KING'S College LONDON
Founded 1829

RNA Interference:

from petunias to a panacea?

Jonathan R.M. Thackray

April 2004

MSc General Biochemistry

and Molecular Biology

Department of Life Sciences,

King's College, London

Table of Contents

Introduction

RNA interference (RNAi) is a natural biological mechanism whereby double-stranded RNA (dsRNA) inhibits gene expression in a highly sequence-specific manner, preventing expression of a single gene, without affecting expression of other genes. The gene is "knocked down", but not actually deleted from the chromosome ("knocked out"). This occurs through the degradation of messenger RNA (mRNA) transcribed from the gene, preventing translation of mRNA into protein; only mRNA whose sequence matches the introduced dsRNA is degraded.

The phrase "RNA interference" was coined in 1998 by Fire, Mello *et al*[1] when they were investigating the effects of injecting a dsRNA mixture of sense and anti-sense RNA into *C. elegans*, trying to suppress gene expression using anti-sense RNA. The idea of using anti-sense DNA or RNA to silence gene expression was not new. Zamecnik and Stephenson[2] in 1978, used an anti-sense oligodeoxynucleotide to silence a specific mRNA and prevent its translation into protein.

Co-suppression was first discovered in petunias in 1990 by Richard Jorgensen

However, Fire and Mello were surprised to see a response which was ten-fold more potent with double-stranded RNA than by using single stranded sense RNA or agnti-sense RNA alone. RNA interference, as observed in *C. elegans*, appeared to be closely related to similar effects that were previously known as co-suppression or post-transcriptional gene silencing (PTGS) in the pigmentation of petunias as discovered in 1990 by Jorgensen[3]. The effects of virus-induced gene silencing (VIGS) in plants and 'quelling' in fungi[4] were also observed. It was suspected that similar cellular mechanisms were involved in these various silencing phenomena.

RNAi is an incredibly potent mechanism, requiring just a few molecules of dsRNA per cell to trigger gene silencing[1]. It appears to be an evolutionary well-conserved biological mechanism, occurring in many organisms, including *Arabidopsis* and other plants, *Drosophila*[5], *C. elegans*[6], *T. brucei*[7], hydra[8], planaria[9], zebrafish[10], mice[11] and human cells.

One has to ask the question why this particular pathway exists at all, and what is its natural role and purpose? Does the cell use it to regulate gene expression in addition to existing mechanisms? Possible roles that it may play include defending cells against RNA viral infection (exogenous threats), suppressing mobilization of transposons (endogenous threats), and regulating expression of endogenous genes in development.

Since RNA interference has only been recently discovered, there are many possible future avenues for application. Its specificity makes it an ideal tool for knocking down single genes for studying gene function or for gene therapy. There appear to be numerous potential clinical and medical applications. For example, Jacque *et al*[12] and have used RNA interference to modulate HIV-1 replication in cells.

There was much excitement when RNA interference was first discovered, that it would provide a simple way to knock down one or more chosen genes and create new phenotypes in any organism with only a day's work. If used in a high-throughput screening set up, could this allow probing of gene function across many genes in an organism at once? Could RNAi be a researcher's dream, and a geneticist's panacea?

As we will see, many hurdles to using RNAi as an effective technique have been overcome, but whilst the phenomena appears simple, many subtleties are involved, and the proteins and biochemical mechanisms have yet to be fully understood. As Gregory J. Hannon[13] notes in his Nature Review paper in 2002, "*We are only beginning to appreciate the mechanistic complexity of this process and its biological ramifications*."

Overview of the RNA interference mechanism

RNA interference has been shown to be a two step process. Although each step happens independently of the other, either step may be used individually or as part of other cellular pathways. Firstly, an enzyme named DICER (or a homolog thereof) cleaves the introduced dsRNA into a number of small, single-stranded RNAs, which are known as short interfering RNAs (siRNAs). These double stranded oligonucleotides are approximately 21-23 nucleotides long, and have an overhang of two nucleotides at the 3' end.

Secondly, the siRNAs which are produced by DICER cleaving the dsRNA, join a RNA endonuclease to form a ribo-protein complex known as RISC (RNA-induced silencing complex), and act as guide RNAs for this complex. The complex appears to specifically target the mRNA that matches the sequence of the siRNA which has bound to the enzyme. When the complex encounters the target mRNA, endonucleolytic cleavage occurs, inducing specific degradation of the mRNA and preventing translation into protein.

This RNA-directed response regulates the expression of a specific gene, in response to introduced dsRNA, whilst other gene expression remains unaffected. If even a single nucleotide is different between the siRNA and the mRNA to be cleaved, then the RNA inteference for that gene being expressed will not occur, or will be massively diminished[14].

