA accumulation and DNA methylation (Zilberman et al., 2003); the *Arabidopsis* SDE4 locus, which is of unknown biochemical function but is responsible for retroelement TS SINE-specific siRNA formation (Hamilton et al., 2002); and the *Arabidopsis rts1* mutation, which causes about 50% reduction in target promoter DNA methylation. However, not all TGS mutations affect the PTGS pathways and vice versa, suggesting that the two pathways diverge at some point (Vaucheret and Fagard, 2001). RNA-dependent DNA methylation has been reported only in plants until now. It is unknown whether it also occurs in animals.

1.3.2.2 Heterochromatin silencing

Generally, in eukaryotic systems, histone modifications make the chromatin structure inert to transcription by heterochromatin formation. In almost all organisms heterochromatin formation requires that histone H3 of the chromatin to 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 bidning of the chromodomain containing HP1 to histone H3 methylated on Lys9 is highly specific and of very high affinity (Bannister et al., 2001; for review, see Richards and Elgin, 2002). 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

over a large distance due to cooperative protein-protein interactions of chromatin-remodeling factors, the components of which have not been fully identified yet. However, heterochromatic formation is initiated at places containing repeated DNA sequences.

Recent research of RNAi mechanism in *Schizosaccharomyces. pombe* surprisingly revealed the connection between RNAi and heterochromatin formation (Reinhart and Bartel, 2002). A Dicer and an Argonaute homolog are found in the genome of *S. pombe*, implying that siRNAs, miRNAs, or another class of small RNAs might play an important role in fission yeast. Like in other organisms, endogenous small RNAs with the features of Dicer cleavage products, i.e.,~22-nt RNAs with 5'-phosphate and 3'-hydroxyl groups were cloned from *S. pombe*. Surprisingly, small RNA matching the *S. pombe* centromeric repeats were found. The majority of the centromeric RNAs are from the dh repeat, an element that can confer heterochromatic silencing on another locus and is sufficient for centromere function along with the centromeric central core. These small RNAs do not appear to be miRNAs since transcription of their genomic sequences would not produce step-loop structures akin to those of the miRNA precursors. Instead, the small RNAs are reminiscent of siRNAs, corresponding to transcripts generated from both DNA strands of the repeat region.

Involvement of RNAi machinery in heterochromatic silencing was further demonstrated by the experiments using deletion of Dicer, argonaute, and RNA-dependent RNA polymerase genes in *S. pombe*. These deletions result in the aberrant accumulation of complementary transcripts from centromeric heterochromatic repeats. This is accompanied by transcriptional de-repression of transgenes integrated at the centromere, loss of histone H3 lysine 9 methylation, and impairment of centromere function. Centromeric repeats that are transcribed at low levels and produce dsRNA are sufficient to induce heterochromatin formation at an ectopic site in *S. pombe*, and this recruitment of repressive chromatin is strictly dependent on the RNAi machinery (Volpe et al., 2002). A mechanism by which repeated sequences and RNAi trigger silent chromatin assembly is unknown. Possibly, it involves base-pairing of homologous DNA sequences. The connection between RNAi and heterochromatin assembly has suggested a model for the RNA-mediated epigenetic structuring of eukaryotic genomes. Double-stranded RNA is believed to be processed into small RNAs, which in turn provide specificity for targeting histone-modifying activities and epigenetic modification of the genome through homology recognition.

1.3.2.3 DNA elimination

The most dramatic effect of siRNA-mediated heterochromatin formation followed by chromosomal DNA elimination and rearragnement has been recorded in the ciliated protozoan *T. pyriformis* (Mochizuki et al., 2002; Taverna et al., 2002). Among unicellular organisms, *T. pyriformis* is unique because of its nuclear dimorphism. The two nuclei, the micronucleus and macronucleus, serve different functions. During conjugation, the micronucleus gives rise to the macronucleus, and this transition is accompanied by chromosome rearrangements in which specific regions of DNA are eliminated. A gene *TWI1*, the *T. pyriformis* homologe of *piwi* is required for DNA elimination. Small RNAs were also found to be specifically expressed prior to chromosome rearrangements during conjugation, and these RNAs were not observed in TWI1 knockout cells. These results suggested that these small RNAs might function to specify sequences to be eliminated by a mechanism similar to RNAi.

These *S. pombe* as well as *T. pyriformis* data show how dramatic the epigenetic consequences for the genome could be following the formation of siRNA molecules in cells. The RNAi/miRNA machinery is reported to control many important features of cellular biology, namely stem cell maintenance (Cox et al., 1998), cell fate determination (Bohmert et al., 1998b), nonrandom chromosome segregation (Schmidt et al., 1999). It is not difficult to imagine that we might witness RNAi related signals also participate in other chromosomal functions.

1.4 Proteins involved in RNAi

1.4.1 Dicer

Among all the proteins identified as involved in RNAi, Dicer is the most important. It occupies the central position in the siRNA and miRNA pathways by generating the effector molecules. Dicer belongs to the RNase III ribonuclease family. These nucleases cut specificly dsRNA and generate dsRNA products with 5’-phosphate and 3'-hydroxyl groups, and two nucleotide 3’ overhang termini. This enzyme family can be subdivided to three classes, based upon domain structure (Figure 4). Bacterial RNase III, representing sub-family I, contains a single catalytic domain and a dsRNA-binding domain (dsRBD). The *E. coil* RNase III promotes maturation of ribosomal RNAs (rRNAs), tRNAs, and mRNAs, and can also initiate mRNA degradation. Drosha nuclease sub-family contains two catalytic and dsRBD domains and N-terminal Pro-rich and RS-rich domains (Drosha will be discussed in the next section). A third sub-family, which also contains two catalytic domains

and, in addition, helicase, PAZ, and DUF283 motifs, is the Dicer family. As discussed earlier, Dicer is involved in the initiation step of RNAi, generating siRNAs or miRNAs, and the two RNase III domains are mainly responsible for the cleavage activity.

Dicer is evolutionarily conserved and exists in all eukaryotes studied to date except budding yeast. There is one gene encoding Dicer in human, mouse, and *C. elegans*, two in *Drosophila*, and at least four in *Arabidopsis*. It appears that Dicers from different organisms and also different Dicers in the same organism might function differently. In *Drosophila*, Dicer shows ATP dependence for dsRNA cleavage (Bernstein et al., 2001), which is rather unique for the RNase III family enzymes. However, we have shown that mammalian Dicer cleaves dsRNA without ATP dependence (Zhang et al., 2002). In Drosophila S2 cell extracts, siRNA production is associated most strongly with one of the two *Drosophila* Dicers (Dicer-2) but not the other (Dicer-1) (Liu et al., 2003). In plants, there are two classes of siRNAs generated by Dicer, short ones of about 21-22 nts and longer ones which are about 24-25 nts (Hamilton et al., 2002). It is assumed that different Dicers generate these different length products.

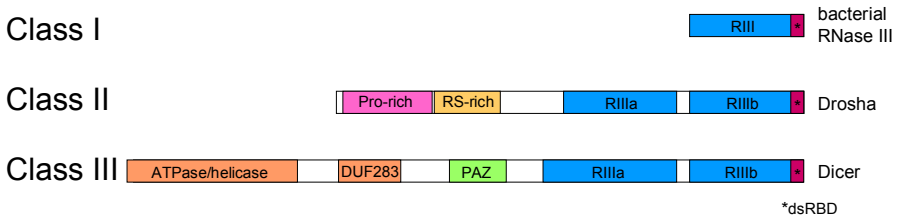

Figure 4 . Schemes of the RNase III superfamily proteins. Three classes of RNase III family proteins represented by human Dicer, human Drosha and eubacterial RNase III. Individual protein domains are indicated in different colors.

Besides RNase III domains, the function of other domains of Dicer is not well known. There is still no demonstration helicase activity in Dicer preparations. It is possible that this activity is involved in unwinding of the siRNA and transfer of its single strand to RISC (Schwarz et al., 2003). The PAZ domain was originally assumed to be responsible for a protein-protein interaction with Argonaute proteins, which also contain this domain. However, our data showed that these are the

PIWI domain of AGO and RNase III domain of Dicer, which are responsible for this interaction. The structure of the PAZ domain from Argonaute proteins in *Drosophila* was recently solved, and it revealed that PAZ is a nucleic acid binding domain (Lingel et al., 2003; Song et al., 2003b; Yan et al., 2003) (the PAZ domain structure will be discussed in a later section about Argonaute proteins, and in Chapter 3). Bioinformatic studies have revealed that some Dicer molecules have extremely divergent PAZ domains or are even devoid of it, e.g. the *Arabidopsis* DCL4. More recently, we have prepared mutants in the PAZ domain of human Dicer and found that this domain is involved in the dsRNA processing (Chapter 3). Dicer knock out mouse has recently been generated, and shown to be embryonic lethal, probably due to its involvement in biogenesis of miRNAs, which regulate development (Bernstein et al., 2003).

Recently, the crystal structure of the RNase III catalytic domain has been established and a model of the dsRNAs cleavage proposed (Blaszczyk et al., 2001). In this model, the enzyme functions as a dimer and the two antiparallel RNase III catalytic domains contribute to two compound catalytic centers. Sequences of one of the Dicer RNase III domains reveal deviations from the catalytic consensus in the bacterial RNases III. Since introduction of these alterations into bacterial RNase III resulted in defects in the enzyme function, the catalytic centre of the second RNase III domain of Dicer might be non-functional. Based on the bacterial RNase III structure and a model of its function, the antiparallel alignment of Dicer RNase III domains on a dsRNA substrate was proposed to produce four compound catalytic centers, with two of them being inactive due to the alterations of essential amino acids. In this way, cleavage would occur at 22-bp intervals rather than 10-bp as in the bacterial enzyme. This seemingly right model was challenged by mutagenesis studies of the recombinant Dicer, which are presented in this thesis Chapter 3. Based on our work, a single compond catalytic center model is proposed for both bacterial RNase III and human Dicer (see Chapter 3 and figures within).

1.4.2 Drosha (mammalian RNase III)

As mentioned above, members of the second class of RNase III proteins, comprised of Drosha and homologs, contain two RNase III domains, a dsRBD, and a long N-terminal segment. The N-terminal part of human and mouse Droshas contains two domains thought to be involved in protein-protein interactions, namely a proline rich region (PRR) and a serine-arginine-rich (RS) domain. RS domains are commonly found in RNA metabolism/splicing factors (Graveley and

Maniatis, 1998). Drosha homologs have been identified in flies, worms, humans and mice, but not yet in plants. The human Drosha has been reported to function in processing of highly structured ribosomal RNA precursors, much like the *E. coli* RNase III. The degree to which a panel of antisense RNAs against human Drosha decreased its expression correlated very well with an increase in the 32S and 12S ribosomal precursors (Wu et al., 2000). As mentioned earlier, Drosha has emerged as an important player in the maturation of miRNAs. MiRNAs are transcribed as long primary miRNAs (pri-miRNA). Drosha clips pri-miRNA in the nucleus into shorter, ~70-nt pre-miRNAs. The pre-miRNAs are then exported from the nucleus and processed by Dicer into mature ~22-nt miRNAs (Lee et al., 2003). Thus, Drosha processes two distinct types of substrates with hairpin-like secondary structures: pre-rRNAs and pri-miRNAs.

1.4.3 Argonaute family

Argonaute proteins were first identified in *Arabidopsis* mutants that produced altered leaf morphology. They constitute a large, evolutionarily conserved gene family (Bohmert et al., 1998a). Argonaute proteins are ~100-kD highly basic proteins characterized by the presence of two homology regions, the PAZ domain and the Piwi domain, the latter being unique to this group of proteins. Argonaute proteins can be separated, based on their sequence, into two subclasses: those that resemble *Arabidopsis* AGO1, and those that more closely resemble *Drosophila* piwi.

Argonaute family genes have been isolated from several organisms in screens for mutants that are deficient in RNAi and related phenomena. These genes include *C. elegans rde-1* (Tabara et al., 1999), *Arabidopsis AGO1* (Bohmert et al., 1998a) and *Neurospora QDE2* (Fagard et al., 2000). Disruption of the *S. pombe* single Argonaute gene leads to the loss of silencing of heterochromatic repeats at the centromere (Reinhart and Bartel, 2002). The *Drosophila piwi* mutation also affects transcriptional gene silencing (Pal-Bhadra et al., 2002). In addition, a *Tetrahymena* piwi-related gene, *TWI1*, is required for DNA elimination (Mochizuki et al., 2002). Still other Argonaute proteins have been implicated in developmental control, and recent studies suggest that silencing and developmental functions may be linked. However, in some cases, genetic studies have indicated exclusive roles in either silencing or developmental control, keeping open the possibility that Argonaute proteins may exert some of their biological functions through processes unrelated to RNAi. Mutants of the best characterized *C. elegans* Argonaute gene, *rde-1*, are strongly resistant to RNAi but are developmentally normal (Tabara et al., 1999). Two other family members, *alg-1*

and *alg-2* functionally overlap and show strong developmental phenotypes, but are dispensable for RNAi in the soma (Grishok et al., 2001).

Drosophila contains four characterized Argonaute proteins (piwi, aubergine/sting, dAgo1 and dAgo2) plus one predicted from the genomic DNA (dAgo3). Piwi, aubergine, dAgo1 and dAgo2 have been implicated in RNAi-like silencing phenomena. dAgo1 was shown to be required for efficient RNAi in *Drosophila* embryos (Williams and Rubin, 2002). Piwi has been shown to be necessary for PTGS and some aspects of transcriptional gene silencing (Pal-Bhadra et al., 2002). Strains with mutations in piwi, aubergine/sting, and Ago1 also exhibit developmental phenotypes (Kennerdell et al., 2002; Morel et al., 2002; Smulders-Srinivasan and Lin, 2003). The fourth Argonaute protein in *Drosophila*, dAgo2, has been shown to be necessary for RNAi as a component of the RISC complex (Hammond et al., 2001a). No dAgo2 mutants have yet been reported, so it remains unknown whether dAgo2 is involved in the developmental control.

The first mammalian Argonaute protein to be studied in any detail was the rabbit eIF2C, which was identified as a component of a protein fraction that enhances translation (Cikaluk et al., 1999). However, it should be noted that no solid evidence exists for this protein playing a role in translation. There are 7 Argonaute proteins in human and eight in mouse. Both mouse and human have three proteins that fall into the Piwi subfamily. Humans have four genes that fall into the *Ago1* subfamily. Mouse has one additional *Ago1* family gene, which may have an orthologous human gene that has not yet become apparent in the current human gene register. Human EIF2c1/hAgo1, also known as GERp95 for its association with Golgi and ER, was also identified in a screen for genes involved in Wilms' tumors. Human EIF2C2/hAgo2 was shown to reside in a complex with Gemin3, an RNA helicase that joins the growing number of helicases implicated in RNAi, and Gemin4, a protein of unknown function (Mourelatos et al., 2002).

The structure of PAZ domain of Argonaute family proteins has been recently solved. Two groups solved the PAZ domain structure of *Drosophlia* Ago1 and Ago2 by NMR (Lingel et al., 2003; Yan et al., 2003) and one by the crystallization of Ago2 (Song et al., 2003b). The PAZ domain contains a variant of the OB (oligonucleotide and oligosaccharide-binding)-fold, a structure that is implicated most frequently in single-stranded nucleicacid binding. To investigate whether the PAZ domain binds nucleic acid directly, all three groups studied its binding to single- and double-stranded siRNAs and DNA. Although there is some discrepancy about the species of nucleic acid that preferentially binds to PAZ domains, it is clear that all observed interactions are of low affinity, with dissociation constants in the micromolar range. The study of Ago1 reported that the PAZ

domain interacts with RNA preferentially over DNA (Lingel et al., 2003; Yan et al., 2003). One Ago2 study agreed with this finding (Song et al., 2003b), while the other finds that both single and double-stranded DNA bind with the affinity equal to RNA (Yan et al., 2003). Although none of the solved structures contained RNA, the binding site is predicted to involve highly conserved residues in an inter-subdomain cleft. Mutational studies confirmed that this cleft contains aromatic residues critical for RNA binding. All three studies demonstrated that binding is independent of nucleic acid sequence. Because critical residues and the binding surface are conserved, it is possible that PAZ domains of all members of the Argonaute and Dicer family proteins adopt a similar fold with a nucleic acid binding function.

The binding assay and UV cross-linking experiments of the Ago PAZ domain with siRNA supported the notion that PAZ domain could be functioning in end-recognition the 3'-overhang of siRNA. The apparent end-sensing ability of the Ago PAZ domain led to the speculation that the PAZ domain of Dicer could function in end-recognition through a direct interaction with one or more characteristic terminal structures of siRNA or dsRNA. The PAZ domain is found only in Dicer and Argonaute proteins. These proteins function at two steps that must distinguish genuine siRNAs from other RNAs in the cell. Although RISC contains single stranded siRNAs, it has not yet been determined whether RISC also transiently associates with double stranded siRNAs. In either case, some components of RISC must be able to detect the chemical structure of bona fide intermediaries in the pathway. The PAZ-containing Argonaute proteins, shown to be core components of RISC complexes, are suitable candidates for this function.

1.4.4 R2D2 and Rde-4

R2D2 is a very recently identified protein which links the initiation and effector phases of RNAi in *Drosophila* S2 cell extracts (Liu et al., 2003). This protein cofractionated with the dsRNA processing activity from S2 cell extracts through multiple chromatographic steps. R2D2 is a 36 kDa protein with two dsRNA binding domains. It forms a stable complex with Dicer-2 in S2 cells. Significantly, R2D2 displays 33% sequence similarity to the Dicer-associated *C. elegans* protein RDE-4, which was originally identified genetically as an RNAi factor that is important for the initiation but not the maintenance of RNAi.

Purified recombinant Dicer-2 catalyzed siRNA production with equal efficiency in the presence or absence of R2D2. In contrast, gel-shift and UV crosslinking assays revealed that the Dicer-

2/R2D2 complex but not Dicer-2 alone associates stably with siRNAs; furthermore, siRNA binding by the Dicer-2/R2D2 complex was nearly abolished by mutation in the dsRNA binding domains of R2D2. The apparent role of R2D2 in siRNA retention after processing provided the first indication that R2D2 controls siRNA fate rather than its production, prompting the idea that R2D2 participates in channeling siRNAs into the RISC complex. This was further proven by the effect of Dicer-2/R2D2 complex on the effector step of RNAi. The siRNA production machinery could be crudely but effectively separated from RISC complexes by the polyethylene glycol (PEG) precipitation. Addition of Dicer-2 alone to the RISC-containing PEG supernatant stimulated only weakly the RISC activity in response to a long dsRNA trigger. However strong stimulation was observed upon addition of the Dicer-2/R2D2 complex. Mutation of the R2D2 dsRNA binding domains abolished the stimulation of RISC activity. Importantly, similar effects were observed when pre-cleaved siRNAs were used to trigger RISC activity, indicating that RISC does not efficiently utilize siRNAs in the absence of R2D2. Ago2, the RISC component, is not efficiently coselected with biotinylated siRNAs unless functional Dicer-2/R2D2 complex is present. The clear implication of these findings is that R2D2 helps to mediate the transition between initiation and effector phases of RNAi.

Although R2D2 is similar to RDE-4, their roles may be slightly different. Rde-4 is found in a complex with Dicer1, Argonaute protein Rde-1, and an additional DexH-box helicase. The proposed function of Rde-4 in *C. elegans* is the recognition and presentation of exogenous dsRNA triggers to Dicer. Rde-4 seems to interact only with foreign trigger dsRNA, not endogenous miRNAs or siRNAs. Mutations in Rde-4 substantially reduced the amount of siRNAs derived from an injected dsRNA trigger. This defect in RNAi could be partially bypassed by a direct injection of siRNAs (Parrish and Fire, 2001).

1.4.5 RNA dependent RNA polymerase

In both plants and *C. elegans*, RNAi/PTGS requires proteins similar in sequence to a tomato RNA-directed RNA polymerase (RdRP) (Schiebel et al., 1998). In *Arabidopsis*, RdRP homologue SDE1/SGS2 is required for transgene silencing, but not for virally induced gene silencing (Dalmay et al., 2000; Mourrain et al., 2000b). This may suggest that SDE1/SGS2 act as an RdRP, since viral replicases could substitute for this function in VIGS. In *Neurospora*, RdRP homologues QDE-1 is required for efficient quelling (Cogoni and Macino, 1999). EGO-1, one of the *C. elegans* RdRP, is essential for RNAi in the germline of the worm (Smardon et al., 2000), and another RdRP

homologue, RRF-1/RDE-9, is required for silencing in the soma. All RdRP proteins could be involved in amplifying the RNAi signal. However, only the tomato and *Neurospora* enzymes have been demonstrated to posses RNA polymerase activity, and biochemical studies are required to establish definitively the role of these proteins in RNAi (Cogoni and Macino, 1999a). These genetic studies have led to a model for transitive RNAi in which siRNAs might prime the synthesis of additional dsRNA by RdRPs. Although RdRP activity has been reported in *Drosophila* embryo extracts (Lipardi et al., 2001), there is no obvious RdRP homolog found in the fly genome. Additionally, numerous experiments demonstrated that there is no transitive RNAi in flies and RdRP is not required for RNAi in *Drosophila* extracts (Roignant et al., 2003; Schwarz et al., 2002).

1.4.6 RNA and DNA helicases

Through genetic studies in various organisms, many RNAi related proteins are identified as DNA or RNA helicases. In *Chlamydomonas. reinhardtii* , a RNAi-resistant mutants, *mut6*, was isolated and the gene encodes a protein which is a memeber of the DEAH box RNA helicase family protein (Wu-Scharf et al., 2000). The quelling-defective mutant in *Neurospora*, *qde3*, was cloned and the sequence encodes a 1,955-amino acid protein. This protein shows homology with the family of RecQ DNA helicases, which includes the human proteins for Bloom syndrome and Werner syndrome (Cogoni and Macino, 1997). In *C. elegans*, two mutants involved in RNAi are also revealed to have 3'-5' exonuclease domain (*mut7*) (Tijsterman et al., 2002) or ATPase with RNA-binding and helicase activities (*smg2)* (Domeier et al., 2000). Helicases also had been identified in plants (*sde3* in *Arabidopsis*) (Dalmay et al., 2000). Although possible roles in RNAi for some of these protiens were proposed, e.g. MUT6 might involved in the degradation of misprocessed aberrant RNAs (Wu-Scharf et al., 2000), their functions are mostly unknown and further biochemical experiments are needed to reveal their exact roles in RNAi.