Figure 1 shows an overview of how the two stage process occurs, showing the siRNAs with the 2nt overhang being formed by DICER and forming a complex with the RISC enzyme.

Abbreviations used in this document:

nt – nucleotide, bp – base pair, RNAi – RNA interference, mRNA – messenger RNA, siRNA – short interferring RNA, miRNA – micro RNA, dsRNA – double stranded RNA, ssRNA – single stranded RNA, snoRNA – small nucleolar RNA, stRNA – short temporal RNA, PTGS – post transcriptional gene silencing, RdRP – RNA-dependent RNA polymerase, VIGS – virus induced gene silencing.

The cleavage of dsRNA into siRNAs by a DICER enzyme or homolog appears to be a distinct process, and can occur separately from the silencing directed by the RISC enzyme, and is therefore uncoupled from the second stage of mRNA degradation.

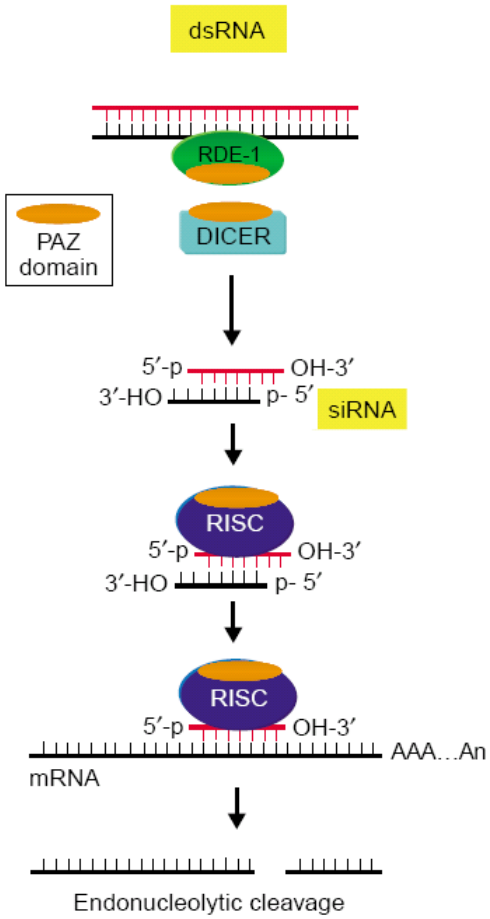

Figure 1. Mechanism of RNA interference (diagram taken from Voinnet, O[50].)

RNA interference is directed and controlled by dsRNA which matches the sequence of the mRNA to be cleaved and degraded, preventing translation into protein. Double-stranded RNA that contains both sense and anti-sense sequences can be introduced exogenously into the cell using a number of methods, which are described in detail later. Alternatively, Paddison *et al*[15] have shown that dsRNA can be synthesised intracellularly by using short 'hairpin' RNAs folded back on themselves, which are then cut by DICER to the correct size for siRNA.

There are a number of proteins that act as specificity factors influencing the progression of each stage of RNA interference. Tabara *et al*[16] have found that in order for the initiation of RNA interference in *C. elegans* to occur, the RDE-1 protein is required to be present, but it is not required for any of the further stages in RNAi. Other mutant *C. elegans* phenotypes missing the rde-2 or rde-3 genes have lost the RNA interference pathway, but also show increased mobilization of endogenous transposons, suggesting that RNA interference can suppress transposon mobilization.

It slices, it dices. Identification of DICER enzyme and homologs

In 1999, Hamilton and Baulcombe[17] determined that short dsRNAs of about 25 nucleotides in length were a key component of RNA interference. They noticed that these were present in plants which were undergoing virus-induced gene silencing, but were not present in those not being silenced. They proposed that these small RNAs may have been synthesised from an RNA template, a longer strand of dsRNA.

In 2001, Bernstein *et al*[18] were the first to identify and name the enzyme in *Drosophila melanogaster* that is responsible for cleaving long dsRNA strands into short interfering RNAs. They named this enzyme DICER after its behaviour of dicing dsRNA into successive short RNAs. Hammond *et al*[19] found that DICER is much more efficient at cleaving long dsRNAs; those with fewer than ~200nt triggered silencing very inefficiently. No cleavage intermediates have thus far been detected *in vitro* or *in vivo*, so DICER appears to cleave dsRNA to siRNA lengths in a single dicing action.

Homologs of DICER have been found in other organisms, and function in a similar way, thus indicating that this is an evolutionarily conserved mechanism. DICER homologs have been found in *C. elegans*, *Arabidopsis thaliana*, *Spodoptera frugiperda* (armyworm), *Neurospora crassa* (fungi), *mus musculus* and humans.

Provost *et al*[20] have cloned and expressed the human DICER enzyme as a protein of mass 218kDa. Tang *et al*[21] found that in *Arabidopsis thaliana*, the Carpel Factory (caf1) gene encodes an ortholog of DICER, and rice genomes appear to encode four different DICER-like proteins, including caf1.

Ketting *et al*[22] have identified the DICER ortholog in *Caenorhabditis elegans*, which is coded for by the dcr-1 gene (K12H4.8), and showed that it is required for functional RNA interference, and that short RNA molecules are involved in the regulation of developmental timing. This was confirmed by Grishok *et al*[23] who inactivated the dcr-1 gene in *C. elegans* and found that RNAi was impeded in dcr-1 defective mutants.

DICER has subsequently been shown to be an RNase III endonuclease. Three major domains are contained within the DICER enzyme: an N-terminal helicase domain, a Piwi/Argonaute/Zwille (PAZ) domain, and dual C-terminal (bidentate) RNase III motifs. Contained within the latter domain is a dsRNA-binding domain (dsRBD). It is possible that the helicase is required to unwind the dsRNA before cleavage can occur. The domains contained within the DICER enzyme are shown in Figure 2, with a scale indicator for 100 amino acids.

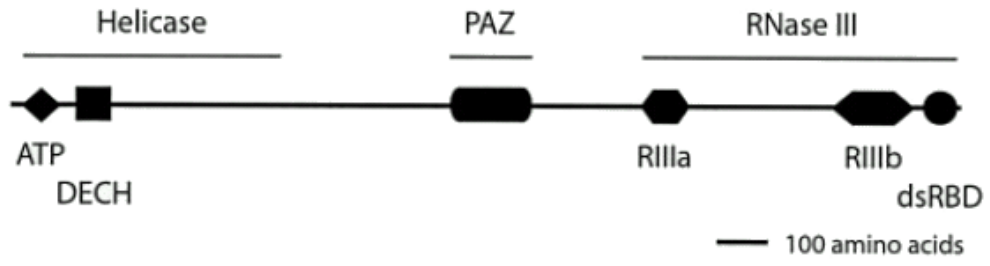

Figure 2. Domains within the human DICER enzyme (From Provost *et al*[20])

The processing of double-stranded RNA by DICER is an adenosine triphosphate (ATP) dependent process. This was shown by Zamore *et al*[33] by providing hexokinase and glucose in excess to result in the depletion of ATP in a *Drosophila* embryo lysate. By converting the ATP to ADP, no ATP was available and RNAi did not occur in the absence of ATP. The addition of creatine kinase and creatine phosphate to the lysate which was depleted of ATP, restored the occurrence of RNA interference, by increasing the amount of ATP available. Nykanen, Haley and Zamore[24] later showed that ATP is also required to unwind the siRNA double-stranded helix when forming the RISC complex.

So how does DICER bind to and execute the dicing action on dsRNA? DICER contains dual RNase III domains, which Blaszczyk *et al*[25] have suggested from crystallographic and modelling studies a mechanism by which dsRNA may be cleaved. The structure of the PAZ domain from the Argonaute2 protein has been determined by Song *et al*[26] using X-ray crystallography. The structure showed that the PAZ domain contains a variant of an OB fold, which is a recognised motif whereby an enzyme can bind single-stranded nucleic acids.

The PAZ domain is also found in the RDE-1 effector protein and the RISC complex, and Song *et al* noted that it may be used to bind the 3' end of siRNAs, both in DICER and in RISC, since the 3' overhang is single-stranded at this point.

Yan *et al*[27] also determined the PAZ domain structure, but this time from the Ago1 (Argonaute1) protein using the technique of NMR spectroscopy, The structure demonstrated that a 5 nucleotide RNA binds to the PAZ domain. The PAZ domain structure is shown in Figure 3.

DICER has also been implicated in the processing of micro RNAs (miRNAs), which are described in detail later.

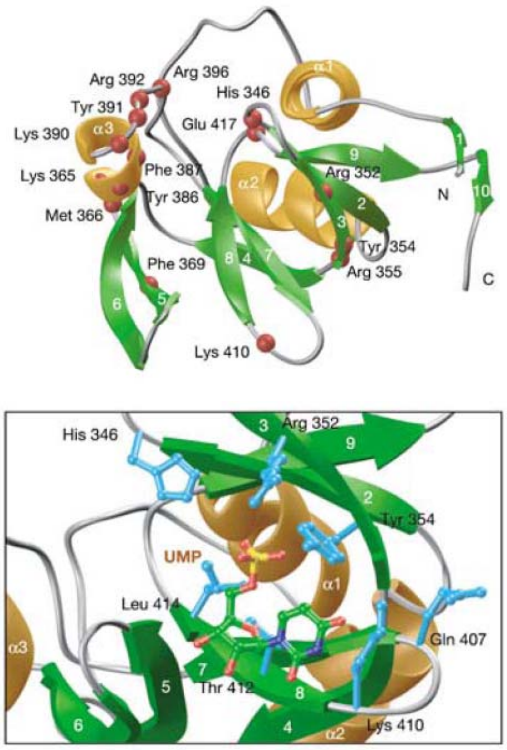

Figure 3. Structure of the Ago1-PAZ domain, and RNA binding site. (From Yan *et al*.)