1.4.7 FMRP

Fragile X syndrome is a common form of inherited mental retardation caused by the loss of *FMR1* expression. The *FMR1* gene encodes an RNA-binding protein (FMRP, Fragile X mental retardation protein) that associates with translating ribosomes and acts as a negative translational regulator. In *Drosophila*, the dFXR (*Drosophila* Fragile-X-related), the fly homolog of FMRP, binds

to and represses the translation of an mRNA encoding the microtuble-associated protein Futsch. In Drosophlia, two labs independently discovered the link between RNAi and dFXR. One group discovered dFXR in association with the purified RISC complexes (Caudy et al., 2002), whereas the second group identified Ago-2 in attempts to purify proteins that associate with endogenous dFXR (Ishizuka et al., 2002). Consistent with this reciprocal interaction, antibodies to dFXR immunoprecipitate RISC activity (Caudy et al., 2002). In addition to Ago-2, Ishizuka et al. (2002) found that dFXR co-purified with the *Drosophlia* homolog of the p68 RNA helicase (Dmp68), two ribosomal proteins, L5 and L11, along with 5S ribosomal RNA. Both RISC and the dFXR protein bind RNA, but the complexes were RNase resistant, suggesting that the proteins are interacting directly, rather than being associated via interaction with cellular RNAs present in the extract. The two studies disagreed on the requirement of dFXR for RNAi. Caudy et al. found that depletion of dFXR by RNAi suppresses RNAi slightly, whereas Ishizuka et al. did not observe any effect. These conflicting results might indicate that dFXR is probably an accessory factor, rather than an essential component of RISC. Not only siRNAs but also miRNAs are present in complexes with dFXR.

The Fragile X family of RNA-binding proteins contain two KH domains and an RGG box, all of which bind RNA. Many RNAs have been found to interact with FMRP (Brown et al., 2001; Brown et al., 1998), including its own message. Both SELEX (systematic evolution of ligands by exponential enrichment) experiments (Darnell et al., 2001) and a directed analysis of the FMRP binding to its message (Schaeffer et al., 2001) indicate that the RGG box selectively binds the G-quartet structures. However, the RNAs resulting from the SELEX experiments are not only potential G-quartets, but also short hairpins with a 14-bp stem. Some mutations that disrupt binding also disrupt this stem loop, and not the potential G-quartet. Thus, some type of secondary structure, perhaps a hairpin, might also be important for the recognition of target RNAs by FMRP.

Although there is a single Fragile X family member in *Drosophila*, mammals have three Fragile-X-related proteins – FMRP, FXR1P and FXR2P. These proteins are closely related in sequence and are expressed in most tissue types. In addition to mental retardation, individuals with Fragile X syndrome have a number of other phenotypes, suggesting that the Fragile X protein family may be important for gene regulation in a number of cell types and not just in the brain. Fragile X family members associate with polyribosomes in an RNA-dependent manner (Khandjian et al., 1996). The I304N point mutation in the FMRP KH domain is associated with disease (De Boulle et al., 1993) and disrupts the binding of FMRP to polysomes (Feng et al., 1997; Tamanini et al., 1999).

The analogous mutation in Drosophila also disrupts dFXR association with ribosomes (Caudy et al. 2002).

1.4.8 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. (2002) constructed and used a special transgenic strain of *C. elegans* (HC57). They identified a systemic RNA interference-deficient (*sid*) locus required to transmit gene silencing between cells, using the green fluorescent protein (GFP) as a marker. 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 *Drosophila*, which is consistent with the apparent lack of systemic RNAi in this organism. However, the presence of SID homologues in humans and mice might hint at the systemic characteristics of RNAi in mammals.

1.5 RNAi as a tool

The excitement connected with the discovering RNAi is not only due to the phenomena itself. RNAi has revolutionized all aspect of biological sciences by offering an excellent new way to knock down genes in various organisms for studying their functions. As a powerful new tool of reverse genetics, RNAi has a potential to allow systematic analysis of gene expression. Especially in mammals, the advance with siRNA-directed knock-downs has sparked a revolution in somatic cell genetics, allowing the inexpensive and rapid analysis of gene function. In addition, it also offers possibilities for a therapeutic gene silencing.

1.5.1 Silencing by long dsRNA

Long dsRNA was used in *C. elegans*, *Drosophila*, plants, and other organisms for probing gene function. It is experimentally very simple to introduce dsRNA into worm or *Drosophila* cells. In *C. elegans*, injection into the intestine or pseudocoelom (body cavity) is almost as efficient as injection into the germ line. Even feeding worms with bacteria that express dsRNA, or soaking worms in dsRNA solutions have been applied with success (Timmons and Fire, 1998). Testing the function of individual genes by RNAi has now extended to analysis of nearly all worm predicted genes. Similar strategies are being pursued in other organisms. DNA-vector-mediated mechanisms to express long dsRNA were also tested. The constructs make use of an RNA polymerase II (pol II) promoters to drive the expression, what allows inducible, tissue- or cell- type specific RNA expression.

1.5.2 Silencing by siRNA

Using siRNAs to trigger gene silencing was first explored in mammalian cells to bypass their nonspecific responses to long dsRNA. Based on experimental analyses, siRNAs are now being optimized for systematic exploration of the function of genes in various organisms. In brief, the application of siRNA for gene silencing involves a careful consideration of the following variables: (i) selecting the target siRNA sequence in the gene; (ii) synthesis of siRNAs versus construction of plasmids encoding siRNAs; (iii) optimizing transfection of siRNAs or plasmids expressing siRNAs in the target cells; and (iv) monitoring the efficacy of gene silencing.

1.5.2.1 Selection of siRNA

Different siRNAs synthesized against different regions of the same target mRNA will show different silencing efficiencies (Holen et al., 2002). To promote efficient gene silencing by siRNA, the consideration of the siRNA sequence is crucial. A number of groups have analyzed 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 *Drosophila* and human cells (Elbashir et al., 2001; Schwarz et al., 2002). No consensus on choosing the siRNA sequence has evolved, but recommendations offerred are the following. 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. For synthetic siRNA, it is suggested to be 21 nucleotides long and have 5'-phosphate and 3'-hydroxyl group for efficiency. The GC content of the siRNAs should be kept between 30 and 70%. Computer programs were developed to offer helpful guidelines to select potential siRNA sequences and determine whether these selected sequences match mRNA sequences other than those of intended target, e.g. by Lin (Jack Lin's siRNA sequence finder; www.lc.sunysb.edu/stu/shiklin/rnai.html) and by Ambion (www.ambion.com) Compared to antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. Due to the paucity of information on the selection of siRNAs and their structures, these general guidelines are only suggestive and do not guarantee the silencing effect. Recently, new studies revealed the asymmetry in the assembly of the RISC. The RISC assembly is governed by an enzyme that selects which strand of an siRNA is loaded into RISC. This strand is always the one whose 5' end is less tightly paired to its complement. For miRNAs, it is the miRNA strand of a short-lived, siRNA duplex-like intermediate that assembles into a RISC complex, casusing miRNAs to accumulate in vivo as single-stranded RNAs. The authors suggested that for designing siRNAs it is important to place the 5' end of the antisense siRNA strand in a mismatch or G:U base pair. The designed duplexes resemble double-stranded miRNA intermediates and produce highly functional siRNAs (Khvorova et al., 2003; Schwarz et al., 2003).

1.5.2.2 Generation of siRNA

The 21-nucleotide siRNAs can be chemically synthesized with appropriately protected ribonucleoside phosphoramidites by 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. The *in vitro* T7 RNA polymerase synthesized siRNAs were also tested (Donze and Picard, 2002). Several groups have also used either the *E. coli* RNase III (Calegari et al., 2002; Yang et al., 2002) or the recombinant human Dicer (Kawasaki et al., 2003; Myers et al., 2003) to cleave the *in vitro* transcribed long dsRNA into siRNAs that can be transfected into mammalian cells. This approach allows generation of siRNAs with multiple specificities to the target and overcomes the siRNA selection ambiguity without activating an interferon response.

1.5.2.2.1 DNA-vector-mediated RNAi

The high cost of synthesizing siRNAs and their transient effect have compelled investigators to explore alternative strategies to generate a continuous supply of a battery of siRNAs. Several groups have devised strategies introducing plasmids with the ability to make siRNAs in vivo (Yang et al., 2002; Yu et al., 2002). DNA-based plasmid vectors have been designed by cloning siRNA templates downstream of an pol III promoter. Pol III promoters (either human or mouse U6 snRNA or human RNase P [H1] RNA promoters) were used to drive expression of short RNAs, because they precisely initiate and terminate RNA transcription (Goomer and Kunkel, 1992). Although U6 and H1 promoters contain the same set of cis-acting-elements, the H1 promoter has a more compact organization. The U6 promoter has a requirement for a guanosine in the +1 position, whereas the H1 promoter is more permissive. RNA pol III recognizes a simple cluster of four or more T residues as a termination signal that accurately and efficiently terminates transcription in the absence of other elements. Although most expression systems use either the U6 or H1 promoter, Kawasaki and Taira (2003) recently described an expression system that uses the $tRNA^{Val}$ promoter.

Two approaches have been developed for expressing siRNAs. In the first, sense and antisense strands constituting the siRNA duplex are transcribed from separate promoters (Lee et al., 2002a; Miyagishi and Taira, 2002), and in the second, siRNAs are expressed as fold-back stem-loop structures (referred to as shRNAs) that give rise to siRNAs after cleavage of the loop by cellular endonucleases (Brummelkamp et al., 2002; Paddison and Hannon, 2002). The shRNA approach is the most commonly used now.

1.5.2.2.2 Virus-vector-mediated RNAi

Virus-based high-efficiency siRNA delivery systems are also being developed. Two types of retrovirus vectors have been used as gene delivery systems: oncoretrovirus vectors that are based on the Moloney murine leukemia virus (MoMuLV) or the murine stem cell virus (MSCV), and lentivirus vectors that are derived from HIV-1. In one study, a U6 expression cassette was incorporated into the long terminal repeat (LTR) of the MoMuLV-based vector, pBabe-puro. Owing to the activity of the reverse transcriptase, which duplicates the LTR, the proviral form contains two copies of the LTR and therefore two copies of the U6 expression cassette. Expression of shRNA against the tumour suppressor p53 silenced p53 stably and resulted in an expected phenotype

(Paddison et al., 2002). The H1 expression cassette was also incorporated into a self-inactivating MSCV vector and successfully targeted endogenous genes. Compared to oncoretrovirus vectors, lentiviruses have two distinct characteristics that make them more effective gene delivery vectors. Unlike oncoretrovirus vectors, HIV-1 based lentivirus vectors can infect both actively dividing and non-dividing, post-mitotic cells (Naldini et al., 1996a; Naldini et al., 1996b). In addition, oncoretroviruses undergo proviral silencing during development, which leads to decreased or abrogated gene expression (Svoboda et al., 2000a). Lentivirus-based vectors are resistant to this silencing and therefore can be used to generate transgenic animals. Human peripheral blood T lymphocytes that were infected with a lentivirus vector expressing a shRNA against the HIV-1 coreceptor CCR5 showed a 10-fold decrease in CCR5 expression and, when challenged with a CCR5-tropic HIV-1 virus, resulted in a 3-7 fold reduction in HIV-1-infected cells (Qin et al., 2003). Although lentivirus vectors hold promise as vehicles for gene therapy, the development of leukaemias in two patients that were undergoing retroviral-based therapy for X-linked severe combined immunodeficiency indicates that better control must be achieved before lentiviruses can be used to deliver hairpin RNAs for therapeutic purposes.

1.5.2.2.3 Other systems

Most of the siRNA expression vectors produced to date use pol III regulatory units, which do not allow tissue-specific siRNA expression. However, Shiagawa and Ishiid (Shinagawa and Ishii, 2003) reported a pol II promoter-based plasmid encoding dsRNA that could eventually express siRNA in a tissue-specific manner. In their novel scheme, a pDECAP vector was used, which expressed a long dsRNA 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 which prevented 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 control protein, SNO, remained unaffected. The same vector was also used to create *ski* knock-down mice, the phenotype of which was similar to that of *ski* knock-out 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 animals. Although the pDECAP vector system looks very promising, these findings still require independent confirmation. Nuclear fragmentation of dsRNA to siRNAs also needs rationale explanation. To date, no evidence exists that Dicer localizes to the nuclear in mammalian cells. With the advent of vector-mediated siRNA delivery methods, it is now possible to make transgenic animals that can silence gene expression stably. This can be done by standard transgene technology or by the infection of embryonic stem cells or blastocysts with lentivirus vectors (Hasuwa et al., 2002; Lois et al., 2002).

The rapid injection of large volumes of a physiological solution containing siRNA into the tail vein of postnatal mice, which was originally described for delivery of plasmid DNA to various organs, has also been used to silence gene expression in various mouse tissues. The study showed that injected siRNAs are stable and that they remain sufficiently concentrated to produce an expected gene silencing (Song et al., 2003a).

1.5.2.3 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. Few newer transfection reagents such as TransIT-TKO (Mirus) and Ambion's Siport 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* (McRobert and McConkey, 2002; Shi et al., 2000). In adult mice, naked siRNAs have been delivered by hydrodynamic transfection methods to combat hepatitis C virus infection in liver (McCaffrey et al., 2002). The preferred way to detect specific gene knock-down by RNAi is to study the depletion of the target protein by immunofluorescence and Western blotting with the specific antibody. In addition, the Northern blot analysis and knockdown phenotype can also be used to detect the effects of siRNA. If the gene is essential, cellular growth is delayed or arrested.

Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.
Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H., and Vance, V. B. (1998). A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A *95*, 13079-13084.
Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J., and Matzke, M. (2002). RNA-directed DNA methylation in Arabidopsis. Proc Natl Acad Sci U S A *99 Suppl 4*, 16499-16506.
Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.
Barber, G. N. (2001). Host defense, viruses and apoptosis. Cell Death Differ *8*, 113-126.
Bashirullah, A., Pasquinelli, A. E., Kiger, A. A., Perrimon, N., Ruvkun, G., and Thummel, C. S. (2003). Coordinate regulation of small temporal RNAs at the onset of Drosophila metamorphosis. Dev Biol *259*, 1-8.
Baulcombe, D. (1999). Viruses and gene silencing in plants. Arch Virol Suppl *15*, 189-201.
Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.
Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.
Billy, E., Brondani, V., Zhang, H., Muller, U., and Filipowicz, W. (2001). Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. Proc Natl Acad Sci U S A *98*, 14428-14433.
Birchler, J. A., Pal-Bhadra, M., and Bhadra, U. (1999). Less from more: cosuppression of transposable elements. Nat Genet *21*, 148-149.
Blaszczyk, J., Tropea, J. E., Bubunenko, M., Routzahn, K. M., Waugh, D. S., Court, D. L., and Ji, X. (2001). Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure (Camb) *9*, 1225-1236.
Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998a). AGO1 defines a novel locus of Arabidopsis controlling leaf development. Embo J *17*, 170-180.
Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998b). AGO1 defines a novel locus of Arabidopsis controlling leaf development. In Embo J, pp. 170-180.
Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., and Cohen, S. M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25-36.
Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W., and Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in Nicotiana benthamiana. Embo J *17*, 6739-6746.
Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., Jin, X., Feng, Y., Wilkinson, K. D., Keene, J. D.*, et al.* (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell *107*, 477-487.
Brown, V., Small, K., Lakkis, L., Feng, Y., Gunter, C., Wilkinson, K. D., and Warren, S. T. (1998). Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein. J Biol Chem *273*, 15521-15527.
Brummelkamp, T. R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science *296*, 550-553.
Calegari, F., Haubensak, W., Yang, D., Huttner, W. B., and Buchholz, F. (2002). Tissue-specific RNA interference in postimplantation mouse embryos with endoribonuclease-prepared short interfering RNA. Proc Natl Acad Sci U S A *99*, 14236-14240.
Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2002). Involvement of small RNAs and role of the qde genes in the gene silencing pathway in Neurospora. Genes Dev *16*, 790-795.
Caudy, A. A., Ketting, R. F., Hammond, S. M., Denli, A. M., Bathoorn, A. M., Tops, B. B., Silva, J. M., Myers, M. M., Hannon, G. J., and Plasterk, R. H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. Nature *425*, 411-414.
Caudy, A. A., Myers, M., Hannon, G. J., and Hammond, S. M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev *16*, 2491-2496.
Chanfreau, G., Buckle, M., and Jacquier, A. (2000). Recognition of a conserved class of RNA tetraloops by Saccharomyces cerevisiae RNase III. Proc Natl Acad Sci U S A *97*, 3142-3147.
Cikaluk, D. E., Tahbaz, N., Hendricks, L. C., DiMattia, G. E., Hansen, D., Pilgrim, D., and Hobman, T. C. (1999). GERp95, a membrane-associated protein that belongs to a family of proteins involved in stem cell differentiation. Mol Biol Cell *10*, 3357-3372.

Cogoni, C., Irelan, J. T., Schumacher, M., Schmidhauser, T. J., Selker, E. U., and Macino, G. (1996). Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. Embo J *15*, 3153-3163.
Cogoni, C., and Macino, G. (1997). Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa. Proc Natl Acad Sci U S A *94*, 10233-10238.
Cogoni, C., and Macino, G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature *399*, 166-169.
Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.
Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D. C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell *101*, 543-553.
Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell *107*, 489-499.
Der, S. D., Yang, Y. L., Weissmann, C., and Williams, B. R. (1997). A double-stranded RNA-activated protein kinase-dependent pathway mediating stress-induced apoptosis. Proc Natl Acad Sci U S A *94*, 3279-3283.
Ding, S. W., Shi, B. J., Li, W. X., and Symons, R. H. (1996). An interspecies hybrid RNA virus is significantly more virulent than either parental virus. Proc Natl Acad Sci U S A *93*, 7470-7474.
Djikeng, A., Shi, H., Tschudi, C., and Ullu, E. (2001). RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. Rna *7*, 1522-1530.
Doench, J. G., Petersen, C. P., and Sharp, P. A. (2003). siRNAs can function as miRNAs. Genes Dev *17*, 438-442.
Domeier, M. E., Morse, D. P., Knight, S. W., Portereiko, M., Bass, B. L., and Mango, S. E. (2000). A link between RNA interference and nonsense-mediated decay in Caenorhabditis elegans. Science *289*, 1928-1931.
Donze, O., and Picard, D. (2002). RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase. Nucleic Acids Res *30*, e46.
Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.
English, J. J., Mueller, E., and Baulcombe, D. C. (1996). Suppression of Virus Accumulation in Transgenic Plants Exhibiting Silencing of Nuclear Genes. Plant Cell *8*, 179-188.
Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.
Fagard, M., Boutet, S., Morel, J. B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A *97*, 11650-11654.
Feng, Y., Absher, D., Eberhart, D. E., Brown, V., Malter, H. E., and Warren, S. T. (1997). FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. Mol Cell *1*, 109-118.
Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.
Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell *90*, 1023-1029.
Goomer, R. S., and Kunkel, G. R. (1992). The transcriptional start site for a human U6 small nuclear RNA gene is dictated by a compound promoter element consisting of the PSE and the TATA box. Nucleic Acids Res *20*, 4903-4912.
Graveley, B. R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell *1*, 765-771.
Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.
Guo, S., and Kemphues, K. J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.
Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. Embo J *21*, 4671-4679.
Hamilton, A. J., and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.
Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., and Hannon, G. J. (2001a). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.
Hammond, S. M., Caudy, A. A., and Hannon, G. J. (2001b). Post-transcriptional gene silencing by double-stranded RNA. Nat Rev Genet *2*, 110-119.

Hannon, G. J. (2002). RNA interference. Nature *418*, 244-251.
Harborth, J., Elbashir, S. M., Bechert, K., Tuschl, T., and Weber, K. (2001). Identification of essential genes in cultured mammalian cells using small interfering RNAs. J Cell Sci *114*, 4557-4565.
Hasuwa, H., Kaseda, K., Einarsdottir, T., and Okabe, M. (2002). Small interfering RNA and gene silencing in transgenic mice and rats. FEBS Lett *532*, 227-230.
Holen, T., Amarzguioui, M., Wiiger, M. T., Babaie, E., and Prydz, H. (2002). Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. Nucleic Acids Res *30*, 1757-1766.
Hunter, T., Hunt, T., Jackson, R. J., and Robertson, H. D. (1975). The characteristics of inhibition of protein synthesis by double-stranded ribonucleic acid in reticulocyte lysates. J Biol Chem *250*, 409-417.
Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.
Hutvagner, G., and Zamore, P. D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.
Ishizuka, A., Siomi, M. C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev *16*, 2497-2508.
Jacobsen, S. E., Running, M. P., and Meyerowitz, E. M. (1999). Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. Development *126*, 5231-5243.
Jones, K. W., Gorzynski, K., Hales, C. M., Fischer, U., Badbanchi, F., Terns, R. M., and Terns, M. P. (2001). Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. J Biol Chem *276*, 38645-38651.
Jones, L., Hamilton, A. J., Voinnet, O., Thomas, C. L., Maule, A. J., and Baulcombe, D. C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. Plant Cell *11*, 2291-2301.
Jorgensen, R. A., Cluster, P. D., English, J., Que, Q., and Napoli, C. A. (1996). Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences. Plant Mol Biol *31*, 957-973.
Kasschau, K. D., Xie, Z., Allen, E., Llave, C., Chapman, E. J., Krizan, K. A., and Carrington, J. C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. Dev Cell *4*, 205-217.
Kataoka, Y., Takeichi, M., and Uemura, T. (2001). Developmental roles and molecular characterization of a Drosophila homologue of Arabidopsis Argonaute1, the founder of a novel gene superfamily. Genes Cells *6*, 313-325.
Kawasaki, H., Suyama, E., Iyo, M., and Taira, K. (2003). siRNAs generated by recombinant human Dicer induce specific and significant but target site-independent gene silencing in human cells. Nucleic Acids Res *31*, 981-987.
Kawasaki, H., and Taira, K. (2003). Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. Nucleic Acids Res *31*, 700-707.
Kennerdell, J. R., and Carthew, R. W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-1026.
Kennerdell, J. R., Yamaguchi, S., and Carthew, R. W. (2002). RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. Genes Dev *16*, 1884-1889.
Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.
Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., and Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.
Khandjian, E. W., Corbin, F., Woerly, S., and Rousseau, F. (1996). The fragile X mental retardation protein is associated with ribosomes. Nat Genet *12*, 91-93.
Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003). Functional siRNAs and miRNAs Exhibit Strand Bias. Cell *115*, 505.
Knight, S. W., and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.
Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.
Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr Biol *12*, 735-739.
Lai, E. C., Burks, C., and Posakony, J. W. (1998). The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts. Development *125*, 4077-4088.
Lai, E. C., and Posakony, J. W. (1997). The Bearded box, a novel 3' UTR sequence motif, mediates negative post-transcriptional regulation of Bearded and Enhancer of split Complex gene expression. Development *124*, 4847-4856.
Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lee, N. S., Dohjima, T., Bauer, G., Li, H., Li, M. J., Ehsani, A., Salvaterra, P., and Rossi, J. (2002a). Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. Nat Biotechnol *20*, 500-505.
Lee, R. C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.
Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.
Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.
Lee, Y., Jeon, K., Lee, J. T., Kim, S., and Kim, V. N. (2002b). MicroRNA maturation: stepwise processing and subcellular localization. Embo J *21*, 4663-4670.
Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.
Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B., and Bartel, D. P. (2003a). Vertebrate microRNA genes. Science *299*, 1540.
Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. (2003b). The microRNAs of Caenorhabditis elegans. Genes Dev *17*, 991-1008.
Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465-469.
Lipardi, C., Wei, Q., and Paterson, B. M. (2001). RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. Cell *107*, 297-307.
Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H. E., Smith, D. P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.
Llave, C., Kasschau, K. D., and Carrington, J. C. (2000). Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. Proc Natl Acad Sci U S A *97*, 13401-13406.
Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.
Lois, C., Hong, E. J., Pease, S., Brown, E. J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science *295*, 868-872.
Lucy, A. P., Guo, H. S., Li, W. X., and Ding, S. W. (2000). Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. Embo J *19*, 1672-1680.
Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.
Mallory, A. C., Reinhart, B. J., Bartel, D., Vance, V. B., and Bowman, L. H. (2002). A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. Proc Natl Acad Sci U S A *99*, 15228-15233.
Martienssen, R. A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science *293*, 1070-1074.
Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.
Matzke, M., Matzke, A. J., and Kooter, J. M. (2001). RNA: guiding gene silencing. Science *293*, 1080-1083.
McCaffrey, A. P., Meuse, L., Pham, T. T., Conklin, D. S., Hannon, G. J., and Kay, M. A. (2002). RNA interference in adult mice. Nature *418*, 38-39.
McRobert, L., and McConkey, G. A. (2002). RNA interference (RNAi) inhibits growth of Plasmodium falciparum. Mol Biochem Parasitol *119*, 273-278.
Meister, G., Hannus, S., Plottner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B., and Fischer, U. (2001). SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. Embo J *20*, 2304-2314.
Miyagishi, M., and Taira, K. (2002). Development and application of siRNA expression vector. Nucleic Acids Res Suppl, 113-114.
Mochizuki, K., Fine, N. A., Fujisawa, T., and Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689-699.
Morel, J. B., Godon, C., Mourrain, P., Beclin, C., Boutet, S., Feuerbach, F., Proux, F., and Vaucheret, H. (2002). Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell *14*, 629-639.
Moss, E. G., and Tang, L. (2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. Dev Biol *258*, 432-442.
Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. Embo J *20*, 5443-5452.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev *16*, 720-728.
Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N.*, et al.* (2000a). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. In Cell, pp. 533-542.
Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N.*, et al.* (2000b). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell *101*, 533-542.
Myers, J. W., Jones, J. T., Meyer, T., and Ferrell, J. E., Jr. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. Nat Biotechnol *21*, 324-328.
Naldini, L., Blomer, U., Gage, F. H., Trono, D., and Verma, I. M. (1996a). Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. Proc Natl Acad Sci U S A *93*, 11382-11388.
Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M., and Trono, D. (1996b). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263-267.
Nykanen, A., Haley, B., and Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.
Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J., and Conklin, D. S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev *16*, 948-958.
Paddison, P. J., and Hannon, G. J. (2002). RNA interference: the new somatic cell genetics? Cancer Cell *2*, 17-23.
Palauqui, J. C., Elmayan, T., Pollien, J. M., and Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. Embo J *16*, 4738-4745.
Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (1999). Cosuppression of nonhomologous transgenes in Drosophila involves mutually related endogenous sequences. Cell *99*, 35-46.
Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol Cell *9*, 315-327.
Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol *12*, 1484-1495.
Parrish, S., and Fire, A. (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. Rna *7*, 1397-1402.
Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). A functional interaction between the survival motor neuron complex and RNA polymerase II. J Cell Biol *152*, 75-85.
Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell *95*, 615-624.
Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. (2002). Ribonuclease activity and RNA binding of recombinant human Dicer. Embo J *21*, 5864-5874.
Qin, X. F., An, D. S., Chen, I. S., and Baltimore, D. (2003). Inhibiting HIV-1 infection in human T cells by lentiviral-mediated delivery of small interfering RNA against CCR5. Proc Natl Acad Sci U S A *100*, 183-188.
Que, Q., and Jorgensen, R. A. (1998). Homology-based control of gene expression patterns in transgenic petunia flowers. Dev Genet *22*, 100-109.
Reinhart, B. J., and Bartel, D. P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.
Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.
Richards, E. J., and Elgin, S. C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.
Roignant, J. Y., Carre, C., Mugat, B., Szymczak, D., Lepesant, J. A., and Antoniewski, C. (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila. Rna *9*, 299-308.
Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. Embo J *20*, 4803-4813.
Schauer, S. E., Jacobsen, S. E., Meinke, D. W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in Arabidopsis development. Trends Plant Sci *7*, 487-491.
Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sanger, H. L., and Wassenegger, M. (1998). Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. Plant Cell *10*, 2087-2101.