Formation of the RISC complex

In the second stage of RNA interference, in order to target the specific mRNA for degradation, the siRNAs produced from dsRNA by the DICER enzyme combine with the RISC multicomponent nuclease (RISC stands for RNA induced silencing complex) to form a ribonucleoprotein complex.

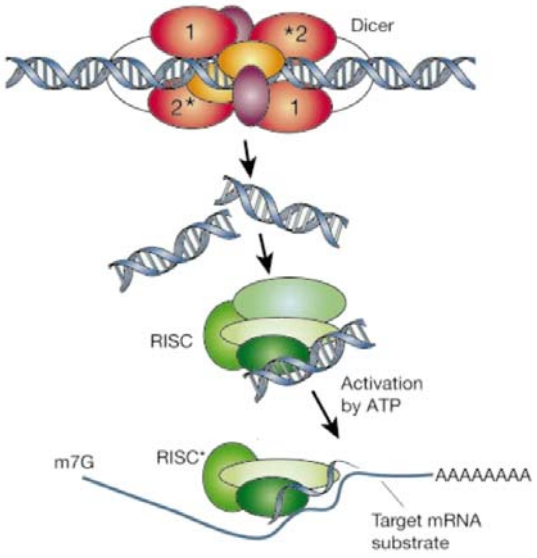

Figure 4. Diagrammatic representation of DICER cleaving dsRNA and RISC degrading mRNA guided by siRNAs, taken from Hannon[13].

The RISC complex is guided by the sequence of the siRNA that has bound, and if a complementary match is made between the siRNA acting as a silencing trigger, and the mRNA to be degraded, then the mRNA is cleaved by RISC endonucleolytically. Hannon *et al*[13] have purified RISC from *Drosophila* S2 cells to yield a ~500kDa ribonucleoprotein complex.

Whilst the approximate function of RISC has been determined, the subunits comprising RISC and precise biochemical mechanisms of the RISC complex are in the process of being determined. How do the siRNAs that join the RISC complex direct the cleavage to occur in the correct place?

Hammond *et al*[28] have discovered one of the subunits in the RISC complex. They purified the RISC complex by the centrifugation of *Drosophila* S2 lysates, where RISC was bound to ribosomes in cell-free extracts. After microsequencing, they discovered that numerous peptides matched a single gene from *Drosophila*. This gene was identified as a homolog of the *rde-1* gene. *rde-1* is a member of the Argonaute family of genes, which has already been shown to be essential for RNAi in *C. elegans*, *Neurospora* and *Arabidopsis thaliana*. They named this new gene *Argonaute2* and tested that the AGO2 protein really was part of RISC using AGO2-

specific antibodies and Western blotting of a chromatography column, which yielded the ~130kDa AGO2 protein. Carmell *et al*[29] have shown that proteins belonging to the Argonaute family (there are ten Argonaute genes in *Arabidopsis* and seven in humans) play many roles, including determining the fate of RNAs which have been processed by DICER, and also affecting developmental control and stem cell maintenance.

Caudy *et al*[30] performed large-scale biochemical purification of *Drosophila* RISC to try to indentify additional RISC subunits, and found an additional two proteins that were co-purified with RISC. One of these is VIG (vasa-intronic gene), which is evolutionarily conserved and has homologs in *C. elegans*, *Arabidopsis*, mammals and *S. pombe*; the other protein found was dFXR, a *Drosophila* homolog of the human Fragile X Mental Retardation protein (FMRP). The precise function of these proteins has yet to be determined within the RISC complex, but some tantalising clues have been uncovered.

Little is currently known about VIG, only having one recognisable motif, an RGG box, which can bind RNA. The Fragile X protein is better characterised; in humans the FMR (Fragile X) protein has been implicated previously in the regulation of gene expression, and has been implicated in RNAi pathways that cause disease. The authors speculate that dFXR, as a subunit of the RISC complex, may be involved in pathways where microRNAs are used for the regulation of other genes via RNAi.

Caudy *et al*[31] have noted that either siRNAs or microRNAs (miRNA) can bind to the RISC complex. The authors subsequently identified another component of the RISC complex, a protein containing multiple staphylococcal/micrococcal nuclease domains and a tudor domain, which they have called Tudor-SN (tudor staphylococcal nuclease). They note that "*tudor-SN is the first RISC subunit to be identified that contains a recognisable nuclease domain, and could therefore contribute to the RNA degradation observed in RNAi*". When all these results are considered together, the RISC complex appears to contain a small RNA (microRNA or siRNA), together with the Argonaute2, Fragile X, VIG and Tudor-SN protein subunits. These subunits are shown in Figure 5.

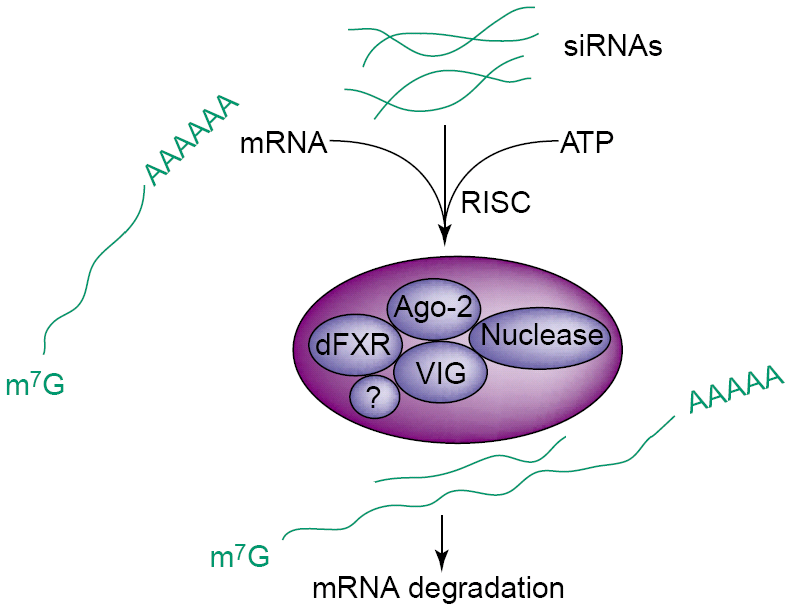

Figure 5. The various protein subunits comprising the RISC complex, which is the effector guided by the siRNA joining the complex. The subunit labelled 'nuclease' has since been identified as tudor staphylococcal nuclease (TSN-1). Taken from Denli *et al*[32].

However, there are many questions which remain to be answered. How does RISC combine with the siRNA or microRNA, and how is it used as a guide to cleave the mRNA? There are indications that the PAZ domain within Ago-2 may bind the 3' end of the siRNA. Is the siRNA unwound to allow the RISC complex to match it to a complementary mRNA sequence, and how is cleavage effected? There are many opportunities for future research in this area.

Properties of the short interferring RNA (siRNA)

Short interfering RNAs produced by DICER are double stranded and have a 2 nucleotide overhang at the 3' end, a phosphorylated 5' end and a terminal hydroxyl group attached at the 3' end.

Zamore *et al*[33] in 2000 were the first to show that cleavage of the dsRNA by DICER occurs at 21 to 23 nucleotide intervals, and is ATP dependent. The siRNAs which were generated by DICER did not require the target mRNA to be present for cleavage to occur, thus proving the decoupled nature of generation of siRNAs and the subsequent association with RISC. RNA-directed RNA polymerases (RdRP) are thought to assist in amplifying the number of siRNAs after the initial introduction of dsRNA.

The high resolution gel in Figure 6 shows the products cleaved after incubation for 0, 20 and 60 minutes of the Rr-luciferase mRNA with each of three dsRNAs, A, B and C. Curiously, one of the short RNAs appears to be only 9 nucleotides long; it is thought that this has occurred because the run of 7 uracil residues 'resets' the ruler which is used for cleavage.