Schmidt, A., Palumbo, G., Bozzetti, M. P., Tritto, P., Pimpinelli, S., and Schafer, U. (1999). Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. Genetics *151*, 749-760.
Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.
Schwarz, D. S., Hutvagner, G., Haley, B., and Zamore, P. D. (2002). Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol Cell *10*, 537-548.
Sempere, L. F., Sokol, N. S., Dubrovsky, E. B., Berger, E. M., and Ambros, V. (2003). Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity. Dev Biol *259*, 9-18.
Shi, H., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C., and Ullu, E. (2000). Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA. Rna *6*, 1069-1076.
Shinagawa, T., and Ishii, S. (2003). Generation of Ski-knockdown mice by expressing a long double-strand RNA from an RNA polymerase II promoter. Genes Dev *17*, 1340-1345.
Sijen, T., Fleenor, J., Simmer, F., Thijssen, K. L., Parrish, S., Timmons, L., Plasterk, R. H., and Fire, A. (2001a). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.
Sijen, T., Vijn, I., Rebocho, A., van Blokland, R., Roelofs, D., Mol, J. N., and Kooter, J. M. (2001b). Transcriptional and posttranscriptional gene silencing are mechanistically related. Curr Biol *11*, 436-440.
Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., Ahringer, J., and Plasterk, R. H. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. Curr Biol *12*, 1317-1319.
Smardon, A., Spoerke, J. M., Stacey, S. C., Klein, M. E., Mackin, N., and Maine, E. M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol *10*, 169-178.
Smulders-Srinivasan, T. K., and Lin, H. (2003). Screens for piwi Suppressors in Drosophila Identify Dosage-Dependent Regulators of Germline Stem Cell Division. Genetics *165*, 1971-1991.
Song, E., Lee, S. K., Wang, J., Ince, N., Ouyang, N., Min, J., Chen, J., Shankar, P., and Lieberman, J. (2003a). RNA interference targeting Fas protects mice from fulminant hepatitis. Nat Med *9*, 347-351.
Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., and Joshua-Tor, L. (2003b). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.
Stark, A., Brennecke, J., Russell, R. B., and Cohen, S. M. (2003). Identification of Drosophila MicroRNA Targets. PLoS Biol *1*, E60.
Svoboda, J., Hejnar, J., Geryk, J., Elleder, D., and Vernerova, Z. (2000a). Retroviruses in foreign species and the problem of provirus silencing. Gene *261*, 181-188.
Svoboda, P., Stein, P., Hayashi, H., and Schultz, R. M. (2000b). Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. Development *127*, 4147-4156.
Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C. C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.
Tamanini, F., Van Unen, L., Bakker, C., Sacchi, N., Galjaard, H., Oostra, B. A., and Hoogeveen, A. T. (1999). Oligomerization properties of fragile-X mental-retardation protein (FMRP) and the fragile-X-related proteins FXR1P and FXR2P. Biochem J *343 Pt 3*, 517-523.
Tang, G., Reinhart, B. J., Bartel, D. P., and Zamore, P. D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev *17*, 49-63.
Taverna, S. D., Coyne, R. S., and Allis, C. D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell *110*, 701-711.
Tijsterman, M., Ketting, R. F., Okihara, K. L., Sijen, T., and Plasterk, R. H. (2002). RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs. Science *295*, 694-697.
Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.
Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., and Sharp, P. A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev *13*, 3191-3197.
Vaistij, F. E., Jones, L., and Baulcombe, D. C. (2002). Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. Plant Cell *14*, 857-867.
Vaucheret, H., and Fagard, M. (2001). Transcriptional gene silencing in plants: targets, inducers and regulators. Trends Genet *17*, 29-35.
Voinnet, O., Lederer, C., and Baulcombe, D. C. (2000). A viral movement protein prevents spread of the gene silencing signal in Nicotiana benthamiana. Cell *103*, 157-167.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D. C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell *95*, 177-187.
Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.
Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol *2*, 70-75.
Williams, R. W., and Rubin, G. M. (2002). ARGONAUTE1 is required for efficient RNA interference in Drosophila embryos. Proc Natl Acad Sci U S A *99*, 6889-6894.
Winston, W. M., Molodowitch, C., and Hunter, C. P. (2002). Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. Science *295*, 2456-2459.
Wu, H., Xu, H., Miraglia, L. J., and Crooke, S. T. (2000). Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. J Biol Chem *275*, 36957-36965.
Wu, H., Yang, P. K., Butcher, S. E., Kang, S., Chanfreau, G., and Feigon, J. (2001). A novel family of RNA tetraloop structure forms the recognition site for Saccharomyces cerevisiae RNase III. Embo J *20*, 7240-7249.
Wu-Scharf, D., Jeong, B., Zhang, C., and Cerutti, H. (2000). Transgene and transposon silencing in Chlamydomonas reinhardtii by a DEAH-box RNA helicase. Science *290*, 1159-1162.
Xu, P., Vernooy, S. Y., Guo, M., and Hay, B. A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.
Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.
Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C. Y., Brodsky, F. M., and Bishop, J. M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proc Natl Acad Sci U S A *99*, 9942-9947.
Yang, S., Tutton, S., Pierce, E., and Yoon, K. (2001). Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. Mol Cell Biol *21*, 7807-7816.
Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.
Yu, J. Y., DeRuiter, S. L., and Turner, D. L. (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. Proc Natl Acad Sci U S A *99*, 6047-6052.
Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.
Zhang, H., Kolb, F. A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. Embo J *21*, 5875-5885.
Zhang, Y., Olbort, M., Schwarzer, K., Nuesslein-Hildesheim, B., Nicolson, M., Murphy, E., Kowalski, T. J., Schmidt, I., and Leibel, R. L. (1997). The leptin receptor mediates apparent autocrine regulation of leptin gene expression. Biochem Biophys Res Commun *240*, 492-495.
Zhou, A., Paranjape, J., Brown, T. L., Nie, H., Naik, S., Dong, B., Chang, A., Trapp, B., Fairchild, R., Colmenares, C., and Silverman, R. H. (1997). Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. Embo J *16*, 6355-6363.
Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.

Chapter 2

Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP

Published Manuscript:

Haidi Zhang, Fabrice A.Kolb, Vincent Brondani, Eric Billy and Witold Filipowicz (2002) ***The EMBO Journal*** **21, 5875-5885.**

The EMBO Journal Vol. 21 No. 21 pp. 5875–5885, 2002

Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP

Haidi Zhang, Fabrice A.Kolb, Vincent Brondani[1], Eric Billy[2] and Witold Filipowicz[3]

Friedrich Miescher Institute for Biomedical Research, PO Box 2543, CH-4002 Basel, Switzerland

[1]Present address: Institute of Medical Microbiology, University of Basel, CH-4003 Basel, Switzerland
[2]Present address: Novartis Pharma AG, CH-4056 Basel, Switzerland

[3]Corresponding author
e-mail: Filipowi@fmi.ch

Dicer is a multi-domain RNase III-related endonuclease responsible for processing double-stranded RNA (dsRNA) to small interfering RNAs (siRNAs) during a process of RNA interference (RNAi). It also catalyses excision of the regulatory microRNAs from their precursors. In this work, we describe the purification and properties of a recombinant human Dicer. The protein cleaves dsRNAs into ~22 nucleotide siRNAs. Accumulation of processing intermediates of discrete sizes, and experiments performed with substrates containing modified ends, indicate that Dicer preferentially cleaves dsRNAs at their termini. Binding of the enzyme to the substrate can be uncoupled from the cleavage step by omitting Mg^{2+} or performing the reaction at 4°C. Activity of the recombinant Dicer, and of the endogenous protein present in mammalian cell extracts, is stimulated by limited proteolysis, and the proteolysed enzyme becomes active at 4°C. Cleavage of dsRNA by purifed Dicer and the endogenous enzyme is ATP independent. Additional experiments suggest that if ATP participates in the Dicer reaction in mammalian cells, it might be involved in product release needed for the multiple turnover of the enzyme.

Keywords: gene silencing/ribonucleases/RNAi/RNase III/siRNA

Introduction

In many eukaryotes, double-stranded RNA (dsRNA) inhibits gene expression in a sequence-specific manner by triggering degradation of mRNA. This effect, referred to as RNA interference (RNAi), has many features in common with post-transcriptional gene silencing (PTGS) in plants and quelling in *Neurospora crassa*. RNAi/PTGS appears to function as a defence system against viruses and in preventing transposon movement. RNAi also offers a way to inactivate genes of interest and, thus, provides a powerful tool to study gene function (reviewed by Hannon, 2002; Hutvágner and Zamore, 2002a).

Genetic and biochemical studies have revealed that RNAi/PTGS is a very complex reaction, probably involving dozens of different proteins and intersecting with other cellular processes. Ongoing work is now beginning to unravel some of the mechanistic aspects of RNAi. Experiments carried out in *Drosophila* cell or embryo extracts, recapitulating the RNAi reaction *in vitro*, demonstrated that dsRNA is first processed into ~21 nucleotide (nt), small interfering RNAs (siRNAs) (Hammond *et al.*, 2000; Zamore *et al.*, 2000; Bernstein *et al.*, 2001; Elbashir *et al.*, 2001b). siRNAs are also implicated in RNAi/PTGS in other eukaryotes (Hamilton and Baulcombe, 1999; reviewed by Hannon, 2002; Hutvágner and Zamore, 2002a) and their role in mediating RNAi is supported by the demonstration that synthetic 21mer duplexes induce an RNAi in various organisms, both *in vivo* and *in vitro* (Caplen *et al.*, 2001; Elbashir *et al.*, 2001a,b; Nykänen *et al.*, 2001). siRNAs may affect mRNA degradation by two distinct mechanisms. In *Caenorhabditis elegans*, they appear to act as primers for the cellular RNA-dependent RNA polymerase (RdRP), converting the target mRNA into dsRNA. Such nascent dsRNA is then fragmented, resulting in elimination of the target mRNA and the generation of a new population of siRNAs (Sijen *et al.*, 2001). Although a similar mechanism may also operate in *Drosophila* (Lipardi *et al.*, 2001), the major pathway of mRNA degradation in flies appears to be different, involving a multi-component nuclease, called RISC (Hammond *et al.*, 2000; Nykänen *et al.*, 2001). siRNAs associated with RISC target homologous mRNAs for destruction by cleaving them within the region complementary to the siRNA (Hammond *et al.*, 2000; Zamore *et al.*, 2000; Elbashir *et al.*, 2001b; Nykänen *et al.*, 2001).

Irrespective of the mechanism of mRNA degradation, the generation of siRNAs is central to RNAi, and the RNase III-like enzyme (named Dicer) responsible for processing of dsRNA to siRNAs has recently been identified in *Drosophila*, *C.elegans* and mammals (Bernstein *et al.*, 2001; Billy *et al.*, 2001; Ketting *et al.*, 2001; Knight and Bass, 2001). Dicer is a large (~220 kDa) multi-domain protein present in all eukaryotes studied to date, with the exception of budding yeast. Its domains include a putative DExH/DEAH RNA helicase/ATPase domain, a PAZ signature, two neighbouring RNase III-like domains and a dsRNA-binding domain (RBD) (Figure 1A). dsRBD and RNase III domains are most certainly involved in dsRNA binding and cleavage, but the functions of the remaining domains are not known. In *Drosophila* extracts, RNAi requires ATP (Zamore *et al.*, 2000; Nykänen *et al.*, 2001). Studies performed with *Drosophila* and *C.elegans* extracts, and with the immunoprecipitates (IPs) containing *Drosophila* Dicer, indicated that one of the ATP-dependent steps is the processing of dsRNA to siRNAs (Zamore *et al.*, 2000; Bernstein *et al.*, 2001; Ketting *et al.*, 2001; Nykänen *et al.*, 2001). Consequently, it has been proposed that the helicase/ATPase domain of Dicer may promote

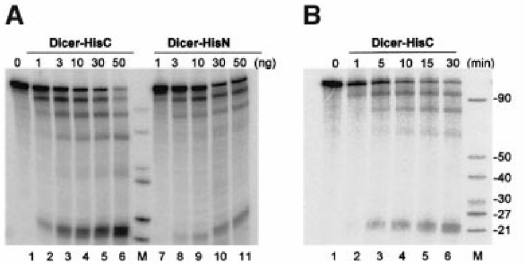

Fig. 2. Recombinant Dicer preparations cleave dsRNA into ~22-nt-long siRNA-like products. (**A**) Effect of increasing amounts (indicated at the top) of Dicer-HisC (lanes 2–6) and Dicer-HisN (lanes 7–11) on dsRNA cleavage. Lane M, RNA size markers (90, 70, 45, 40, 27 and 21 nt). Lane 1, reaction with no Dicer added. (**B**) Kinetics of dsRNA cleavage by Dicer-HisC (25 ng). The size of generated siRNA-like products (21–23 nts) was confirmed by co-electrophoresis with alkaline and RNase T1 ladders.

Dicer, assayed with IPs of the P19 cell extract prepared with the anti-Dicer antibody (Ab) (Billy *et al.*, 2001), is also stimulated by ProtK (Figure 3C). Since activities of the non-treated IPs and the equivalent amount of cell extract were similar (compare lane 0 with E), activation of the former is unlikely to be due to the protease-mediated release of Dicer from beads, association with which could potentially inhibit enzymatic activity. To further eliminate this possibility, attempts were made to detach Dicer from IP beads with an excess of the peptide used for immunization, but they were unsuccessful. Two other proteases, LysC and GluC, also stimulated activity of the protein by 2- to 3-fold (data not shown).

Dicer-HisC used in the experiments shown in Figures 2 and 3 was purified in the absence of reducing agents. Comparison of Dicer-HisC preparations purified in the presence or absence of β-mercaptoethanol and DTT indicated that inclusion of reducing agents lowers the basal activity of the protein by 2- to 3-fold. However, following treatment with ProtK, activities of both preparations were similar (data not shown). Buffers used for the preparation of Dicer-HisN contained DTT and, similarly to Dicer-HisC prepared under reducing conditions, Dicer-HisN had a relatively low basal activity and was strongly activated by proteolysis (Figure 3, compare A with B).

Requirements for Dicer activity

Dicer-HisC prepared under non-reducing conditions was used to optimize cleavage conditions, and used in most of the other experiments described below. Maximal activity was found at pH 6.5–6.9 and at 50–100 mM NaCl concentration; addition of NaCl to 0.2 M inhibited the reaction by >90% (see Supplementary figure A available at *The EMBO Journal* Online). The reaction required the presence of Mg^{2+}, the optimal concentration being 1–5 mM; addition of EDTA completely inhibited activity. dsRNA cleavage, though less efficient, was also observed when Mg^{2+} was replaced by Mn^{2+} and Co^{2+}, but not Ca^{2+}, Ni^{2+} or Zn^{2+} (Supplementary figure B). Requirements of the ProtK-activated enzyme were similar (data not shown). However, the response to temperature of native and ProtK-treated Dicer differed substantially. With non-treated

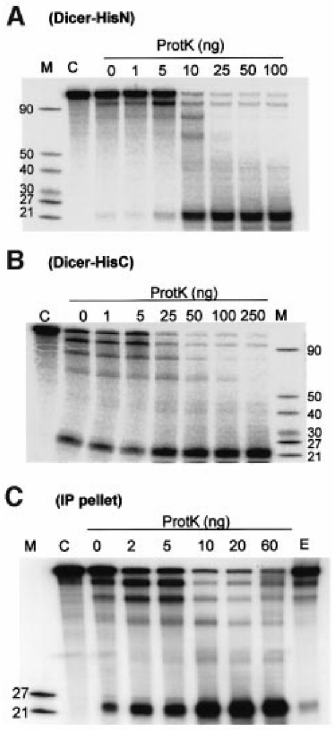

Fig. 3. Proteolysis stimulates activity of Dicer-HisN (**A**), Dicer-HisC (**B**) and IPs containing endogenous Dicer (**C**). Reactions contained 10 ng (A) and 25 ng (B) of recombinant Dicer, or 2 µl of IP beads (C). Amounts of added ProtK are indicated at the top. Lane C in (A) and (B), reactions incubated with no Dicer added. Lane E in (C), reaction containing the amount of P19 cell extract equivalent to the amount of IP beads used in other lanes.

Dicer, no cleavage was observed at 4°C, and cleavage at 25°C occurred ~2× slower than at 37°C. In contrast, the enzyme pre-treated with ProtK beads showed considerable activity even at 4°C (Supplementary figure C).

Estimation of the efficiency of recombinant proteins indicated that under optimal reaction conditions and following 30 min incubation, only ~0.1 mol of the siRNA product is generated per mole of enzyme. ProtK treatment increased this ratio up to ~0.5. In the experiments presented above, the concentration of dsRNA was 0.3–0.5 nM. When cleavage of the 130 bp dsRNA substrate was assayed at 30 nM concentration, ~0.5 mol of siRNA was formed per mole of the native protein. It is unknown whether this low number is due to only a small fraction of the recombinant enzyme being active or if it reflects a low catalytic efficiency of the enzyme (see Discussion).

Processing of dsRNA into siRNAs is ATP independent

Experiments carried out with *Drosophila* and *C.elegans* extracts, and IPs containing the *Drosophila* Dicer, demonstrated that processing of dsRNA into siRNAs is strongly stimulated by addition of ATP (see Introduction). However, our previously performed assays with P19 cell

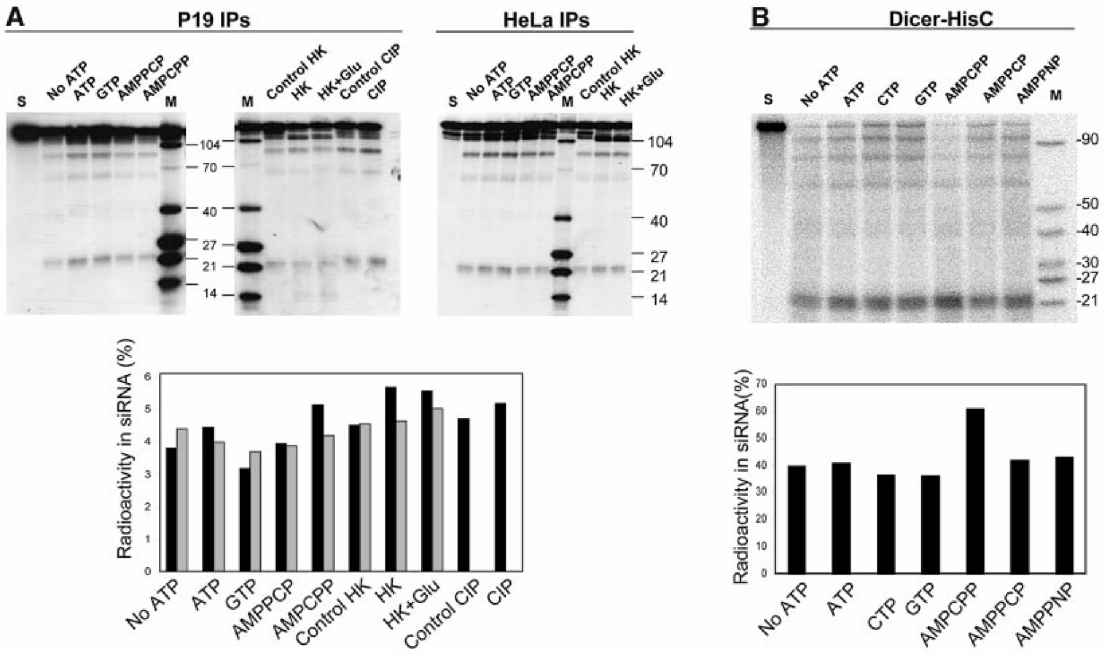

Fig. 4. Processing of dsRNA by human Dicer is ATP independent. (**A**) Effect of different nucleotides (2 mM; complexed with Mg^{2+}) and ATP depletion conditions on processing of dsRNA by Dicer present in IPs of P19 (left and middle gels) and HeLa (right gel) cell extracts. Lanes S, reactions without addition of IP beads and nucleotide. Lanes 'Control CIP' and 'Control HK', reactions pre-incubated without CIP, and HK and glucose addition, respectively. Lower panel, quantification of the P19 (black bars, averages of two independent experiments) and HeLa (grey bars, data from the gel shown) IPs. (**B**) Effect of different nucleotides on processing of dsRNA by Dicer-HisC (25 ng). Lane S, reaction without addition of Dicer and nucleotide. Lower panel, quantification of the data.

extracts indicated that ATP has no major effect on dsRNA cleavage in the mammalian system (Billy *et al.*, 2001). Dicer activity measured in IPs of the mouse P19 and human HeLa cell extracts also did not require addition of ATP, and was not affected by pre-incubation with either calf intestine phosphatase (CIP) or hexokinase (HK) and glucose, treatments performed to eliminate residual ATP (Figure 4A). Likewise, addition of non-hydrolysable ATP analogues, such as AMPPCP or AMPCPP, had no effect on dsRNA cleavage. ATP was also without effect when combined with the ATP regeneration system, creatine phosphate and creatine kinase (data not shown). Parallel experiments carried out with extracts prepared from *Drosophila* KC167 cells demonstrated a strong stimulatory effect of ATP on dsRNA processing in this system (data not shown), consistent with the results of other investigations (Zamore *et al.*, 2000; Bernstein *et al.*, 2001; Nykänen *et al.*, 2001).