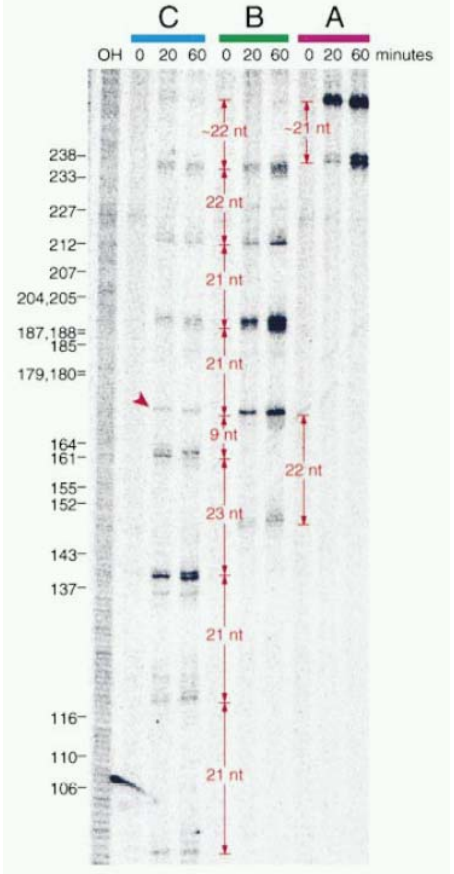

Figure 6. High resolution gel showing the size of the siRNA as being 21-22nt long. Taken from Zamore *et al*[33].

Elbashir *et al*[34] found that the 3' 2nt overhangs of two uracil residues are more efficient for RNAi than siRNAs that have 3' overhangs of AA, CC or GG. Cleavage of the target mRNA occurs at the point defined by the 5' end of the siRNA, rather than at the 3' end.

Whilst it was initially observed that a mRNA was the target for degradation by the RISC complex containing the siRNA, research by Liang, Liu and Michaeli[35] found that small nucleolar RNAs (snoRNAs) can also be degraded by RNA interference, indicating the variety of pathways in which RNA interference is involved. The snoRNAs that they observed are involved in the synthesis of rRNA in the nucleolus of trypanosomes, and the expression of siRNAs complementary to the snoRNAs caused their degradation. The snoRNAs differ from mRNAs in that they have a different 5' end and poly(A)s at the 3' end, so the degradation does not appear to be affected by these features of the RNA.

Intracellular control of RNA interference

Whilst the initial research on RNA interference was done using *C. elegans*, it was suspected that since RNAi was a naturally occuring pathway, it may be present and functional in other organisms. Many organisms have since been studied to determine whether RNAi occurs equally well in them, with differing degrees of success. RNAi has been shown to work well in *C. elegans, Arabidopsis* and other plants*, Drosophila, Planaria* and *Trypanosomes*. However, the studies are taking longer in higher organisms due to their complexity.

In particular, inducing RNA interference in mammalian cells is more difficult, since when dsRNAs longer than 30nt are introduced into a cell, they activate a defence mechanism which produces an interferon cytokine. This causes non-specific RNA degradation and a general shutdown of cell protein synthesis[36], which is mediated through a dsRNA-activated protein kinase (PKR). PKR phosphorylates EIF-2α in reponse to dsRNA, and terminating translation non-specifically. This PKR pathway can also cause apoptosis[37].

The PKR defence means that long (>30nt) dsRNAs cannot be introduced into mammalian cells; however, they can be expressed intracellularly and then cleaved by DICER to produce siRNAs. A number of strategies for producing dsRNA within cells have been investigated by using short hairpins or dual promoters. For example, Wang *et al*[38] have used RNA interference to inhibit gene expression in *T. brucei* using opposing T7 promoters to produce the ssRNA intracellularly, which hybridize to produce the dsRNA. These methods are shown in figure 7.

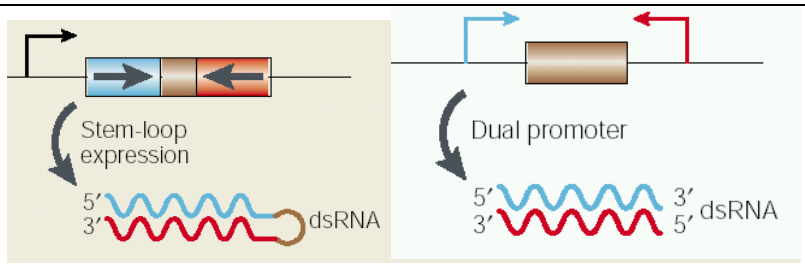

Figure 7. Stem-loop expression and dual promotors are a couple of methods of expressing dsRNA intracellularly. Transcription occurs in the directions indicated by the arrows. Taken from Hammond, Caudy and Hannon[44].

DICER, as well as cleaving long dsRNA into siRNAs, is also responsible for the maturation and cleavage of short hairpin RNAs (shRNAs) which are endogenously encoded in the genome. *let-7* is known as a small, highly conserved RNA in *C. elegans*, that is encoded in the genome and transcribed as a ~70 nucleotide RNA and processed into a ~21nt RNA[39]. Paddison *et al*[15] aimed to "*retarget these small, endogenously encoded hairpin RNAs to regulate genes of choice*" and performed extensive experiments using expression plasmids to produce custom shRNAs suitable for silencing specific targets. They note that "*the ultimate utility of encoded short hairpins will be in the creation of stable mutants that permit the study of the resulting phenotype*".