We investigated whether activity of the recombinant human Dicer is stimulated by ATP. As shown in Figure 4B, ATP had no appreciable effect on cleavage of dsRNA. Similarly, addition of ATP analogues, or 1 mM GTP, CTP, UTP, dATP, ADP, NAD or NADP, had no influence on Dicer activity (Figure 4B; data not shown). Pre-incubation of the enzyme with either HK and glucose, or CIP, did not affect either the kinetics or the extent of the cleavage, further arguing against the requirement for ATP (Supplementary figure D).

Dicer contains an evolutionarily conserved DExH/DEAH ATPase/RNA helicase domain (Figure 1A). The reported ATP requirement of the dsRNA cleavage in *Drosophila* extracts is generally attributed to a potential role of this domain in translocation of the enzyme along the dsRNA, or to a structural rearrangement of the substrate required for the cleavage (Bernstein *et al.*, 2001; Ketting *et al.*, 2001; Nykänen *et al.*, 2001; Hutvágner and Zamore, 2002a). We wanted to eliminate the possibility that the observed ATP independence of the recombinant mammalian enzyme is due to ATP remaining associated with the ATPase/helicase domain, even following HK or CIP treatment. To this end, we prepared the Dicer-HisC mutant K70A, bearing the K to A substitution in the P-loop motif of the ATPase/helicase domain (Figure 1). In all investigated ATPases/helicases, and also other P-loop-containing proteins, this mutation strongly inhibits nucleotide binding and enzymatic activity (Gross and Shuman, 1998; Levin and Patel, 1999, and references therein). The DExH motif of Dicer, DECH, contains an evolutionarily conserved cysteine. Hence, the mutant and wild-type (wt) proteins were purified under both reducing and non-reducing conditions, and all preparations were assayed for dsRNA processing and

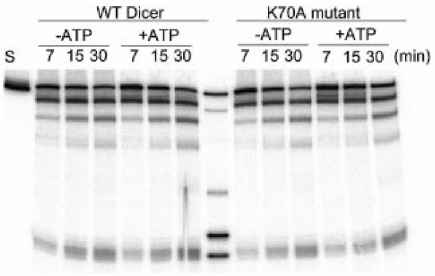

Fig. 5. Mutation K70A in the P-loop of the putative ATPase/helicase domain of Dicer has no effect on dsRNA cleavage. Reactions contained 25 ng of Dicer-HisC purified under reducing conditions, and were incubated for the times indicated in the absence or presence of 3 mM ATP.

ATPase activities. Nuclease assays have revealed that wt and mutant proteins purified under either reducing (Figure 5) or non-reducing (data not shown) conditions have similar activity, independent of the presence or absence of ATP. However, to date, we have been unable to demonstrate ATPase activity for any Dicer preparation, also when reactions included ssRNA, dsRNA, ssDNA or dsDNA (data not shown). Likewise, assays with the P19 cell extract IPs provided no reliable evidence of ATPase activity associated with Dicer.

We have also analysed effects of ATP, its analogues, and different ATP depletion conditions on the activity of wt and K70A mutant proteins at the 30 nM 130 bp dsRNA concentration. The results were similar to those presented above. Likewise, processing of other substrates, a 700 bp dsRNAs or 130 bp dsRNAs strongly differing in the GC content (71 versus 29%), was not dependent on ATP hydrolysis (data not shown). In summary, our findings argue strongly against ATP being required for the cleavage of dsRNA by the human Dicer.

Interactions of Dicer with dsRNA studied on native gels

For more insight into the mechanism of dsRNA cleavage by Dicer, we investigated the formation of complexes between Dicer and dsRNA, using non-denaturing gels. The enzyme was incubated with dsRNA under conditions that do not allow cleavage: either at 4°C in the presence of Mg^{2+}, or at 25°C in the presence of EDTA. As shown in Figure 6, in both conditions, the incubation of 130 bp dsRNA with increasing amounts of Dicer resulted in the formation of complexes whose mobility decreased as a function of increasing protein concentration. The appearance of several distinct bands suggests that more than one Dicer molecule can bind to the 130 bp dsRNA. Taken together with the results shown in Supplementary figures B and C, these findings indicate that binding and cleavage steps of the Dicer-catalysed reaction can be uncoupled by carrying out the incubation at 4°C, or by omitting Mg^{2+}.

We have also carried out native gel analysis of reactions incubated at 37°C in the presence of Mg^{2+}, under conditions that allow cleavage of dsRNA. In parallel, aliquots of these reactions were analysed on denaturing gels. This comparison, performed with dsRNA substrates of different lengths and either native or ProtK-beads-

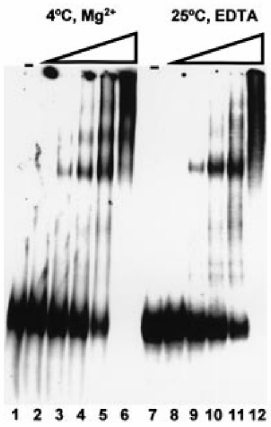

Fig. 6. Complexes between Dicer and dsRNA, visualized on non-denaturing gels. Increasing amounts of Dicer-HisC (0, 2, 6, 20, 60 and 200 ng, respectively, in lanes 1–6 and 7–12) were incubated with the ^{32}P-labelled 130 bp dsRNA either at 4°C in the presence of 2 mM Mg^{2+} (lanes 1–6) or at 25°C in the presence of 2 mM EDTA (lanes 7–12).

activated Dicer, revealed that in general less radioactivity migrates at the position corresponding to siRNA on native gels as compared with denaturing gels (Figure 7A and B). This difference was particularly evident for reactions containing higher concentrations of Dicer, both native and ProtK treated (Figure 7; data not shown). To explain this apparent inconsistency, the native gel lanes representing reactions performed with the 70 bp dsRNA and either native or proteolysed Dicer were cut into 1 cm pieces and the recovered RNA was subjected to electrophoresis on a denaturing gel. As shown in Figure 7C, for the ProtK-treated enzyme, the siRNA-like 22 nt RNA was found in association with a complex marked X in Figure 7, while for the native enzyme, ~50% of ~22 nt RNA was associated with slow-migrating complexes labelled Y and Z. Additional analyses, allowing better separation of complexes Y and Z, indicated that both contain ~22 nt RNA (data not shown). Although further work is needed to establish the identity of complexes X, Y and Z, it is possible that complex Z, reproducibly seen with reactions incubated under cleavage conditions but not following incubations at 4°C or in the presence of EDTA (Figure 6; data not shown), represents Dicer associated with the siRNA product, while complex Y is an aggregate thereof. Association of Dicer with its reaction product would provide a plausible explanation for the apparent low catalytic activity of the enzyme (see Discussion).

Dicer preferentially cleaves dsRNAs at their termini

Previous work has indicated that dsRNAs shorter than ~100 bp are generally much less effective inducers of RNAi *in vivo* and *in vitro* than longer RNAs (Tuschl *et al.*, 1999; Hammond *et al.*, 2000; Parrish *et al.*, 2000; Elbashir *et al.*, 2001b). Likewise, the efficiency of dsRNA processing into siRNAs in *Drosophila* extracts was found to be length dependent, with 29 bp dsRNA being processed very inefficiently (Bernstein *et al.*, 2001; Elbashir *et al.*, 2001b). To provide a possible explanation

for these findings, dsRNAs ranging in length from 30 to 130 bp were compared for their ability to act as substrates for the recombinant Dicer. Processing of dsRNAs of 30, 40, 50, 70 and 130 bp to ~22 nt RNAs occurred with comparable efficiencies (Figure 8A). For 130, 70 and 50 bp substrates, intermediates of processing with lengths diagnostic of the removal of one or more ~22 nt siRNA units were readily visible. In addition, processing of 40 (Figure 8A) and 30 bp RNA (data not shown) resulted in the accumulation of short fragments, whose length (~18 and ~7–8 nt, respectively) suggests that they represent terminal cut-off products. Accumulation of these terminal fragments, as well as intermediates of discrete sizes, is consistent with a step-wise excision of siRNAs from dsRNA termini.

We prepared several derivatives of the 130 bp dsRNA to assess the importance of free RNA termini for the cleavage reaction (Figure 8B). dsRNAtetra bears GAAA tetra-loops at either end. This sequence confers high stability on the loop through an additional base pair between the first guanosine and the third adenosine. dsRNAdna is a 77 bp RNA bearing either 26 or 27 bp RNA–DNA duplexes at the ends. Using a 50 bp RNA–DNA duplex, we verified that Dicer is unable to cleave the RNA–DNA hybrid (results not shown). dsRNArna serves as a control for dsRNAdna and contains 26 and 27 nt terminal RNA oligoribonucleotides in place of oligodeoxynucleotides.

All the RNAs listed above acted as processing substrates (Figure 8B). However, comparison of the kinetics of processing of dsRNAtetra and dsRNAdna with that of dsRNA and dsRNArna revealed that cleavage of the former substrates is initiated with a 5–10 min delay. This could be due to hindrance by the terminal tetra-loops and RNA–DNA duplexes of the interaction with Dicer, necessitating that the enzyme binds, possibly with slower kinetics, to internal regions of the substrate. To address this possibility, dsRNAtetra and dsRNAdna were pre-incubated with the enzyme for 15 min at 4°C, conditions allowing binding but not cleavage of the substrate, prior to the incubation at 37°C. As seen in Figure 8B, the pre-incubation eliminated the delay in siRNA accumulation, indicating that binding of Dicer to the internal sites may indeed be a limiting factor during processing of terminally modified substrates. Consistent with the initiation of cleavages at different internal sites, accumulation of processing intermediates was considerably less pronounced for dsRNAtetra than control dsRNA. However, treatment of dsRNAdna yielded discrete intermediates similar to those seen during cleavage of dsRNA or dsRNArna. Hence, dsRNAdna, despite having non-cleavable RNA–DNA ends, apparently allows a selection of site(s) preferentially used for initiation of processing by the enzyme. Dicer possibly binds to the terminal RNA–DNA duplexes and helps in aligning additional enzyme molecules at the internal RNA–RNA region of the substrate.

dsRNA substrates used in the experiments described above contained blunt termini bearing 5′-triphosphates on both RNA strands. However, siRNA duplexes produced in the Dicer cleavage reaction are expected to contain 2 nt 3′-overhangs bearing 5′-monophosphate groups (Elbashir *et al.*, 2001b), and such terminal structures should be present in processing intermediates acting as Dicer substrates in consecutive rounds of cleavage of long dsRNAs. Using sets of 30 bp dsRNAs with either 5′-monophosphate or 5′-triphosphate ends and either blunt or 2 nt 3′-overhang ends, we found that all these RNAs are processed with similar efficiencies (data not shown).

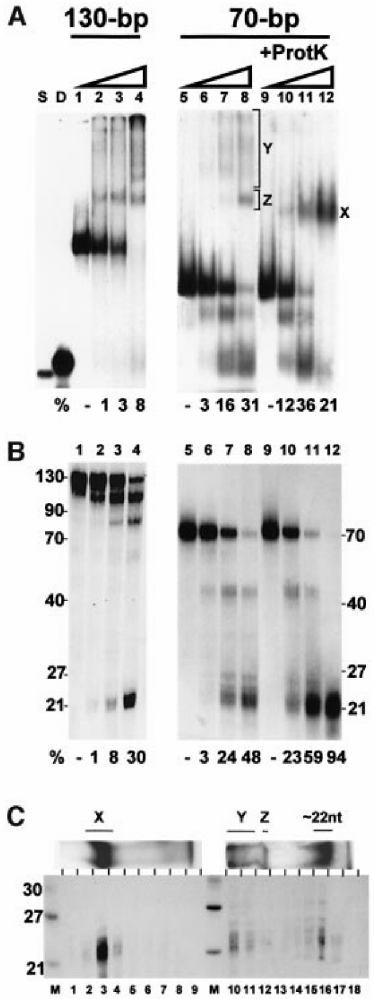

Fig. 7. Analysis of Dicer cleavage reactions on native (**A**) and denaturing (**B**) gels, and demonstration that complexes X, Y and Z contain siRNA products (**C**). (A and B) Increasing amounts of Dicer-HisC (0, 10, 30 and 100 ng), either native (lanes 1–8) or pre-incubated with ProtK beads (lanes 9–12), were incubated with 1 fmol of the ^{32}P-labelled 130 or 70 bp dsRNA at 37°C for 30 min. Halves of each reaction were analysed on two different gel types. Lanes S and D, ss and ds 21 nt siRNA markers. The percentage of radioactivity representing siRNA products is indicated below each lane. (C) Lanes from the non-denaturating gel (shown at the top), representing analysis of reactions containing 70 bp dsRNA and 100 ng of either proteolysed (lanes 1–9) or native (lanes 10–18) Dicer, were cut into 1 cm pieces. RNA was eluted, ethanol precipitated with a tRNA carrier and electrophoresed on an 8 M urea gel. Positions of complexes X, Y and Z are indicated.

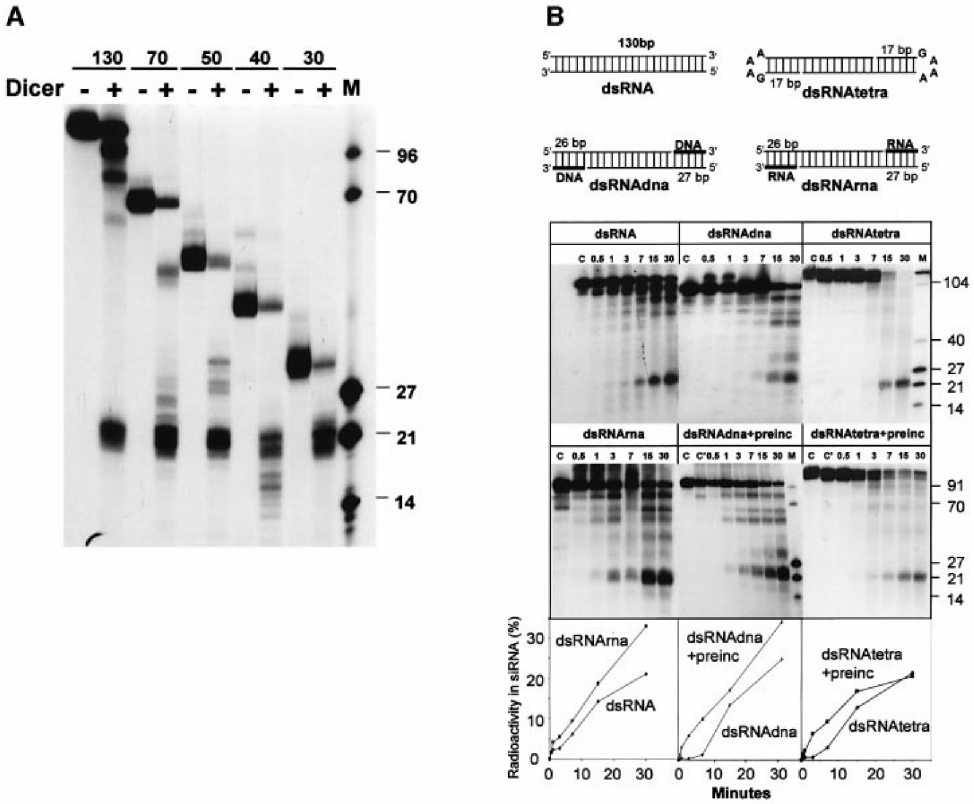

Fig. 8. Processing of dsRNA of different lengths (**A**) or terminal structures (**B**). dsRNAs of indicated length (5 fmol) were incubated either in the absence or presence of 40 ng of Dicer-HisC. Equivalent amounts of counts were applied to the denaturing gel. (B) Terminal GAAA tetra-loops and RNA–DNA duplexes delay cleavage of dsRNA substrates. Upper section, schematic structure of different substrates. Lengths of relevant ds regions are indicated. Lower section, kinetics of processing of different substrates, and effect of 15 min pre-incubation at 4°C on kinetics. The identity of substrates and times of incubation (in min) are indicated at the top. Lanes C, control incubations without Dicer-HisC. Lanes C′, control reactions subjected to only pre-incubation at 4°C. Quantification of the data is shown at the bottom. All kinetic experiments have been reproduced at least twice.

Discussion

This work represents a first comprehensive characterization of the activity of purified Dicer, a key enzyme in RNAi/PTGS. We have established procedures for purifying the N- and C-terminally His_6-tagged recombinant human proteins, Dicer-HisN and Dicer-HisC. Notably, Dicer-HisC was purifed by successive fractionation on Co-containing resin and Ni/NTA beads. Despite a similar principle of affinity separation, only a combination of both matrices reproducibly yielded a high-purity protein; preparations applied to only one type of beads contained numerous impurities. Dicer-HisC, used in most of the experiments described in this work, was prepared under either reducing or non-reducing conditions. The latter preparations were more active then those prepared in the presence of reducing agents. Surprisingly, pre-incubation of recombinant proteins with ProtK and some other proteases markedly stimulated their activity. Activity of the endogenous Dicer present in IPs of P19 cell extracts was also stimulated by pre-incubation with ProtK. A dissection of the mechanisms underlying the effects of proteolysis and reducing agents on enzyme activity requires further experimentation. Provost *et al.* (2002) have also independently succeeded in preparing an active recombinant human Dicer.

Many reaction requirements and properties of Dicer are similar to those established for the RNase III from *E.coli* and yeast. All enzymes require divalent cations for cleavage of the substrate. As with the *E.coli* RNase III (Li *et al.*, 1993), Mg^{2+} could be substituted by Mn^{2+} and Co^{2+}. Mg^{2+} and Mn^{2+}, but not Co^{2+}, are active with the *Schizosaccharomyces pombe* enzyme (Rotondo and Frendewey, 1996). As previously demonstrated for the *E.coli* RNase III (Chelladurai *et al.*, 1993), Mg^{2+} ions are required by Dicer for cleavage of the dsRNA substrate but not for its binding. Likewise, incubation at 4°C allows binding to, but not cleavage of, dsRNA by Dicer (Figure 6). These properties allow separation of the dsRNA binding and cleavage steps of the reaction.

Previous experiments carried out with cell extracts originating from *Drosophila* and *C.elegans*, and with IPs containing the *Drosophila* Dicer, demonstrated that cleavage of dsRNA into siRNAs is strongly stimulated by the addition of ATP (Zamore *et al.*, 2000; Bernstein *et al.*, 2001; Ketting *et al.*, 2001; Nykänen *et al.*, 2001).

However, ATP appeared not to have much effect on dsRNA cleavage in the mammalian system (Billy *et al.*, 2001). We have now measured the activity of different preparations of the recombinant Dicer, and also IPs prepared from extracts of mouse P19 and human HeLa cells with anti-Dicer Abs. Cleavage of dsRNA was never stimulated by addition of ATP. Cleavage was also not affected by the addition of an excess of non-hydrolysable ATP analogues or treatments performed to eliminate any residual ATP.

A role for ATP in siRNA formation, demonstrated for *Drosophila* and *C.elegans* extracts, has generally been linked to the ATPase/helicase domain of Dicer and its hypothetical function in either moving the enzyme along the dsRNA or a local unwinding of the substrate, possibly required for the cleavage (Bernstein *et al.*, 2001; Ketting *et al.*, 2001; Nykänen *et al.*, 2001; Hutvágner and Zamore, 2002a). We found that the K to A change in the P-loop of the Dicer ATPase/helicase domain, a mutation known to eliminate or at least strongly inhibit nucleotide binding and activity of enzymes containing related domains (Gross and Shuman, 1998; Levin and Patel, 1999), has no effect on dsRNA cleavage by Dicer. Although this result provides an additional argument against the requirement of ATP for the dsRNA cleavage, it has to be interpreted with caution. To date, we have been unable to demonstrate ATPase activity in recombinant Dicer preparations, and hence to directly assess functionality of the K70A mutation. Additional factors may be required to activate the ATPase-like domain of Dicer. Another possibility is that, despite employing different forms of recombinant proteins and purification protocols, none of the Dicer preparations used in this work was fully functional.

Some additional observations may serve as arguments against the dsRNA cleavage *per se* being ATP dependent. First, processing of various substrates, including perfectly complementary dsRNAs, by RNase III, a 'prototype' enzyme of Dicer, does not require ATP (reviewed by Nicholson, 1999). Secondly, a putative homologue of Dicer in *Dictyostelium* is apparently devoid of the ATPase/helicase domain. In contrast, a domain with homology to the Dicer ATPase/helicase domain is present in RdRP-like proteins of *Dictyostelium*, suggesting that ATP may be required downstream of the dsRNA cleavage reaction (Martens *et al.*, 2002). Detailed analysis of the ATP requirement for RNAi in *Drosophila* extracts identified two ATP-dependent steps in the pathway. In addition, ATP was utilized for maintenance of 5′-phosphates on siRNAs (Nykänen *et al.*, 2001).