A more sophisticated and controllable approach has been developed by Gupta *et al*[40]. By placing the expression of short-hairpin RNAs under the control of a U6 promotor for an RNA polymerase III, this allows the control of expression of shRNAs and therefore of silencing in mammalian cells using RNA interference. The inducible system used a ecdysone-responsive transcriptional element, which is a common system used for mammalian cells, and this was delivered using a retrovirus. The group targetted the p53 tumor suppressor gene for silencing, as shown in figure 8.

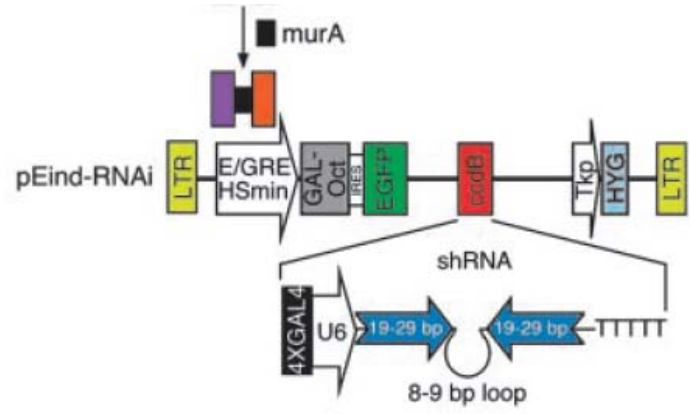

Figure 8. Expression of short hairpin RNAs (shRNAs) under control of a U6 promotor. pEind-RNAi is a self-inactivating retroviral ecdysone-inducible vector. From Gupta *et al*[40].

p53 was chosen as a target for silencing since antibodies are available to monitor levels of the protein, and an effective shRNA had already been calculated for this gene. It was found that p53 was suppressed by RNA interference and could be controlled in a dose-dependent way by addition of the inducer ecdysone. Once the induction was stopped, the silencing of p53 halted and levels of the protein were restored to normal.

These results the variety of methods available to initiate silencing, and illustrate that it is possible to maintain an increasingly fine-grained control over the silencing of specific genes using RNA interference.

The efficacy of siRNAs; no side effects?

For RNA interference to be a useful tool, it is important that siRNAs do not produce cause any effects other than stopping the expression of the target gene. This could occur through cross-hybridization of the antisense strand to the mRNA of non-target genes. In addition to the problems already discussed regarding introducing dsRNA into mammalian cells, research is ongoing to ensure that RNAi causes no other side-effects to a phenotype.

Semizarov *et al*[41] have written extensively about this, and have noted that "*if an siRNA produces a phenotype such as apoptosis or cell cycle arrest because of cross-hybridization, sequence-specific protein binding, or a general dsRNA response, then the target gene may be erroneously associated with that phenotype*". In other words, scientists using RNAi as a technique need to be careful that the correct gene is being silenced through other controls.

Semizarov *et al* used DNA microarrays to analyse the "global view" of the gene expression occurring, and to notice any non-specific changes in gene expression, which was not observed. They found that siRNAs at concentrations of approximately 100nM can induce the unwanted expression of genes which are involved in apoptosis and stress response. Reduction in the concentration of siRNA to 20nM prevented this response.

In similar research, Jackson *et al*[42] used a microarray to profile genome-wide changes in expression, when using siRNAs to silence two genes involved in signal transduction, insulin-like growth factor receptor (IGF1R) and mitogen-activated protein kinase 1 (MAPK14). They found that the mRNA for non-targeted genes (i.e. other than IGF1R or MAPK14) could be affected by siRNA, where the off-target genes that were silenced had only 15 contiguous nucleotides that were identical to the siRNA. They also found that an siRNA duplex, designed to silence two off-target transcripts, KPNB3 and FLJ20291, also silenced MAPK14 as well as the indented targets. These off-target transcripts shared only 14 contiguous nucleotides with MAPK14, and 15 nt in total (see figure 9).