If ATP is not needed for dsRNA cleavage itself, what other steps leading to the increased siRNA yields, as demonstrated in *Drosophila* and *C.elegans* extracts, could be influenced by this factor? ATP hydrolysis, catalysed by either Dicer or another protein interacting with it, could be required for release of the siRNA product from the enzyme and its transfer to the downstream RNAi component, such as a RISC nuclease. Alternatively, and possibly in conjunction with the product release, ATP hydrolysis could help in a structural rearrangement of Dicer required for a next round of substrate binding and catalysis. Our findings that significantly less radioactivity migrates at a position corresponding to siRNA on native as compared with denaturing gels, and that the deficit of the siRNA product on native gels is accompanied by accumulation of complexes probably representing Dicer (or its ProtK-generated fragment) associated with siRNAs (Figure 7), support the first scenario. This possibility is further strengthened by the recent observation of Hutvágner and Zamore (2002b) that pre-let-7 RNA, which must undergo Dicer-catalysed processing in the *Drosophila* extract, is considerably more effective in targeting degradation of a model RNA than the preformed siRNA-like let-7 duplex. The structural rearrangement explanation is supported by our observations that the activity of different recombinant Dicer preparations, and of the endogenous protein, is stimulated by proteolysis (Figure 3), and that the proteolysed enzyme becomes active at 4°C (Supplementary figure C). The catalytic site is possibly occluded by another domain of Dicer, and access to it entails a temperature-dependent rearrangement, a requirement at least partially alleviated by proteolysis.

We consider that ATP is unlikely to function in translocation of the enzyme along the dsRNA substrate. In *Drosophila* extracts, excision of miRNAs from hairpin precursors is ATP dependent, as is the processing of long dsRNAs (Hutvágner *et al.*, 2001). However, miRNA processing involves excision of only a single 21 nt miRNA segment, with no obvious need for Dicer movement along the substrate. In this case also, ATP is more likely to be required for product release and/or a conformational change in the enzyme. Consistent with this possibility, we have found that excision of let-7 RNA from its precursor by the recombinant human Dicer does not require ATP (our unpublished results).

In conclusion, we offer two possible explanations for the apparent discrepancy between the results obtained with mammalian and lower metazoan systems. According to the first, ATP plays a role in increasing the efficiency of siRNA formation in all metazoa by promoting a structural rearrangement of the enzyme and/or product release. With mammalian cell extracts and purified human Dicer, as used in this work, siRNA yields may be ATP independent because these preparations lack some RNAi reaction components, possibly present in extracts of *Drosophila* and *C.elegans*. This explanation would be consistent with the apparent low catalytic efficiency of the mammalian enzyme. A second possibility is that mechanisms of dsRNA cleavage by Dicer are fundamentally different in mammals and lower metazoa. For example, *Drosophila* and *C.elegans* enzymes might cleave dsRNA by a processive mechanism requiring ATP hydrolysis, while the mammalian Dicer functions distributively and independently of ATP. Comparison of properties and factor requirements of purified Dicer proteins from *C.elegans*, *Drosophila* and mammals should help to distinguish between these alternatives.

The accumulation of processing intermediates having lengths diagnostic of the removal of one or more ~22 nt siRNA units, and experiments performed with substrates containing terminal tetra-loops and RNA–DNA duplexes, indicate that recombinant Dicer preferentially cleaves dsRNAs at their termini. A similar conclusion can be drawn from experiments carried out with P19 and HeLa cell IPs. The ATP-dependent formation of discrete size intermediates, supporting a stepwise excision of siRNAs from dsRNA termini, was also recently observed in

C.elegans extracts (Ketting *et al.*, 2001). Previous experiments performed with *Drosophila* extracts did not reveal any cleavage intermediates, suggesting that *Drosophila* enzyme cuts dsRNA by a highly processive mechanism (for references, see Zamore, 2001). Nevertheless, it is also very likely that the *Drosophila* enzyme processes dsRNA progressively from the ends since the mRNA cleavage during RNAi in *Drosophila* extracts occurs at characteristic ~22 nt intervals (Zamore *et al.*, 2000).

The demonstrated preference of Dicer to cleave dsRNAs at their termini may be of physiological significance. Such a property might prevent accidental cleavage of extended hairpins located in internal regions of mRNAs and other cellular molecules (Morse *et al.*, 2002). However, as investigated in this work, Dicer is also able to initiate processing, though less effectively, at internal sites of dsRNA substrates. This is consistent with the established function of Dicer in excision of miRNAs from inner regions of hairpin precursors (Grishok *et al.*, 2001; Hutvágner *et al.*, 2001; Ketting *et al.*, 2001).

We have found that dsRNAs as short as 30 bp act as efficient substrates of the purified human Dicer. Previous work with *Drosophila* extracts indicated that 29 bp dsRNAs are very inefficiently cleaved to siRNAs in embryo extracts (Elbashir *et al.*, 2001b), but more recently Paddison *et al.* (2002) demonstrated a very effective processing, in extracts originating from suspension cells, of 29 bp and even shorter hairpins. Since 30 bp dsRNAs are cleaved by Dicer only once, they will be useful for a direct assessment of structural features required for RNA processing by the enzyme. Initial experiments performed with a selection of 30 bp substrates, either terminally or internally labelled, demonstrated that neither the recessed nature of the 5′ end, nor its phosphorylation status, is important for the efficient cleavage (our unpublished results). They also indicated that products generated by the recombinant Dicer contain 3′-protruding ends, a feature expected for the reaction catalysed by the RNase III-like enzyme, and also consistent with the findings that siRNA duplexes containing 2 nt 3′-overhangs are optimal effectors of RNAi both *in vitro* and *in vivo* (Caplen *et al.*, 2001; Elbashir *et al.*, 2001b,c).

Materials and methods

Dicer cDNAs and expression vectors

A cDNA HH03019 (DDBJ/EMBL/GenBank accession No. AB023145), encoding the human Dicer amino acids 401–1922, cloned in *Sal*I and *Not*I sites of pBluescript II SK+, was obtained from N.Kusuhara, Kazusa Research Institute, Chiba. A region corresponding to Dicer positions 1–1494 was amplified by RT–PCR, using total HeLa cell RNA and primers ACGCGTCGAC*ATG*AAAAGCCCTGCTTTGC (sequence-based DDBJ/EMBL/GenBank accession No. AB028449) and CTGAATTCTGCTTCCATCGTG (*Sal*I and *Eco*RI sites are underlined), following the Superscript II RNaseH–reverse transcriptase protocol (Life Technologies). The RT–PCR fragment was cut with *Sal*I and *Eco*RI, and cloned into HH03019 (pre-digested with *Sal*I, and partially with *Eco*RI) to yield pBS-Dicer. The cDNA insert of pBS-Dicer, and also PCR-generated and otherwise manipulated inserts of plasmids described below, have been sequenced. The ATG selected as an initiation codon for plasmid constructions is an in-frame ATG of the Dicer mRNA 5′ leader. It is preceded by six ATGs, which are either not in-frame or are followed by termination codons. The selected ATG (context TGA*ATG*A) is followed 30 bp downstream by a better context ATG (sequence AGC*ATG*G), which may potentially act as an initiator.

To add a His_6 tag to the Dicer C-terminus, the downstream *Bam*HI–*Not*I fragment in pBS-Dicer was replaced by the equivalent fragment (obtained by PCR, using appropriate primers) containing the sequence CATCACCATCACCATCAC upstream of the stop codon, yielding pBS-Dicer/HisC. Dicer and Dicer-HisC cDNAs were recloned from pBS plasmids into the pENTR1A vector, using the Gateway system (Life Technologies), yielding pENTR-Dicer and pENTR-Dicer-HisC. The non-tagged Dicer cDNA was then switched to the pDEST10 vector by the reaction catalysed by LR Clonase mix (Life Technologies), yielding pDEST10-Dicer, used for expression of the N-terminally His_6-tagged Dicer-HisN. Dicer-HisN contains an additional 31 amino acids (DYDIPTTENLYFQGITSLYKKAGFKGTNSVD), encoded by the vector, positioned between the His_6 tag and the N-terminus. Dicer-HisC cDNA was switched to the pDEST8 vector for expression of the C-terminally His_6-tagged Dicer-HisC. Bacmids for insect cell transformation were generated using the Bac-To-Bac Baculovirus Expression System (Life Technologies). The Dicer P-loop K70A mutant was generated in pBS-Dicer/HisC, using a Quikchange site-directed mutagenesis kit (Stratagene).

Preparation of recombinant Dicers

For protein overexpression, insect Sf9 cells were infected with a virus (virus/cell ratio 1:1) and collected 3 days later. Cells (~4×10^7) were resuspended in 4 ml of buffer W100 (Tris–HCl pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10% glycerol and 0.5% Triton X-100) containing 1× protease inhibitor mix without EDTA (Roche) and broken by passing through a 0.45 × 10 gauge needle. Lysates were centrifuged at 17 000 *g* for 10 min and then at 58 000 r.p.m. (SW60 Beckman rotor) for 1 h. For purification of Dicer-HisC, the final supernatant was mixed with the Talon affinity resin (400 µl; Clontech) pre-washed with buffer W100. After 3 h at 4°C, the resin was washed with buffer W100 (3 × 4 ml), buffer W800 (buffer W100 containing 800 mM NaCl; 3 × 4 ml) and then again with buffer W100 (3 × 4 ml). Protein was eluted with 4 ml of buffer W100 containing 40 mM imidazol. The eluate was directly added to 400 µl of Ni/NTA beads (Qiagen) pre-washed with buffer W100, containing 40 mM imidazol and 1 mM DTT instead of β-mercaptoethanol. Following 3 h at 4°C, the beads were washed similarly to Talon beads and the protein eluted with 2 ml of buffer W100 containing 100 mM imidazol and 1 mM DTT. The eluate was dialysed against buffer D (as buffer W100 but containing 50% glycerol, 0.1% Triton X-100 and 1 mM DTT) (2 × 2 h, 1 l each time) and stored at –20°C. To eliminate any Ni or Co ions potentially released from affinity beads, aliquots of purified Dicer were subjected to additional extensive dialysis against buffer D containing 20 mM EDTA, followed by dialysis against buffer D containing 1 mM EDTA. This treatment had no effect on either cleavage or ATPase activity of Dicer.

For purification of Dicer-HisC under non-reducing conditions, an identical procedure was used, except that all buffers were devoid of DTT or β-mercaptoethanol.

Purification of Dicer-HisN was carried out under reducing conditions and all buffers contained 1 mM DTT. Initial steps were identical as in the purification of Dicer-HisC, except that a high-speed centrifugation and Talon steps were omitted. The 17 000 *g* supernatant was directly mixed with pre-washed Ni/NTA beads. After 12 h at 4°C, the beads were packed into a column. The column was consecutively washed with buffers W100, W800 and W100, and the protein eluted with buffer W100 containing 100 mM imidazol. Fractions containing Dicer activity (3 × 1 ml) were collected and the buffer was changed back to W100 using the Millipore Biomax 30K concentrator. The sample was applied to a Mini-Q PC 3.2/3 column (Amersham) equilibrated with buffer W100. The column was washed extensively with buffer W100 and protein eluted with a gradient of 10 ml of buffer W100 and 10 ml of buffer W1000 (buffer W100 containing 1 M NaCl). Fractions (2 × 0.4 ml) corresponding to the peak of activity eluting at ~0.4 M NaCl were collected. They were concentrated, concomitant with a buffer exchange to W100, in a Millipore concentrator. Proteins were stored at –20°C in buffer W100 containing 50% glycerol, 0.1% Triton X-100 and 1 mM DTT. Protein concentration was calculated using Bradford reagent with BSA as a standard.

Cell extracts and IPs

P19 and HeLa cell growth, and preparation of cytoplasmic extracts and IPs with the anti-Dicer Ab D347 coupled to protein A beads, were as described previously (Billy *et al.*, 2001).

dsRNA substrates

Details of their preparation are specified in Supplementary data.

Assays of Dicer activity

dsRNA processing assays (10 µl) contained 30 mM Tris–HCl pH 6.8, 50 mM NaCl, 3 mM $MgCl_2$, 0.1% Triton X-100, 15–25% glycerol and 1 mM DTT when indicated. Amounts of Dicer and other included components and modified conditions are specified in the legends. If not indicated otherwise, reactions contained 3–5 fmol of dsRNA internally labelled with [α-^{32}P]UTP (final sp. act. 30 or 150 Ci/mmol; Amersham), and were incubated for 30 min at 37°C. For ATP depletion, reactions were pre-incubated for 20 min at 37°C with 2 mM (for purified Dicer) or 10 mM (for IPs) glucose and 1 U of hexokinase (HK, Sigma), or with 1 U of CIP (Roche). We have verified, using P19 cell extracts and a bioluminescence ATP assay kit (Sigma), that the procedure with 10 mM glucose decreases ATP to <100 nM. For analysis of activity in IPs, reactions contained 2–5 µl of IP beads. BSA (10 µg) was additionally added to reactions in the experiment shown in Figure 3C. For proteolytic activation, reactions were pre-incubated for 10 min at 37°C, under assay conditions, with the indicated amounts of ProtK (Sigma) prior to dsRNA addition, or 1 µg of Dicer-HisC was pre-incubated with ProtK–agarose beads (Sigma; 3 mg containing 0.8 U of ProtK, for 5 min at 37°C). Reactions were analysed on 10% polyacrylamide/8 M urea gels, which were processed for autoradiography or quantification, using the Storm 860 PhosphorImager (Molecular Dynamics). For quantification, a percentage of radioactivity corresponding to siRNA products was calculated in relation to either the input or a sum of the bands, representing the substrate, intermediates and siRNA products present in individual lanes. For calculation of molar ratios of Dicer to generated siRNA products, the Dicer-HisC purity was assumed to be 75%.

For analysis of complexes on native gels, assays contained Dicer-HisC prepared under non-reducing conditions, 1 fmol of ^{32}P-labelled dsRNA, and either 2 mM $MgCl_2$ or 2 mM EDTA. Samples were electrophoresed (5 V/cm) for 5 h at 4°C on a 4% polyacrylamide gel containing 45 mM Tris–borate pH 8.0 and 0.1% Triton X-100.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

We thank J.Hofsteenge for valuable discussions and P.Provost for exchanging unpublished results. F.A.K. is supported by a long term EMBO fellowship. Friedrich Miescher Institute is a part of Novartis Research Foundation.

References

Bernstein,E., Caudy,A.A., Hammond,S.M. and Hannon,G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.

Billy,E., Brondani,V., Zhang,H., Muller,U. and Filipowicz,W. (2001) Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc. Natl Acad. Sci. USA*, **98**, 14428–14433.

Caplen,N.J., Parrish,S., Imani,F., Fire,A. and Morgan,R.A. (2001) Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl Acad. Sci. USA*, **98**, 9742–9747.

Cerutti,L., Mian,N. and Bateman,A. (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem. Sci.*, **25**, 481–482.

Chelladurai,B., Li,H., Zhang,K. and Nicholson,A.W. (1993) Mutational analysis of a ribonuclease III processing signal. *Biochemistry*, **32**, 7549–7558.

Elbashir,S.M., Harborth,J., Lendeckel,W., Yalcin,A., Weber,K. and Tuschl,T. (2001a) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.

Elbashir,S.M., Lendeckel,W. and Tuschl,T. (2001b) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, **15**, 188–200.

Elbashir,S.M., Martinez,J., Patkaniowska,A., Lendeckel,W. and Tuschl,T. (2001c) Functional anatomy of siRNAs for mediating efficient RNAi in *D.melanogaster* embryo lysate. *EMBO J.*, **20**, 6877–6888.

Grishok,A. *et al.* (2001) Genes and mechanisms related to RNAi regulate expression of the small temporal RNAs that control *C.elegans* evelopmental timing. *Cell*, **106**, 23–34.

Gross,C.H. and Shuman,S. (1998) The nucleoside triphosphatase and helicase activities of vaccinia virus NPH-II are essential for virus replication. *J. Virol.*, **72**, 4729–4736.

Hamilton,A.J. and Baulcombe,D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, **286**, 950–952.

Hammond,S.M., Bernstein,E., Beach,D. and Hannon,G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, **404**, 293–296.

Hannon,G.J. (2002) RNA interference. *Nature*, **418**, 244–251.

Hutvágner,G. and Zamore,P.D. (2002a) RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.*, **12**, 225–232.

Hutvágner,G. and Zamore,P.D. (2002b) A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, **297**, 2056–2060.

Hutvágner,G., McLachlan,J., Pasquinelli,A.E., Balint,E., Tuschl,T. and Zamore,P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838.

Jacobsen,S.E., Running,M.P. and Meyerowitz,E.M. (1999) Disruption of an RNA helicase/RNAse III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development*, **126**, 5231–5243.

Ketting,R.F., Fischer,S.E., Bernstein,E., Sijen,T., Hannon,G.J. and Plasterk,R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C.elegans*. *Genes Dev.*, **15**, 2654–2659.

Knight,S.W. and Bass,B.L. (2001) A role for the RNAse III enzyme DCR-1 in RNA interference and germ line development in *C.elegans*. *Science*, **293**, 2269–2271.

Levin,M.K. and Patel,S.S. (1999) The helicase from hepatitis C virus is active as an oligomer. *J. Biol. Chem.*, **274**, 31839–31846.

Li,H.L., Chelladurai,B.S., Zhang,K. and Nicholson,A.W. (1993) Ribonuclease III cleavage of a bacteriophage T7 processing signal. Divalent cation specificity, and specific anion effects. *Nucleic Acids Res.*, **21**, 1919–1925.

Lipardi,C., Wei,Q. and Paterson,B.M. (2001) RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell*, **107**, 297–307.

Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P. and Nellen,W. (2002) RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol. Biol. Cell*, **13**, 445–453.

Matsuda,S., Ichigotani,Y., Okuda,T., Irimura,T., Nakatsugawa,S. and Hamaguchi,M. (2000) Molecular cloning and characterization of a novel human gene (*HERNA*) which encodes a putative RNA-helicase. *Biochim. Biophys. Acta*, **1490**, 163–169.

Morse,D.P., Aruscavage,P.J. and Bass,B.L. (2002) RNA hairpins in noncoding regions of human brain and *C.elegans* mRNA are edited by adenosine deaminases that act on RNA. *Proc. Natl Acad. Sci. USA*, **99**, 7906–7911.

Nicholson,A.W. (1999) Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol. Rev.*, **23**, 371–390.

Nicholson,R.H. and Nicholson,A.W. (2002) Molecular characterization of a mouse cDNA encoding Dicer, a ribonuclease III ortholog involved in RNA interference. *Mamm. Genome*, **13**, 67–73.

Nykänen,A., Haley,B. and Zamore,P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, **107**, 309–321.

Paddison,P.J., Caudy,A.A., Bernstein,E., Hannon,G.J. and Conklin,D.S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.*, **16**, 948–958.

Parrish,S., Fleenor,J., Xu,S., Mello,C. and Fire,A. (2000) Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol. Cell*, **6**, 1077–1087.

Provost,P., Dishart,D., Doucet,J., Frendewey,D., Samuelsson,B. and Rådmark,O. (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.*, **21**, 5864–5874.

Ray,S., Golden,T. and Ray,A. (1996) Maternal effects of the short integument mutation on embryo development in *Arabidopsis*. *Dev. Biol.*, **180**, 365–369.

Rotondo,G. and Frendewey,D. (1996) Purification and characterization of the Pac1 ribonuclease of *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **24**, 2377–2386.

Schwarz,D.S. and Zamore,P.D. (2002) Why do miRNAs live in the miRNP? *Genes Dev.*, **16**, 1025–1031.

Sijen,T., Fleenor,J., Simmer,F., Thijssen,K.L., Parrish,S., Timmons,L.,

Chapter 3

Single processing center models for human Dicer and bacterial RNase III

Published Manuscript:

Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W (2004)
Cell. 118(1):57-68.

Cell, Vol. 118, 57–68, July 9, 2004, Copyright ©2004 by Cell Press

Single Processing Center Models for Human Dicer and Bacterial RNase III

Haidi Zhang,[1,3] Fabrice A. Kolb,[1,3] Lukasz Jaskiewicz,[1,3] Eric Westhof,[2] and Witold Filipowicz[1,*]
[1]Friedrich Miescher Institute for Biomedical Research
PO Box 2543
4002 Basel
Switzerland
[2]Institut de Biologie Moléculaire et Cellulaire
CNRS
Strasbourg
France

Summary

Dicer is a multidomain ribonuclease that processes double-stranded RNAs (dsRNAs) to 21 nt small interfering RNAs (siRNAs) during RNA interference, and excises microRNAs from precursor hairpins. Dicer contains two domains related to the bacterial dsRNA-specific endonuclease, RNase III, which is known to function as a homodimer. Based on an X-ray structure of the *Aquifex aeolicus* RNase III, models of the enzyme interaction with dsRNA, and its cleavage at two composite catalytic centers, have been proposed. We have generated mutations in human Dicer and *Escherichia coli* RNase III residues implicated in the catalysis, and studied their effect on RNA processing. Our results indicate that both enzymes have only one processing center, containing two RNA cleavage sites and generating products with 2 nt 3′ overhangs. Based on these and other data, we propose that Dicer functions through *intra*molecular dimerization of its two RNase III domains, assisted by the flanking RNA binding domains, PAZ and dsRBD.

Introduction

Dicer is a large endoribonuclease responsible for processing double-stranded RNAs (dsRNAs) to ~20 bp-long small interfering RNAs (siRNAs) acting as effectors during RNA interference (RNAi), and also for excision of microRNAs (miRNAs) from the hairpin precursors. Being a key enzyme for RNAi and for miRNA function, Dicer proteins have been found in all eukaryotes studied to date, with the exception of baker's yeast. The number of genes encoding Dicer-like proteins varies from four in *Arabidopsis* to one in vertebrates (reviewed by Hannon and Zamore, 2003). Mutations in Dicer proteins in different organisms have developmental phenotypes (reviewed by Hannon and Zamore, 2003; Schauer et al., 2002), likely resulting from the compromised formation of miRNAs (Carrington and Ambros, 2003). In zebrafish and mouse, the Dicer-encoding gene is essential (Bernstein et al., 2003; Wienholds et al., 2003).

*Correspondence: filipowi@fmi.ch
[3]These authors contributed equally to this work

Dicers are approximately 200 kDa multidomain proteins. Typically, their domains include a DExH RNA helicase/ATPase domain, the DUF283 and PAZ signatures, two neighboring RNase III-like domains (RIIIa and RIIIb), and a dsRNA binding domain (dsRBD) (Figure 1A). Although no mutagenesis has been performed to date on Dicer proteins, the dsRBD and RNase III domains are most certainly involved in dsRNA binding and cleavage. This is supported by the findings that ~20 bp products of Dicer processing contain 2 nt 3′ overhangs and 5′-p and 3′-OH termini, characteristic features of RNase III-mediated reactions (Elbashir et al., 2001; Nicholson, 2003). Functions of the remaining Dicer domains are not known. The PAZ domain is also found in Dicer-interacting proteins involved in RNAi and miRNA function, referred to as *P*AZ and *P*iwi *D*omain (PPD) proteins (reviewed by Carmell et al., 2002). Structural studies of the PAZ domain of the *Drosophila* PPD proteins Ago1 and Ago2 revealed similarity to the oligonucleotide binding (OB) fold, consistent with the RNA binding activity of the domain (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). The presence of the helicase/ATPase domain could be related to the findings that generation of siRNAs by *C. elegans* and *Drosophila* Dicers is greatly stimulated by the addition of ATP (Bernstein et al., 2001; Ketting et al., 2001; Liu et al., 2003; Nykänen et al., 2001). However, no such effect is observed for the mammalian enzyme (Zhang et al., 2002). Moreover, Dicer of *Dictyostelium* is devoid of the helicase/ATPase domain (Martens et al., 2002).

Irrespective of the specific role of individual domains, Dicer emerges as a very complex and dynamic enzyme, interacting with other cellular proteins. Apart from Ago proteins, shown recently to associate with Dicer directly (Tahbaz et al., 2004), these are two related small proteins, R2D2 of *Drosophila* (Liu et al., 2003) and RDE-4 of *C. elegans* (Tabara et al., 2002), and the *Drosophila* ortholog of the human fragile X mental retardation protein, dFMR1 (Ishizuka et al., 2002). Recent studies with recombinant human and *Drosophila* Dicers revealed some additional properties of the enzyme (Liu et al., 2003; Provost et al., 2002; Zhang et al., 2002). For example, the human enzyme has a strong preference for cleaving siRNAs from the ends of dsRNA substrates. In addition, activity of both the endogenous and recombinant human Dicer is strongly stimulated by proteolysis, suggesting that access to the catalytic center of the enzyme is perhaps regulated by other domains of the protein (Zhang et al., 2002).

Sequence similarity to RNase III was crucial for identification of Dicer as a protein involved in generating siRNAs (Bass, 2000; Bernstein et al., 2001; Billy et al., 2001; Ketting et al., 2001). According to current classification (Blaszczyk et al., 2001; Nicholson, 2003), Dicer belongs to the class 3 enzymes of the RNase III superfamily. The class 1 members include RNases III from bacteria and fungi, and the class 2 encompasses metazoan RNases III, exemplified by Drosha. In contrast to Dicer and metazoan RNases III, prokaryotic and lower eukaryotic RNases contain one RNase III domain (Figure

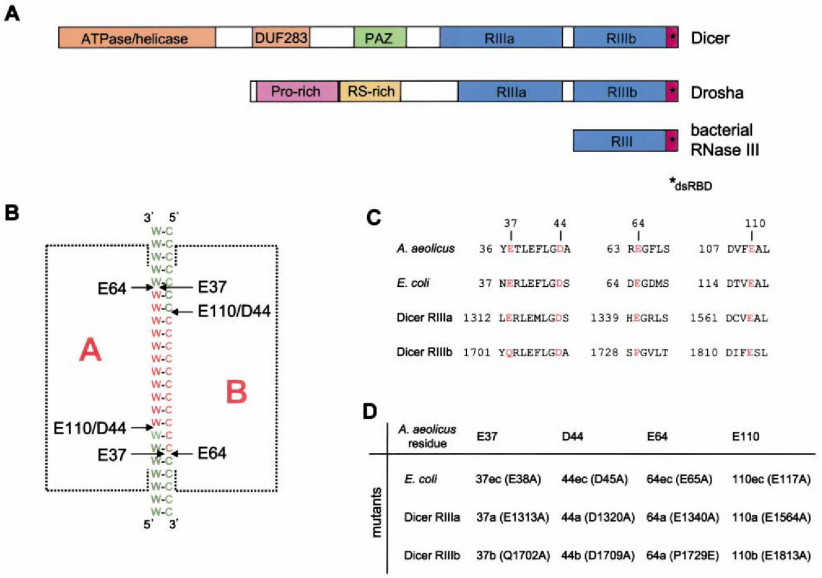

Figure 1. Schemes of the RNase III Superfamily Proteins and Summary of RNase III and Dicer RNase III Domain Sequences and Mutants

(A) Three classes of the RNase III family proteins represented by human Dicer (class 3), human Drosha (class 2), and bacterial RNase III (class 1). Individual protein domains are indicated in different colors.

(B) The postulated mechanism of dsRNA cleavage by Aa-RNase III (Blaszczyk et al., 2001). (A) and (B) represent two subunits of the homodimer enzyme, which was proposed to bind dsRNA (shown as a stack of the Watson-Crick [WC] base pairs) in the intersubunit cleft and to contain two compound catalytic centers. Residues E37 and E64, and D44 and E110 of each center, proposed to be responsible for the generation of products with 2 nt 3′ overhangs, are indicated. Residues E40 and D107, involved with D44 and E110 in coordinating the metal ion, are not shown.

(C) Alignment of conserved regions of Aa-RNase III (SwissProt accession O67082), Ec-RNase (P05797), and Dicer domains RIIIa and RIIIb (Q9UPY3; Zhang et al. 2002), containing residues equivalent to the Aa-RNase III E37 and D44, E64 and E110 (in red).

(D) Summary of single amino acid mutants in RNase III domains of human Dicer and Ec-RNase III.

1A). However, they function as homodimers, as established by biochemical work (reviewed by Nicholson, 2003) and structural studies on the bacterium *Aquifex aeolicus* (Aa) RNase III (Blaszczyk et al., 2001). Based on the X-ray structure of the catalytic domain of Aa-RNase III and mutagenesis of the *Escherichia coli* RNase III (Ec-RNase III), Blaszczyk et al. (2001) proposed a model of the dsRNA cleavage by the enzyme. In this model, shown schematically in Figure 1B, the RNase III dimer contains two compound catalytic centers, positioned at the ends of the postulated dsRNA binding cleft, each cutting two nearby phosphodiester bonds, positioned on opposite RNA strands. Within each catalytic center, two clusters of acidic residues, one comprising Glu40, Asp44, Asp 107, and Glu110, and another residues Glu37 and Glu64, would be responsible for cleavage of individual diester bonds. The former four residues coordinate single metal ions, Mn^{2+} or Mg^{2+}, present in each enzyme monomer. Based on the spacing of the catalytic centers, the model predicts generation of 9 bp products with 2 nt 3′ overhangs, consistent with the size of products generated by RNase III in vitro (Blaszczyk et al., 2001).

The bacterial RNase III structure and activity models raised many speculations regarding the mechanism of dsRNA cleavage by Dicer and the need to explain the size difference—approximately 10 versus 20 bp—in the products of RNase III and Dicer reactions. Is Dicer functioning as a pseudodimer, with RIIIa and RIIIb domains of one molecule interacting with each other, or as a true dimer in which RIII domains associate together intermolecularly in either homologous or heterologous combinations? Irrespective of the model, the size difference between RNase III and Dicer products was suggested to be due to inactivation of one of the two catalytic centers, with an evolutionarily conserved glutamate changed to proline in the position equivalent to the Aa-RNase III Glu64, a putative catalytic residue (Blaszczyk et al., 2001; Hannon, 2002; Nicholson, 2003; Zamore, 2001; see Figure 1C).

We generated mutations in all human Dicer and Ec-RNase III residues implicated in the catalysis and studied their effect on processing of dsRNA and hairpin substrates. Our results demonstrate that both enzymes have only one dsRNA processing center, containing two catalytic sites and generating products containing 3′-protruding ends. We also studied sedimentation properties of the human Dicer and the effect of mutations in its PAZ and dsRBD domains. Collectively, our data indicate that Dicer functions as an intramolecular dimer of RIIIa and RIIIb domains, assisted by two RNA

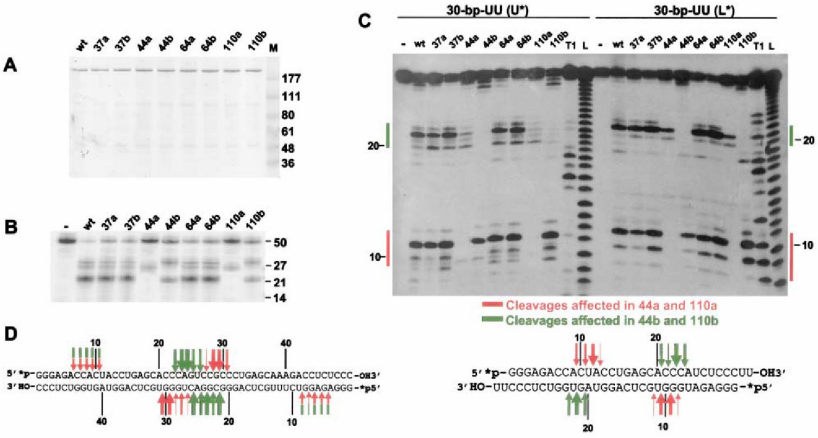

Figure 2. Activity of wt and Mutant Dicer Proteins

(A) SDS-8% PAGE of purified proteins. The gel was stained with GelCode Blue Stain (Pierce). Lane M, protein size markers (in kDa).

(B) Processing of the internally ^{32}P-labeled 50 bp dsRNA. Lane "-",dsRNA incubated without addition of the protein. Positions of oligoribonucleotide markers are indicated.

(C) Processing of the 30 bp dsRNA containing 2 nt 3′ overhangs and ^{32}P-labeled at the 5′ end of either upper (30 bp U*) or lower (30 bp L*) strands. Lanes T1 and L, RNase T1 and alkaline ladders. Positions of cleavages affected in 44a and 110a, and 44b and 110b mutants are indicated in red and green, respectively and marked in the scheme in the lower image. Thickness of arrows represents cleavage intensity.

(D) Schematic representation of processing of the 50 bp dsRNA substrate ^{32}P-labeled at the 5′ end of either the upper or lower strand. For the autoradiogram of the gel, see Supplemental Figure S3 available on *Cell* website. Cleavages marked with the composite red/green arrows represent secondary events occurring only when processing of both strands in the central region of the substrate has taken place (see text and Figure S3).

binding domains, dsRBD and PAZ, the latter domain being likely involved in the recognition of the 3′-overhang end.

Results

Processing of dsRNAs by Dicer RNase III Domain Mutants

For insight into the mechanism of Dicer cleavage we generated single amino acid mutations in its residues equivalent to the Glu37, Asp44, Glu64, and Glu110, which have been proposed to comprise the catalytic centers in Aa-RNase III. With the exception of Dicer residue Pro1731, which was substituted by glutamine, all other residues were changed to alanine. To simplify mutant description and to facilitate comparison of the Aa-RNase III, Ec-RNase III, and human Dicer data, we follow the *Aquifex* protein numbering and refer to the two RNase III domains of Dicer as RIIIa and RIIIb. Accordingly, mutations in Dicer residues Glu1313, Asp1320, Glu1340, and Glu1652 are referred to as 37a, 44a, 64a, and 110a, respectively, and mutations in residues Gln1702, Asp1709, Pro1729, and Glu1813, as 37b, 44b, 64b, and 110b, respectively (for summary of all RIII domain mutations, see Figure 1D). Mutant proteins were overexpressed in insect cells and purified as described previously (Zhang et al., 2002), yielding preparations of comparable purity (Figure 2A).

Activity of wild-type (wt) and mutant proteins was first tested with the 50 bp and 70 bp internally ^{32}P-labeled dsRNA substrates, which were previously shown to undergo effective processing by the recombinant human enzyme. Surprisingly, four of the mutant proteins (37a, 64a, 37b, and 64b) processed the substrates into 21 nt fragments with an activity comparable to that of the wt Dicer (Figure 2B, Supplemental Figure S1 available at http://www.cell.com/cgi/content/full/118/1/57/DC1, and data not shown). The remaining mutants were considerably less active. With the 50 bp substrate, which gives a clearer readout, mutants 44b and 110b yielded lower levels of ~21 nt RNAs and also products of ~29 nt, while mutants 44a and 110a only generated products having length of 25–27 nt. Notably, each pair of mutants appeared to generate a subset of products formed by the wt enzyme (Figure 2B, and see below).

We also generated three sets of double mutants in Dicer RIII domains. Mutants 44a110a and 44b110b combine mutations 44 and 110 in Dicer domains RIIIa and RIIIb, respectively. Mutants 44ab and 110ab combine mutations in equivalent positions in each Dicer RIII domain, and mutants 37ab and 64a37b are combinations of mutations which had no apparent effect on Dicer activity. 44ab110ab is a quadruple mutant combining mutations 44a110a and 44b110b. Mutants 44ab and 110ab, and the quadruple mutant were found to be completely inactive, while other double mutants mimicked precisely the activity of the single amino acid mutants from which they were derived; identical results were obtained for both the 70 bp (Supplemental Figure S1 available on *Cell* website) and the 50 bp (data not shown) substrates.

We have demonstrated previously that Dicer preferen-

tially cleaves off siRNAs from the termini of dsRNA substrates (Zhang et al., 2002). We took advantage of this observation to map cleavage sites introduced by the wt and mutant proteins in the 30 bp substrate. The 30 bp dsRNA can be cleaved by Dicer only once yielding the ~21 nt siRNA-like products and the cut-off fragments (Zhang et al., 2002; and our unpublished results). The 30 bp RNA can be accessed by Dicer from either end but the resulting "end-specific" cleavages can be monitored independently from each other when single strands of the substrate are labeled at the terminus. As shown in Figure 2C, incubation of the wt Dicer with the 30 bp substrate containing either the upper or lower strand labeled at the 5′ terminus yielded series of processing products diagnostic of the enzyme approaching the substrate either from the label-containing (20–23 nt fragments) or opposite (9–12 nt fragments) end. Consistent with the analysis of the internally labeled substrates (see Figure 2B), processing of the 5′-end-labeled 30 bp RNA was unaffected by mutations 37a, 64a, 37b, and 64b (Figure 2C). In contrast, mutations 44a and 110a prevented processing at the labeled-end proximal but not distal sites, while the reverse was true for mutations 44b and 110b. This was irrespective of whether the 5′ label was on the upper or the lower strand of the substrate (Figure 2C). The 30 bp substrates used in this experiment contained 2 nt 3′ overhangs at both ends but similar results were obtained with the blunt-ended 30 bp dsRNA (data not shown). Likewise, processing of the terminally labeled 30 bp dsRNA substrate with a different sequence and base composition yielded a very similar cleavage pattern (Supplemental Figure S2 available on *Cell* website).

The most straightforward interpretation of the data presented above is that residues equivalent to the Aa-RNase III Asp44 and Glu110 of both Dicer RNase III domains are part of one processing center responsible for the dsRNA cleavage, with residues of the domain RIIIa being required for the cleavage of one RNA strand and residues of RIIIb for the nearby cleavage of the second strand, resulting in the formation of products with 3′ overhangs. This conclusion is further supported by analysis of terminally labeled 50 bp substrate processing by Dicer mutants (Figure 2D; for autoradiogram of the gel, see Supplemental Figure S3 available on *Cell* website). This analysis also provides an explanation for the differences in the length of processing products of the internally labeled 50 bp dsRNA generated by the RIIIa and RIIIb domain mutants (see Figure 2B). Since inactivating mutations in domain RIIIa, 44a and 110a, prevent cleavage at sites separated by ~21 nt from RNA 3′-hydroxyl ends (sites shown in red in Figure 2D), the only products formed are those resulting from processing at the nearby sites (shown in green) on complementary strands; for the substrate containing blunt ends, as with the 50 bp dsRNA, these products are expected to be approximately 23 to 27 nt in length, consistent with the data shown in Figure 2B. Conversely, the inactivating mutations in the domain RIIIb, 44b and 110b, prevent cleavage at sites close to the center of the dsRNA (sites shown in green in Figure 2D), and only the products of processing at sites on the complementary RNA strands (shown in red in Figure 2D) accumulate; these products are expected to have lengths of approximately 21 and 29 nt, also in agreement with the data in Figure 2B.

Analysis of terminally labeled 50 bp substrates revealed additional sets of cleavages, marked with the composite red/green arrows in Figure 2D. Positions of these cleavages and the fact that they are observed with the wt protein and its mutants 37 and 64, but not mutants 44 and 110, in both RIII domains (see Supplemental Figure S3 available on *Cell* website), indicate that they represent secondary events occurring only when processing of both strands in the central region of the substrate has taken place. The requirement for the primary processing event explains why no labeled products diagnostic of the cleavage at complementary sites on opposite RNA strands could be identified (Supplemental Figure S3 available on *Cell* website).

Processing of Pre-*let-7* RNA by Dicer Mutants

We extended the analysis of Dicer RIII domain mutants to miRNA precursors, which represent another class of cellular substrates of Dicer (Hannon and Zamore, 2003). Long primary transcripts containing miRNA sequences are first processed in the nucleus by Drosha into ~70 nt long hairpins referred to as pre-miRNAs. The pre-miRNAs are then matured by Dicer to ~21 nt miRNAs in the cytoplasm (Lee et al., 2003). We generated the precursor of *let-7* miRNA, expected to correspond to the Drosha cleavage product (Basyuk et al., 2003; Lee et al., 2003; see Figure 3C), by the self-processing of the in vitro hybrid transcript, which contains the hammerhead ribozyme upstream of the pre-*let-7* sequence. Processing of the internally labeled pre-*let-7* RNA by the wt protein yielded the double-stranded siRNA-like product, as established by gel electrophoresis under nondenaturing conditions (Figure 3A). Analysis of processing of pre-*let-7* RNA, terminally labeled at either the 5′ or 3′ end, by different Dicer mutants revealed that the cleavage is unaffected by mutations 37a, 64a, 37b, and 64b, and that mutations 44a and 110a, and 44b and 110b strongly inhibit processing at adjacent sites on the descending and ascending hairpin arms, respectively (Figures 3B and 3C). These data corroborate conclusions derived from the dsRNA substrate experiments. Collectively, they demonstrate that residues equivalent to the Aa-RNase III Asp44 and Glu110 of both Dicer RIII domains are part of one processing center containing two RNA cleavage sites functioning independently of each other, and that amino acids equivalent to Aa-RNase III residues Glu37 and Glu64 are not essential for the catalysis. The data also indicate that during both the pre-*let-7* RNA (Figure 4) and dsRNA (e.g., see Figure 2D) processing, Dicer accesses the substrate in a polar fashion, with the RIIIa domain always cutting the RNA strand bearing a 3′-hydroxyl, at approximately 21 nt from the end.

Activity of the *E. coli* RNase III Mutants

The Dicer mutagenesis and dsRNA and pre-miRNA processing data are inconsistent with the proposed model of dsRNA cleavage by the bacterial RNase III. They are also in conflict with some previous experimental data indicating that the Ec-RNase III residues equivalent to

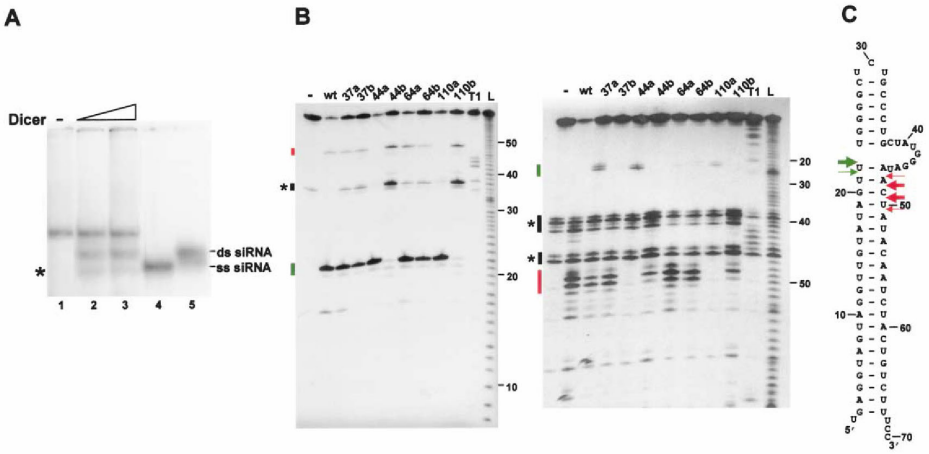

Figure 3. Processing of Pre-*let-7* RNA by Recombinant Dicer Enzymes

(A) Processing of internally ^{32}P-labeled pre-*let-7* RNA by the wt Dicer monitored by nondenaturing PAGE. The substrate (3 fmol) was incubated in the absence (lane 1) or presence (lanes 2 and 3, containing 10 and 30 ng of Dicer, respectively). Lanes 4 and 5, 21 nt single-stranded and double-stranded siRNA markers, respectively. The band marked with the asterisk corresponds the top region of the hairpin.

(B) Mapping of cleavage sites in the 5′-end (left image) or 3′-end (right image) ^{32}P-labeled pre-*let-7* RNA. The RNA was incubated in the absence (lanes "-") or presence of indicated Dicer preparations. Dicer-induced cleavage sites on ascending and descending hairpin arms are indicated by green and red bars, respectively. Unspecific, nonenzymatic cleavages at UA and UG phosphodiesters in the bulge are marked with asterisks.

(C) Secondary structure of pre-*let-7* RNA and positions of cleavage sites as determined in B. Cleavages affected by mutations 44a and 110a, and 44b and 110b, are in red and green, respectively.

the Aa-RNase III Glu37 and Glu64 are essential for the catalysis. Since the latter conclusion was based on a rather indirect assessment of effects of mutations on enzyme activity, namely determining the Ec-RNase III-dependent expression of the λ*N-lacZ* reporter in vivo (Blaszczyk et al., 2001), we reinvestigated activity of Ec-RNase III mutants by more direct methods. Four single amino acid mutants of Ec-RNase III (Glu38Ala, Asp45Ala, Glu65Ala, and Glu117Ala, referred to as 37ec, 44ec, 64ec, and 110ec, respectively; see Figure 1D), in positions corresponding to Aa-RNase III residues Glu37, Asp44, Glu64, and Glu110, were constructed and proteins overexpressed and purified (Figure 4A). Their activity was compared with that of the wt protein, using dsRNA, R1.1 RNA, and R1.1[WC] RNA as substrates. R1.1 and R1.1[WC] RNAs, shown in Figure 4E, represent well-characterized Ec-RNase III substrates. R1.1 RNA corresponds to the phage T7 RNA fragment processed only at the descending arm of the hairpin, and R1.1[WC] RNA is modified such that it is processed at both hairpin arms (Li and Nicholson, 1996; Nicholson, 2003).

Reactions with the internally labeled 70 bp (Figure 4B) and 50 bp (data not shown) dsRNAs revealed that mutations 44ec and 110ec very strongly compromise activity of the enzyme; low levels of products and intermediates were only observed when a large excess of mutant proteins was used. This is consistent with the previous findings that mutation of Glu117 (residue substituted in the 110ec mutant) strongly inhibits activity of the *E. coli* enzyme (Nicholson, 2003; and references therein). In contrast, mutants 37ec and 64ec showed activity comparable with that of the wt protein (Figure 4B).

Terminally labeled 30 bp dsRNA, and R1.1 and R1.1[WC] RNAs were used to map cleavage sites by wt and mutant *E. coli* proteins (Figures 4C–4E). The wt Ec-RNase III cleaved 30 bp dsRNA in the central region and, as expected, processed R1.1 and R1.1[WC] RNAs at the descending arm and at both the ascending and descending arms, respectively. Mutants 37ec and 64ec cleaved all substrates in a manner similar to the wt protein, while mutants 44ec and 110ec were inactive. These findings are consistent with the results of Dicer mutagenesis and indicate that the bacterial RNase III, which functions as a homodimer, also has only one processing center, with the Ec-RNase III residues Asp45 and Glu117 (equivalent to Aa-RNase III residues Asp44 and Glu110) of each monomer contributing to the cleavage of two diester bonds separated by two base pairs and present on opposite RNA strands.

We used the R1.1 single cleavage substrate to compare kinetic parameters of the wt Ec-RNase III with that of its mutants 37ec and 64ec. V_{max} values for all three proteins were similar but K_m values for mutants 37ec and 64ec were, respectively, 5 and 3 times higher than the 137 nM value determined for the wt enzyme (Supplemental Table S1 available on *Cell* website), suggesting a small deficiency in the substrate binding. It is possible that this deficit of mutant proteins is an underlying factor explaining the inability of similar mutants to process the λ*N-lacZ* reporter in vivo as reported by Blaszczyk et al. (2001). (We note that in this work only one of the Glu residues was mutated to alanine; the other was mutated to valine.) Since the λ*N-lacZ* reporter processing was measured following the shift of bacteria to 42°C (Blasz-

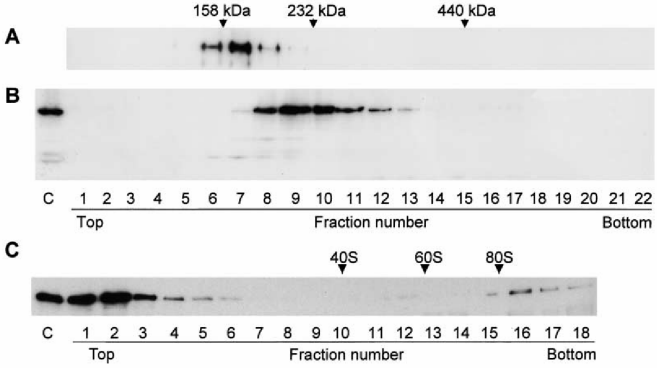

Figure 5. Sedimentation of Purified Dicer (A) and P19 Cell Extract (B and C) on 10%–30% Glycerol Gradients

Position of protein markers (aldolase, 158 kDa; catalase, 232 kDa; ferritin, 440 kDa), and ribosomal subunit and 80S ribosome markers, run in parallel gradients, are indicated. Lanes C, fraction of the input extract. Dicer was visualized by Western blotting. Centrifugation was for 20 hr in (A) and (B), and for 4.5 hr in (C).

Ago1 and Ago2 revealed similarity to the OB fold, consistent with the RNA binding activity of the domain (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). Three Dicer mutant proteins were overexpressed and purified to test the importance of the PAZ domain for RNA processing. In mutants F960A and YY971/2AA, conserved aromatic amino acids, Phe960, and Tyr971 and Tyr972, respectively, are replaced by alanine. Equivalents of the mutated residues were demonstrated to be important for optimal RNA binding of the Ago protein PAZ domains (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003) (for an alignment of PAZ domains of Dicer and PPD proteins, see Supplemental Figure S5 available on *Cell* website). The third mutant is E1036A. Equivalent mutation in *Drosophila* Ago1 had no effect on RNA binding activity (Yan et al., 2003), but in Ago2, the mutation of the corresponding Glu residue to Lys rendered the PAZ protein insoluble (Lingel et al., 2003). In addition, Dicer protein with the C-terminal dsRBD domain deleted, ΔdsRBD, was prepared. When assayed with the 70 bp blunt-ended dsRNA as a substrate, all four mutations markedly decreased Dicer activity. Mutations of the two adjacent tyrosines in the PAZ domain and deletion of the dsRBD had the strongest effect, lowering activity of the enzyme approximately 10-fold (Supplemental Figure S6 available on *Cell* website).

Binding studies with the PAZ domain of Ago proteins indicated that it may recognize the 3′-protruding ends of siRNAs (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). Thus, we compared activity of the wt and mutant Dicer proteins in processing of 30 bp substrates containing either blunt (30 bp) or 2-nt 3′-overhang (30 bp-UU) ends. The wt protein processed the 30 bp-UU RNA with nearly 2-fold higher efficiency than the 30 bp RNA, and each of the three PAZ mutations had a 1.4- to 1.9-fold stronger effect on processing of the 30 bp-UU than on the 30 bp substrate (Figure 6A; for quantification, see the legend). A stronger discrimination between 30 bp and 30 bp-UU substrates was displayed by the ΔdsRBD mutant. While this mutation decreased the cleavage of the 30 bp-UU RNA only 1.9-fold, it decreased 4-fold the cleavage of the 30 bp RNA (Figure 6B). We also compared ability of the wt Dicer and the ΔdsRBD mutant to process the *let-7* RNA precursor containing either 2 nt 3′-protruding (pre-*let-7*) or blunt (pre-*let-7*$_{BL}$) ends. As with the dsRNA substrates, the wt protein showed an ~2.5-fold preference for the pre-*let-7* RNA with the 3′ overhang, and the processing of this substrate was less strongly affected by the dsRBD deletion than the processing of pre-*let-7*$_{BL}$ (Figure 6C). Of note, we found that processing of pre-let-7 RNA by Dicer is not affected by the absence of a phosphate on the recessed 5′-end (data not shown). Taken together, these results indicate that Dicer has a preference for processing of substrates bearing 3′ overhangs; this preference gets accentuated by the deletion of dsRBD. Moreover, the results support a role for the Dicer PAZ domain in the recognition of the substrate 3′-protruding end (see Discussion).

Discussion

Our data indicate that both the human Dicer and the bacterial RNase III contain only a single dsRNA processing center responsible for cleavage of two nearby phosphodiester bonds on opposite RNA strands, generating products containing 2 nt 3′-protruding ends. Residues equivalent to the Aa-RNase III Asp44 and Glu110 of two RNase III domains contribute to this active center, while residues equivalent to Glu37 and Glu64, previously proposed to be required for the RNA cleavage (Blaszczyk et al., 2001), were found to be nonessential. Dispensability of Glu37 and Glu64 equivalents for the catalysis is further supported by the recent finding that mutation of these two residues in the RIIIa domain of the *Drosophila* Dicer does not inactivate the enzyme (Lee et al., 2004). Since in the case of Dicer, two different RNase III domains, RIIIa and RIIIb, contribute to the processing center, analysis of Dicer proteins bearing mutations in individual RIII domains was particularly revealing. This analysis, performed with both dsRNA and pre-*let-7* substrates, demonstrated that mutations of Asp44 and Glu110 equivalents in the RIIIa domain prevent cleavage of one RNA strand but have no effect on cutting the other, while the reverse is true for mutations in the RIIIb domain. Hence, the two catalytic sites present in the RIIIa/RIIIb dimer function independently of each other.

Simultaneous mutations in both Dicer RIII domains (mutants 44a44b or 110a110b) abolished cleavage of each RNA strand. These Dicer double mutants are functionally equivalent to the Ec-RNase III single amino acid

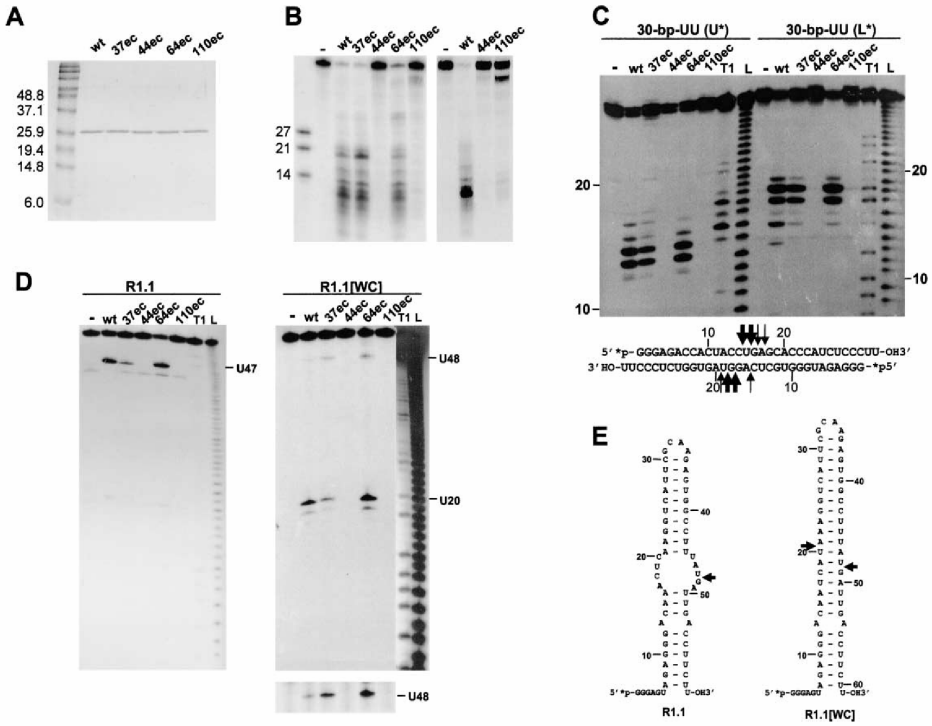

Figure 4. Activity of Recombinant wt and Mutant Ec-RNases III

(A) The purified wt or mutant proteins (0.5 μg) were analyzed by SDS-15% PAGE. Lane M, size markers.

(B) Processing of the 70 bp internally labeled dsRNA (3 fmol). Assays contained either 10 ng (left image) or 70 ng (right image) of the indicated protein.

(C) Processing of the 30 bp dsRNA containing 2 nt 3′ overhangs and terminally labeled on either the upper (30 bp U*) or lower (30 bp L*) strand. Cleavage positions are shown in the bottom image.

(D) Processing of the 5′-^{32}P-labeled R1.1 (left image) and R1.1[WC] (right image) RNAs. RNase T1 and alkaline ladders in the right image represent longer exposure of the same gel. Cleavage positions at U_{47} (R1.1) and U_{20} and U_{48} (R1.1[WC]) are indicated; assignment of cleavages to U_{47} and U_{48} was confirmed by long migration gels. Longer exposure of the U_{48} region is shown at the bottom.

(E) Structure of R1.1 and R1.1[WC] RNAs with marked cleavage sites.

czyk et al., 2001), we have tested whether mutant proteins retain activity at this temperature. The 37ec and 64ec mutants processed dsRNA at 37°C and 42°C with comparable activity (data not shown).

Gradient Sedimentation Analysis of Dicer

As shown above, each of the two RNase III domains of Dicer, RIIIa and RIIIb, contributes amino acids to one processing center generating products with a short 3′ overhang. Thus, Dicer likely operates as a monomer, which functionally represents an intramolecular pseudodimer. We used glycerol gradients to determine whether Dicer sediments as a monomer or a larger complex. Analysis of purified Dicer indicated that the enzyme fractionates at the position corresponding to a molecular mass of ~180 kDa (Figure 5A). The enzyme in a complex with the 30 bp dsRNA substrate having 3′ overhangs sedimented at similar position (Supplemental Figure S4 available on *Cell* website), as did Dicer present in the extract prepared from insect cells overproducing the protein (data not shown). We also analyzed properties of endogenous Dicer in extracts of mouse teratocarcinoma P19 cells (Billy et al., 2001). The protein was found to sediment at ~230 kDa, suggesting that it may be associated with other cellular component(s) of a relatively small size (Figure 5B). Shorter centrifugation of the gradient revealed that an appreciable amount of Dicer also sediments in a region of 80S or larger, suggesting that a fraction of Dicer in P19 cells is associated with ribosomes or another large complex (Figure 5C). The data support a model in which domains RIIIa and RIIIb of a single Dicer molecule interact together to form an intramolecular pseudodimer.

Activity of Mutants in Dicer PAZ and dsRBD Domains

Human Dicer and most of its orthologs in other species contain a PAZ domain positioned N-terminally to RIIIa (Hannon and Zamore, 2003; Schauer et al., 2002). This approximately 150 amino acid long domain is also found in PPD proteins (Carmell et al., 2002), and structural studies of the PAZ domain of *Drosophila* PPD proteins

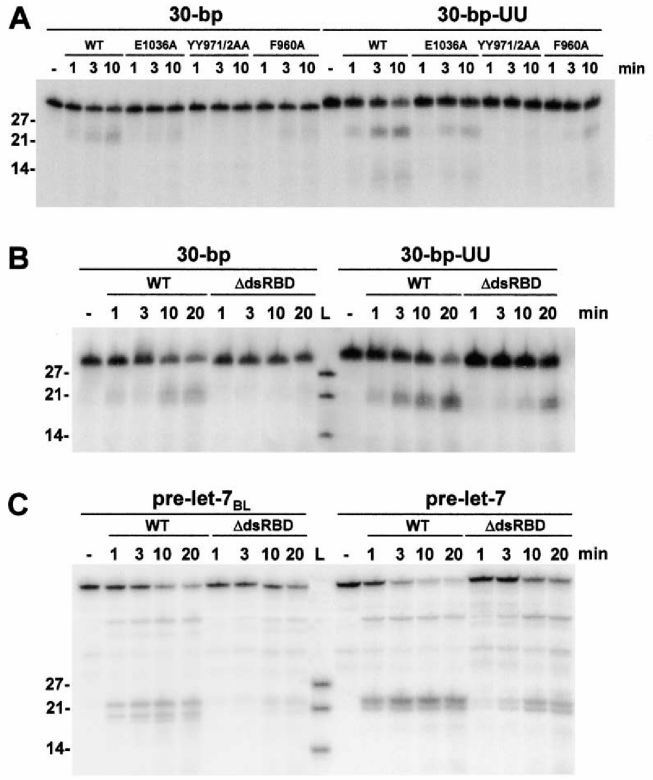

Figure 6. Activity of Dicer PAZ and dsRBD Domain Mutants

(A) Kinetics of processing of the internally labeled 30 bp dsRNA containing either blunt ends (30 bp) or 2 nt 3′ overhangs (30 bp-UU) by the wt Dicer and the PAZ domain mutants. Proteins added to the reaction and times of incubation are indicated at the top. PhosphorImager quantification revealed that the wt Dicer cleaves 1.92-fold and 1.54-fold more of the input 30 bp-UU then 30 bp RNA at time points 3 min and 10 min, respectively (ratios based on averages of three independent experiments similar to those shown in (A) and (B). The PAZ mutations E1036A, YY971/2AA, F960A had 1.4-, 1.4-, and 1.9-fold stronger effect on processing of 30 bp-UU then 30 bp RNA, respectively (values represent averages of 3 and 10 min time points from two independent experiments). Note that the 30 bp-UU RNA contains four additional U residues (see Figure 2C), which are labeled with ^{32}P. Appropriate corrections were made to calculate moles the substrate and the reaction products. (B and C) Comparison of activities of wt and ΔdsRBD Dicers to cleave 30 bp and 30 bp-UU substrates (B) and pre-*let*-7 RNA containing either blunt (pre-*let*-7_{BL}) or 3′-protruding (pre-*let*-7) ends (C). In (B), the ΔdsRBD mutation decreased cleavage of the 30 bp-UU RNA 1.9-fold, and cleavage of 30 bp RNA 4.0-fold (averages from three independent experiments). Reactions in (A–C) contained 40 ng of Dicer proteins and 3 fmol of RNA.

mutants 44ec and 110ec. Since bacterial RNase III functions as a homodimer (Nicholson, 2003), the mutations affect processing of both RNA strands. Notably, mutation 44ec inhibited RNA cleavage more severely than mutation 110ec. It is possible that Asp44 (Asp45 in Ec-RNase III; changed to alanine in the 44ec mutant), which is engaged in metal ion binding via the water molecule in Aa-RNase III (Blaszczyk et al., 2001), acts as a general base that deprotonates the metal bound water (see also Nicholson, 2003).

Based on the results of RNase III and Dicer mutagenesis, and other findings presented in this work, we propose new models for substrate cleavage by both enzymes. In the bacterial Aa-RNase III model (Figures 7A–7C, and Supplemental Figure S7 available on *Cell* website), the dsRNA substrate is rotated by approximately 30° with regard to its position in the model of Blaszczyk et al. (2001). The following arguments support the new placement of the substrate and its availability for the cleavage by the enzyme. (1) In the new position, each of the two scissile phosphodiester bonds is in the proximity of four acidic residues (Glu40, Asp44, Asp107, and Glu110), which coordinate the metal ion (Mn^{2+} or Mg^{2+}) and constitute the catalytic site for the cleavage of one strand. (2) In the model, each catalytic site of the dimer contains a metal ion required for the catalysis. The distance to the scissile bonds (see legend to Figure 7) is consistent with both metal ions participating in the catalysis, each involved in cleaving one of the two RNA strands. In the reported crystal structure of Aa-RNase III, the two metal ions coordinated by four acidic residues in each monomer are the only metal ions identified. No metal ions were found coordinated by residues Glu37 and Glu64, generating speculations about possible repositioning of the substrate in order for the second diester bond cleavage to take place (Nicholson, 2003). (3) The enzyme surface along the new dsRNA placement contains many conserved amino acid residues, likely contributing to substrate binding (Figure 7C; for positions of conserved residues in the alignment of RNase III domains of different proteins, see Supplemental Figure S8 available on *Cell* website).

In the new model, no catalytic role is assigned to residues equivalent to Aa-RNase III Glu37 and Glu64. However, these amino acids are not far from the RNA helix and could, in the presented model, contact the RNA substrate either directly or indirectly via an ion. Such a proximity might explain an increase of K_m values for Ec-RNase III mutants 37ec and 64ec. Notably, Glu37 and Glu64 residues are not 100% conserved in RNase III domains, in contrast to residues equivalent to Asp44 and Glu110. In different members of the RNase III superfamily, Glu, Gln, and Asp residues are present at the Glu37 position, and Glu, Gln, Pro, and Val residues are found at the Glu64 position. For example, in the *Thermotoga maritima* RNase III, the crystal structure of

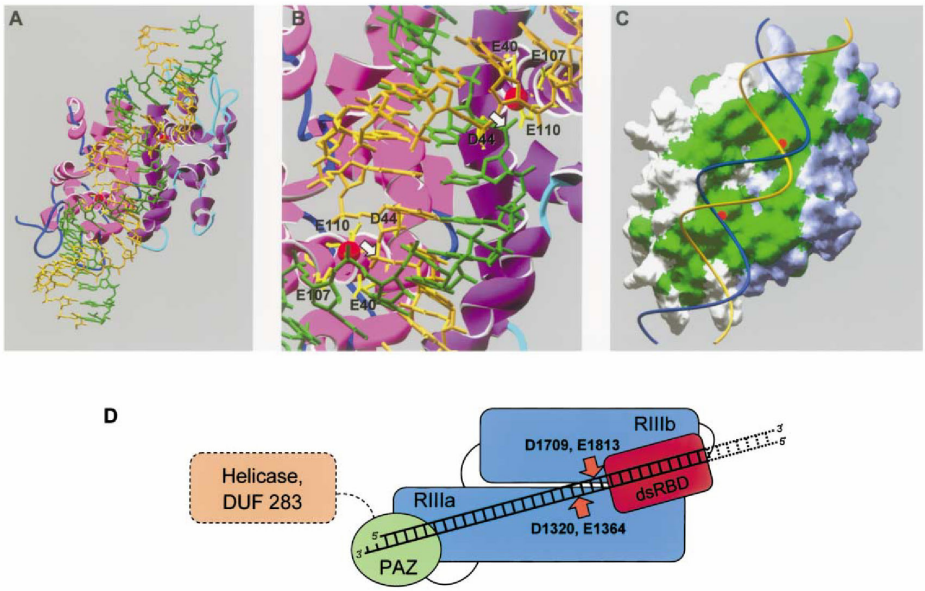

Figure 7. Models of RNase III and Dicer Interactions with dsRNA

(A and B) Frontal view of a complete (A) and an enlarged central region (B) of the Aa-RNase III catalytic domain dimer (Blaszczyk et al., 2001), interacting with the dsRNA helix representing the 30 bp dsRNA shown in Figure 2C. Proteins are represented as ribbons; the two monomers are in different shades of purple. Mn^{2+} ions are as red spheres. Residues coordinating the Mn^{2+} ion in each monomer are shown as yellow sticks. Mn^{2+} ions present in each proposed catalytic site are 21.5 Å apart, while the distance between the two scissile phosphodiester bonds (marked with arrows), across the minor groove, is 17.2 Å; Mn^{2+} ions and phosphorus atoms of cleavable bonds in the model are approximately 5 Å apart. These distances are likely to change upon dsRNA binding to the enzyme. In the same vein, we cannot exclude that additional ions are present in the activated complex. In our model, the dsRNA substrate is rotated by approximately 30° clockwise as compared to its position, parallel to the subunit interface, in the model of (Blaszczyk et al., 2001).

(C) Distribution of conserved amino acids in 75 bacterial and fungal RNase III proteins. Residues conserved in more than 70% of proteins are in green. The modeled-in dsRNA is shown as ribbons. Two monomers are shown in different colors.

(D) A model of dsRNA processing by Dicer. Individual domains of Dicer are in different colors. Helicase/ATPase and DUF283 domains, with no assigned function, are delineated with a broken line. The RIIIa domain is drawn as a larger rectangle than RIIIb to illustrate its larger size. Its placement also illustrates the fact that this domain cleaves the 3′-OH-bearing RNA strand.

which has been recently determined (Protein Data Bank ID 1O0W), Glu64 is replaced by valine. Importantly, a model of the dsRNA-protein complex similar to that for the Aa-RNase III could also be built for the *T. maritima* enzyme (our unpublished results).

Mutagenesis has demonstrated that Dicer domains RIIIa and RIIIb contribute together to one "dimeric" catalytic center, analogous to that of the bacterial RNase III homodimer. Since during gradient sedimentation, Dicer behaves as a monomer rather than a dimer, also when associated with the substrate, we propose that the two interacting RIII domains originate from a single Dicer molecule. Bacterial and fungal RNases III assemble into dimers in the absence of substrate and X-ray analysis has revealed the important contribution of hydrophobic interactions to this process (Blaszczyk et al., 2001). Sequence alignments and modeling of the conserved core regions of Dicer RIIIa and RIIIb domains indicated that interactions similar to the hydrophobic "ball-socket" contacts identified in bacterial enzymes (Blaszczyk et al., 2001; Protein Data Bank ID 1O0W) may also occur between Dicer RIIIa and RIIIb domains (our unpublished observations). Additional support for the intramolecular dimer, containing a single processing center, is offered by the observation that long dsRNA substrates are always shortened successively by ~20 bp siRNA-like units and never by longer segments of dsRNA (for example, see Figure 2 in Zhang et al., 2002). The dependence of secondary processing events on prior cleavages in the center of the 50 bp substrate (Figure 2D and Supplemental Figure S3 available on *Cell* website), also argues for the stepwise excision of the ~20 bp segments by an enzyme with a single processing center.

In most Dicer proteins, the two RIII domains are flanked by the PAZ and dsRBD domains, positioned N- and C-terminally, respectively (see Figure 1A). The PAZ domains of Dicer and PPD proteins were originally thought to mediate interaction between these proteins, but recent work has shown that RIII and Piwi domains of Dicer and PPD proteins are responsible for the interaction (Tahbaz et al., 2004). Recent structural studies of the PAZ domain of *Drosophila* PPD proteins Ago1 and Ago2 revealed its RNA binding potential (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003), raising a possibility that the PAZ domain of Dicer contributes to substrate recognition (Song et al., 2003). Indeed, we

found that mutations of four conserved amino acids, either singly (mutants F960A and E1036A) or in combination (mutant YY971/2AA), have marked inhibitory effect on the activity of the protein. A role for the PAZ domain in dsRNA processing is further supported by the finding that the purified C-terminal fragment of Dicer, encompassing RIIIa, RIIIb, and dsRBD domains, shows no dsRNA cleavage activity (our unpublished results).

We have shown previously that human Dicer preferably cleaves off siRNA products from dsRNA ends and that blocking the ends with RNA tetraloops or RNA-DNA duplexes significantly delays the cleavage reaction (Zhang et al., 2002). Moreover, studies of miRNA precursor processing in vertebrate cells have revealed that Dicer preferentially excises miRNAs from pre-miRNA hairpin intermediates produced by Drosha, possibly due to the availability of proper free ends generated by this enzyme (Lee et al., 2003). The results presented in this work directly demonstrate that Dicer not only requires free dsRNA ends but that it also cleaves substrates with 2 nt 3′-overhangs more effectively than those containing blunt ends. Although for the wt Dicer this preference was found to be relatively small, deletion of the dsRBD domain made the enzyme more dependent on the presence of the 3′-overhang during the processing of both dsRNA and pre-*let-7* RNA substrates. Most probably, the deletion of dsRBD, a domain generally known to mediate unspecific interactions with dsRNA, makes Dicer's interaction with the substrate more dependent on its specific structural features and on protein domains other than its dsRBD. Our findings that mutations in the PAZ domain strongly inhibit Dicer activity, and that this inhibition is greater for the dsRNA containing 3′ overhangs than that with blunt ends, strongly suggest that the PAZ domain of Dicer participates in the recognition of the terminal 3′ overhangs. This conclusion is in agreement with the demonstrated preference of the PAZ domain of *Drosophila* Ago1 and Ago2 for single-stranded RNAs or double-stranded siRNAs with 2 nt 3′ overhangs, as opposed to blunt-ended siRNAs (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). It is likely that the recognition of staggered RNA ends is a major physiological function of the PAZ domain in both Dicer and PPD proteins.

One of the features distinguishing reactions catalyzed by Dicer and the bacterial RNA III is the length of the products, approximately 20 bp and 10 bp, respectively. In previous models, based on the data of Blaszczyk et al. (2001), this size difference was suggested to be due to inactivation of one of the two processing centers, postulated for the RNase III class enzymes by these authors (see Introduction). Our demonstration that both Dicer and bacterial RNase III contain only one active processing center makes this explanation unlikely. Which Dicer domain(s) could be responsible for measuring the distance between the end and the cleavage site? Since mutations in the Dicer PAZ domain were found to have an effect on the processing efficiency but not on the length of the products (Figure 6A, Supplemental Figure S6 available on *Cell* website), we consider it improbable that this domain is a major size-determining factor. It is more likely that the atypical N-proximal RIII domain of Dicer, RIIIa, perhaps assisted by PAZ, measures the distance between the substrate terminus and the cleavage site. RIIIa of all Dicer proteins is much longer than the RIIIb domain and all other typical RNase III catalytic domains, including the N-proximal RNase III domain of Drosha proteins (see Figure 1A).

A model of the dsRNA cleavage by Dicer, incorporating the results of mutagenesis and gradient analyses, and all considerations discussed above, is presented in Figure 7D. The most important feature of the model is the presence of a single dsRNA cleavage center, which is formed by the RIIIa and RIIIb domains of the same Dicer molecule, and which processes the dsRNA approximately 20 bp from its terminus. The PAZ domain recognizes the substrate terminus with the 3′ overhang, and the RIIIa domain, possibly in conjunction with PAZ, measures the distance to the cleavage site. The placement of the RIII domains illustrates the asymmetry of the catalytic region, with RIIIa cleaving the 3′-hydroxyl- and RIIIb cleaving the 5′-phosphate-bearing RNA strand.

The recombinant Dicer utilized in this study is catalytically inefficient. When assayed at excessive (30–100 nM) substrate concentrations, it generates 0.5–1.0 mole of the siRNA product per mole of enzyme in a 30 min incubation (Zhang et al., 2002; our unpublished results). The low activity of purified Dicer is at least partially due to the reaction product, which remains associated with the enzyme (Zhang et al., 2002). Recent studies of the RNA-induced silencing complex (RISC) in *Drosophila* demonstrated that Dicer accompanies siRNA along the RISC assembly pathway and is an essential component of mRNA-cleaving "holo-RISC", sedimenting at ~80S (Lee et al., 2004; Pham et al., 2004; Tomari et al., 2004). The ~80S Dicer-containing complex identified in this study (Figure 5C) may represent the mammalian equivalent of *Drosophila* holo-RISC. The finding that Dicer assembles into RISC along with the siRNA offers an explanation for the low activity of the purified enzyme. In the absence of other RISC components, Dicer would remain associated with its siRNA cargo and be incapable of catalyzing multiple cleavage reactions. The finding that Dicer forms part of RISC (Lee et al., 2004; Pham et al., 2004; Tomari et al., 2004) also provides a rationale for the observation that binding of the PPD protein hAgo2 to Dicer inhibits the RNase activity of the enzyme (Tahbaz et al., 2004). Possibly, one function of Ago proteins, established components of RISC (reviewed by Hannon and Zamore, 2003), is to prevent cleavage of dsRNA by Dicer present in the complex to avoid that a single RISC is associated with more than one siRNA.

Experimental Procedures

Preparation of RNA Substrates

Unless indicated otherwise, all substrates contained 5′-p and 3′-OH ends. dsRNAs, containing either blunt or 2 nt 3′-protruding ends, were obtained by annealing of appropriate RNAs obtained by the T7 polymerase in vitro transcription. Nonradioactive RNAs were synthesized using the Ambion T7 MegaShortScript transcription kit. After transcription, samples were treated with DNase I, extracted with phenol, and purified by denaturating 10% polyacrylamide gel electrophoresis (PAGE). Following dephosphorylation by calf intestine phosphatase (CIP), RNAs were 5′-end-labeled, using T4 polynucleotide kinase and [γ-^{32}P]ATP. The internally ^{32}P-labeled dsRNAs were prepared similarly, except that transcription was performed in the presence of [α-^{32}P]UTP, and [γ-^{32}P]ATP was replaced by 10

mM cold ATP during the 5′-phosphorylation. Complementary RNA strands were annealed at 95°C for 3 min in 20 mM NaCl, transferred to 75°C, and then slowly cooled down to 20°C. The 3′-end-labeling of pre-*let-7* RNA was performed with T4 RNA ligase and [5′-^{32}P]pCp; following ligation, the 3′-phosphate was removed by treatment with CIP, and RNA purified by PAGE. Prior to use, pre-*let-7* RNA was dissolved in water and renaturated by incubation at 90°C for 1 min, followed by incubation at 25°C for 15 min in 30 mM Tris-HCl, [pH 6.8] containing 50 mM NaCl, 2 mM $MgCl_2$, 0.1% Triton X-100, and 10% glycerol. The more AU-rich 30 bp dsRNA used in the experiment shown in Figure S2 was assembled from synthetic oligoribonucleotides (Dharmacon), following their 5′-end labeling.

RNA Processing Assays

Unless indicated otherwise, processing of terminally labeled RNAs was performed as described (Zhang et al., 2002), and incubations were for 15 min at 37°C and contained 0.1 pmol of RNA and either 100 ng of Dicer preparation (0.23 pmol; assuming 50% purity of enzyme preparations) or 0.2 ng (3.8 fmol of dimer) of Ec-RNase III. Unless indicated otherwise, cleavage reactions with internally labeled dsRNAs contained 100 ng of Dicer protein and 125 fmol of the substrate, and were incubated for 30 min. Products were analyzed by denaturing 15% PAGE. RNase T1 ladders of the 5′-end-labeled RNAs were 3′-dephosphorylated by treatment with T4 polynucleotide kinase. Native 4% PAGE was performed as described previously (Zhang et al., 2002). Quantification was done using the Storm 860 PhosphorImager (Molecular Dynamics).

Other Procedures

These are described in the Supplemental Material (available on *Cell* website).

Acknowledgments

We thank N. Hynes and J. Hofsteenge for valuable discussions and V. Brondani for his contribution to the gradient sedimentation analysis. F.A.K. is the recipient of a fellowship from the Human Frontier Science Program. The Friedrich Miescher Institute is a part of the Novartis Research Foundation.

Received: February 9, 2004
Revised: April 28, 2004
Accepted: May 3, 2004
Published: July 8, 2004

References

Bass, B.L. (2000). Double-stranded RNA as a template for gene silencing. Cell *101*, 235–238.

Basyuk, E., Suavet, F., Doglio, A., Bordonne, R., and Bertrand, E. (2003). Human let-7 stem-loop precursors harbor features of RNase III cleavage products. Nucleic Acids Res. *31*, 6593–6597.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat. Genet. *35*, 215–217.

Billy, E., Brondani, V., Zhang, H., Muller, U., and Filipowicz, W. (2001). Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. Proc. Natl. Acad. Sci. USA *98*, 14428–14433.

Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L., and Ji, X. (2001). Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure *9*, 1225–1236.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. *16*, 2733–2742.

Carrington, J.C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. Science *301*, 336–338.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. *15*, 188–200.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244–251.

Hannon, G.J., and Zamore, P.D. (2003). Small RNAs, big biology: biochemical studies of RNA interference. In RNAi A Guide to Gene Silencing, G.J. Hannon, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 87–108.

Ishizuka, A., Siomi, M.C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev. *16*, 2497–2508.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69–81.

Li, H., and Nicholson, A.W. (1996). Defining the enzyme binding domain of a ribonuclease III processing signal. Ethylation interference and hydroxyl radical footprinting using catalytically inactive RNase III mutants. EMBO J. *15*, 1421–1433.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465–469.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921–1925.

Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P., and Nellen, W. (2002). RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase. Mol. Biol. Cell *13*, 445–453.

Nicholson, A.W. (2003). The ribonuclease III superfamily: forms and functions in RNA, maturation, decay, and gene silencing. In RNAi A Guide to Gene Silencing, G.J. Hannon, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 149–174.

Nykänen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309–321.

Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., and Sontheimer, E.J. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. Cell *117*, 83–94.

Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. (2002). Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J. *21*, 5864–5874.

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

Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat. Struct. Biol. *10*, 1026–1032.

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

Tahbaz, N., Kolb, F.A., Zhang, H., Jaronczyk, K., Filipowicz, W., and Hobman, T.C. (2004). Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. EMBO Rep. *5*, 189–194.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831–841.

Wienholds, E., Koudijs, M.J., van Eeden, F.J., Cuppen, E., and Plasterk, R.H. (2003). The microRNA-producing enzyme Dicer1 is essential for zebrafish development. Nat. Genet. *35*, 217–218.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468–474.

Zamore, P.D. (2001). Thirty-three years later, a glimpse at the ribonuclease III active site. Mol. Cell *8*, 1158–1160.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. *21*, 5875–5885.

Chapter 4

General Discussion

The aim of this work was the characterization of the activity of the purified human Dicer, a key enzyme involved in RNAi pathway, as well as getting a further insight into the exact cleavage mode of its RNase III domains.

Our initial data represent a comprehensive characterization of the purified human Dicer (chapter 2). A procedure for purifying the N- and C-terminally His_6-tagged recombinant proteins was established; it consists of a successive fractionation on Cobalt-containing resin and Ni/NTA beads. Dicer processing activity was tested in an *in vitro* assay using radiolabeled dsRNAs and optimized conditions. Many reaction requirements and properties of Dicer are similar to those established for the RNase III from *E.coli* and yeast. All the enzymes require divalent cations for cleavage of the substrate, but not for its binding. Surprisingly, pre-incubation of the recombinant Dicer protein with Proteinase K and some other proteases stimulated its activity. Activity of the endogenous Dicer present in IPs of P19 cell extracts was also stimulated by pre-incubation with Proteinase K. These data suggest that access to the catalytic center of Dicer is perhaps regulated by other domains of the protein, but understanding of the mechanism underlying the effect of proteolysis on Dicer activity requires further experimentation.

Cleavage of dsRNA by purified human Dicer was neither stimulated by addition of ATP, nor affected by the addition of an excess of non-hydrolysable ATP analogues. Likewise, treatments eliminating any residual ATP had no effect on Dicer activity. These results are consistent with data obtained with immunoprecipitates prepared with anti-Dicer Abs from cell extract of mouse P19 and human HeLa cells. Therefore, ATP appears not to have much effect on dsRNA cleavage in the mammalian system. Previous experiments carried out in *Drosophila* and *C. elegans* demonstrated that processing of dsRNA to siRNAs is strongly stimulated by the addition of ATP and it has been proposed that the helicase/ATPase domain of Dicer may promote translocation of the enzyme along the dsRNA, or a structural rearrangement of the substrate required for the cleavage. Hence there is a discrepancy between the results obtained in mammalian and lower metazoan systems. Two possible explanations can be offered for this difference. According to the first, ATP plays a role in increasing the efficiency of siRNA formation in all metazoa by promoting a structural rearrangement of the enzyme and/or product release rather than being required for the cleavage itself. With mammalian cell extracts and purified human Dicer, as used in this work, siRNA yields may be ATP independent because these preparations lack some other RNAi reaction components, possibly present in extracts of *Drosophila* and *C. elegans*. This explanation would be consistent with the apparent low catalytic efficiency of the mammalian enzyme. A second possibility is that

mechanisms of dsRNA cleavage by Dicer are fundamentally different in mammals and lower metazoa. For example, *Drosophila* and *C.elegans* enzymes might cleave dsRNA by a processive mechanism requiring ATP hydrolysis, while the mammalian Dicer might function distributively and independently of ATP. Comparison of properties and factor requirements of purified Dicer proteins from *C.elegans, Drosophila* and mammals should help to distinguish between these alternatives.

Another interesting property revealed from this study is that Dicer preferentially cleaves off siRNAs from dsRNA at termini. Such property of Dicer may have physiological significance. It might prevent accidental cleavage of extended hairpins located in internal regions of mRNAs and other cellular molecules. However, as investigated in this work, Dicer is also able to initiate processing, though less effectively, at internal sites of dsRNA substrates.

The X-ray structure of the bacterial RNase III has been recently established, and a model of how this enzyme binds to dsRNA and processes it into ~ 11 bp products was proposed. Since Dicer belongs to the class 3 enzymes of the RNase III superfamily, the bacterial RNase III structure and its activity model suggested a possible mechanism of dsRNA cleavage by Dicer. Moreover, a model was proposed to explain the size difference between the products of RNase III (~11-bp) and Dicer (~22-bp). Taking advantage of the *in vitro* processing assay with a purified Dicer, it was possible for us to investigate the mechanism dsRNA cleavage by Dicer. To address this problem, the proposed catalytic residues in the two Dicer RNase III domains, as based on the bacterial RNase III model, were mutated either singly or in combinations. Surprisingly, a different picture was revealed by the analysis of our mutants, clearly showing that Dicer works differently than expected. These findings raised the question weather Dicer and bacterial Rnase III might process dsRNA by differently mechanisms. Alternatively, the proposed model of the dsRNA cleavage by bacterial RNase III could simply be wrong. To resolve this issue, mutagenesis was performed on the *E .coli* RNase III. The obtained results are consistent with those obtained with Dicer mutants. Taken together, all these data suggest that the previously proposed model is not correct.

Based on the results of RNase III and Dicer mutagenesis, and other findings reported in Chapter 3, we propose new models for the substrate cleavage by both enzymes. In the bacterial RNase III model, the dsRNA substrate is rotated by approximately 30 degrees with regard to its position in the old model of Blaszczyk et al. (2001). In the new model, two residues previously assigned as important for the catalysis were found to have no role in the cleavage. Based on the results of the mutagenesis, an intra-molecular dimer model is proposed for Dicer. In this model, Dicer RNase IIIa and RNase IIIb domains contribute together to one compound catalytic center.

Further support for this model is coming from the results of gradient sedimentation which showed that Dicer behaves as a monomer rather than a dimer.

We provided in this thesis a comprehensive study of the biochemical properties of the Dicer dsRNA processing activity. However, there are still many questions remaining. For example, how does Dicer measure the distance from the dsRNA end to generate 20-bp product. With the newly revealed structure of the PAZ domain of the Argonaute protein, the answer of this question might be that the Dicer PAZ domain is responsible for the length measurement. The oligonucleotide binding fold-like PAZ domain in Dicer might contribute to substrate recognition. To address this issue, mutations in the Dicer PAZ domain were generated. Mutations of four conserved amino acids were shown to have strong inhibitory effect on the dsRNA cleavage activity of Dicer. However, the mutants still cleave dsRNA at the 20-bp distance from the terminus as the wild-type protein. Additional work on the Dicer PAZ domain is needed. Moreover, it is possible that other domains of Dicer, in addition to PAZ, contribute to the end recognition by Dicer. We noticed that the Dicer RNase IIIa domain is much longer than RNase IIIb and other typical RNase III catalytic domains. Therefore, it is possible that this domain also contributes to the determination of ~20-bp end to the cleavage site distance.

Function of Helicase and DUF283 domain still remains unknown. The double stranded siRNAs must be unwound at a certain stage of the reaction to generate a single-stranded siRNA, which will target the mRNA degradation and the helicase domain of Dicer is an obvious candidate for this step. So far we could not demonstrate ATPase activity in Dicer preparations, suggesting that this domain and its function is possibly regulated by additional factors.

Dicer exists in almost all eukaryotic organisms, but there are differences between proteins from various systems, for example, the ATP requirement discussed above. Another interesting example of the differences comes from *Dictyostelium discoideum*: the putative homolog of Dicer is devoid of the ATPase/helicase domain, and in contrast, a domain with a homology to the Dicer ATPase/helicase domain is present in the RdRP-like protein. This suggests that ATP is required downstream of the dsRNA cleavage reaction in *Dictyostelium*, consistant with our findings for the human Dicer. Moreover, in some organisms like plants and *Drosophila*, two or more Dicers were identified. It was shown genetically and biochemically that different Dicers might play different roles; e.g. in *Drosophila*, only Dicer2 seems to be responsible for siRNA generation. Characterisation of different Dicers might provide further information about different biological

pathways involving RNAi machinery and might help to understand organism-specific differences in RNAi.

Although the two steps of RNAi were extensively studied, the connection in between these steps is still far from clear. Dicer, perhaps with assistance of other factors, must transfer the siRNAs to the RISC complex. Probably the PAZ domain shared by Dicer and the RISC associated Ago proteins plays a role in this hand over, but most likely additional factors are also needed, such as a newly identified Dicer interacting protein R2D2, which bridges Dicer and RISC in *Drosophila*.

As the research of RNAi is progressing, and many different biological pathways are showing connection with the RNAi machinery, Dicer emerges as a key protein for RNAi and miRNA reactions. Dicer is a very complex and dynamic enzyme, interacting also with many other cellular proteins, including key RISC proteins of the Argonaute family. Further studies of Dicer interacting proteins and different complexes, and knock out Dicer in various organisms and tissues will be very information about RNAi machinery and biological role of RNAi.

Besides Dicer, many other protein factors were identified as involved in RNAi. Genetic and biochemical studies indicate that these proteins work together as a complex network to achieve RNA interference and related reactions. Since every single of these factors holds a piece of the jigsaw of RNA interference, it is of a great need to study the function of every individual protein or protein family, as well as to reveal the connection between these factors.

CURRICULUM VITAE

Nationality: CHINA
(permanent resident of Canada)

Address: 7016 rue de st valliers appt.1,
Montreal,Quebec,
Canada, H2S 2R2

Telephone: home: (514) 2743976

Email: kaiteic@hotmail.com

Date of birth
and place: 24th May 1973, Beijing , China

Education: 1999-2004, Ph.D student at Basel University,
Basel, Switzerland
1996-1998, Master student at Tottori University,
Yonago, Japan
1991-1995, undergraduate student at
Beijing Medical University (now, Peking
University Medical school), Beijing, China.

Languages: Chinese, English, Japanese.

Working Experince: Sep.1999 – May.2004, Ph.D student at the laboratory of Prof. Witold Filipowiicz, Friedrich Miescher Institute, Basel
April,1998-Aug,1999 Japan National Followship in the department of Neurobiology

Tottori University, school of Medicine.
March,1996-March 1998, Master student at Gene Research Center, Tottori Unvierstiy
Sep,1994-July 1995, Bachelor thesis in National key Lab of Pharmaceutical chemistry, Beijing Medical University.

Publications:

Kolb F.A., ***Zhang H***., Joronczyk K., Tahbaz N., Hobman T.C. and Filipowicz W.: "Human Dicer: Purification, properties, and interaction with PAZ PIWI domain proteins", *Methods Enzymol*. 2005; 392:316-36. .

Zhang H., Kolb F.A., Jaskiewicz L., Westhof E., Fillipowic W. : "Single processing center models for human Dicer and bacterial RNase III", *Cell*, 2004 Jul 9;118(1):57-68..

Tahbaz N, Kolb F.A., ***Zhang H***., Jaronczyk K., Parker H., Filipowicz W. and Hobman T.C.: " Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer", *EMBO Rep*. 2004 Feb;5(2):189-94

Zhang H., Kolb F.A., Brondani V., Billy E., and Filipowicz W.: "Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP", *The EMBO Journal* (2002), 21, 5875-5885

Billy E., Brondani V., ***Zhang H***., Muller U., and Filipowicz W.: "Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines." *Proc.Natl Acad. Sci. USA*(2001), 98, 14428-14433

Zhang H., Nanba E., Yamamoto T., Ninomiya H., Ohno K., Mizuguchi M., Takeshita K.: " Mutational analysis of TSC1 and TSC2 genes in Japanese patients with tuberous sclerosis complex." *J Hum Genet* (1999) 44: 391-396

Zhang H., Yamamoto T., Nanba E., Kitamura Y., Terada T., Akaboshi S., Yuasa I., Ohtani K., Nakamoto S., Takeshita K., and Ohno K. " Novel TSC2 mutation in a patient with pulmonary tuberous sclerosis: lack of loss of heterozygosity in a lung cyst." *Am J Med Genet* (1999) 82: 368-370

Ohno K., Yakura N., ***Zhang H***., Ninomiya H., Nanba E. " Molecular Epidemiology of Tuberous Sclerosis" *Gann Monograph on Cancer Research 46*, (1999) 53-59

Yamamoto T., Nanba E., Ninomiya H., Higaki K., Taniguchi M., ***Zhang H***., Akaboshi S., Watanabe Y., Takeshima T., Inui K., Okada S.,Tanaka A., Sakuragawa N., Millat G., Vanier MT., Morris JA., Pentchev PG., Ohno K. " NPC1 gene mutations in Japanese patients with Niemann-Pick disease type C." *Hum Genet.* (1999) 271(2) 69-72

Toyoshima M., Hara T., ***Zhang H***., Yamamoto T., Akaboshi S., Nanba E., Ohno K., Hori N., Sato K., Takeshita K. " Ataxia-telangiectasia without immunodeficiency: novel point mutations within and adjacent to the phosphatidylinositol 3-kinase-like domain." *Am J Med Genet* (1998) 75(2): 141-4