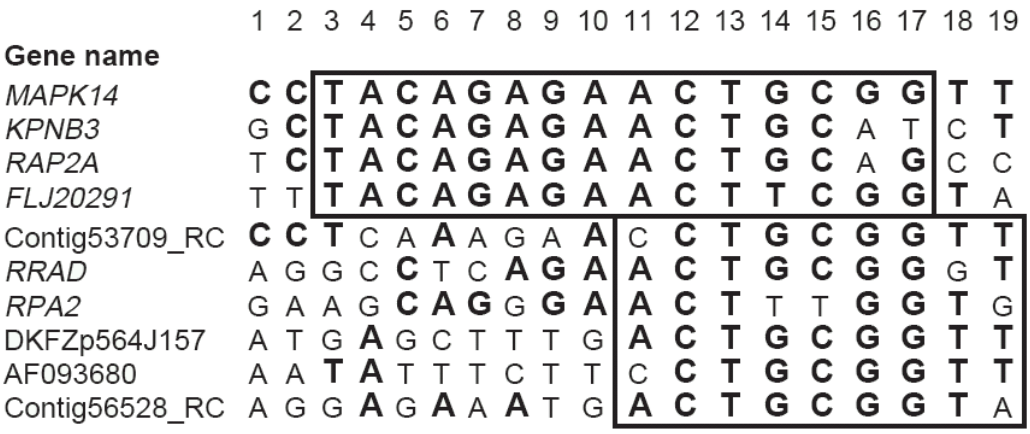

Figure 9. Sequence alignment of genes regulated with similar kinetics to MAPK14; contiguous nucleotides with perfect identity to MAPK14 are marked in bold. Taken from Jackson *et al*[42].

In 2004, Persengiev, Zhu and Green[43] have shown that siRNAs in mammalian cells can non-specifically affect the regulation of more than 1000 other genes, either non-specifically stimulating or repressing these genes, depending on siRNA concentration. This can be explained since dsRNAs can influence multiple transcription and signalling pathways in addition to the dsRNA protein kinase response (PKR).

It is still unclear how the number and location of mismatches between the siRNA and the target mRNA affect the specificity of the RNAi response. These papers indicate there are more complex subtleties in designing siRNAs, and the original hope of being able to "*knock out your favourite gene with only a day's work ... in any organism*"[44] may have been overly optimistic. However, once these factors affecting the specificity of the siRNA as a guide are quantified, the design of siRNAs may be able to take into account off-target regulation and produce siRNAs that are known to only silence the indended gene, to avoid non-specific silencing.

Proteins involved in RNAi discovered from mutant phenotypes

Nearly a dozen genes have so far been identified that affect the RNA interference process in some way: nucleases (mut-7 in *C. elegans*), helicases (qde-3 in *Neurospora crassa*, mut-6 in *C. elegans*), RNA-dependent RNA polymerases (qde-1, ego-1, SDE1, SGS2,) and members of the Argonaute family (rde-1 in *C. elegans*, qde-2, AGO1 in *Arabidopsis thaliana*). Most of these have been determined from mutant phenotypes which were missing one of these genes.

Some of these proteins have already been mentioned in passing in this paper. Figure 10 shows a more complete table of proteins discovered so far which are involved in the RNAi pathway, the domains contained within and the possible function of the domain.

Protein(s) or protein family	**Contain domains**	**Domain function**
Dicer family	RNA helicase	RNA unwinding
	PAZ	bind ssRNA
	RNase III	Ribonuclease
	dsRNA binding	dsRNA binding
Argonaute family	PAZ	bind ssRNA
	PIWI	Unknown
RNA-dependent RNA polymerases	RdRP	RNA-dependent RNA polymerisation
RNA helicases	Putative RNA helicase	RNA unwinding
QDE-3	DNA helicase	DNA unwinding
RDE-4	dsRNA binding	dsRNA binding
MUT-7	RNase D	RNA degradation
Fragile X related protein (dFXR)	KH	Putative RNA binding
	RGG	Putative RNA binding
Vasa intronic gene (VIG)	RGG	Putative RNA binding

Figure 10. Proteins and domains involved in RNA interference and related phenomena. Taken from Denli *et al*[32].

It can be seen that there are a large number of proteins involved in the process of RNAi, and many are still not understood or characterised.

RNA interference, microRNAs and the genome

As previously mentioned, parts of the RNA interference pathway are also involved with other small types of RNA. microRNAs are a large family of small RNAs which are encoded within the genome, and are known to be able to regulate the expression of genes and development[45].

Lee and Ambros[46] discovered a class of genes that encoded RNAs which are essential for proper development in *C. elegans*. These are the lin-4 and let-7 microRNAs, which are single stranded and ~22nt long, and were identified by their mutant phenotypes. miRNAs are located within intergenic regions (IGRs) of the genome, with some miRNAs being highly conserved, across species and phyla boundaries. These miRNAs may have previously been unidentified because they do not contain an open reading frame[47]. In April 2004, there were 714 known miRNAs in the Sanger miRNA registry[48].

miRNAs are expressed as ~70nt pre-cursors hairpin RNAs (pre-miRNAs) that snap-back to anneal to themselves, since half their sequence is complementary to itself in reverse, following Crick-Watson base pairing rules. These pre-cursors are processed by Drosha and DICER enzymes to yield single stranded mature microRNAs[49].

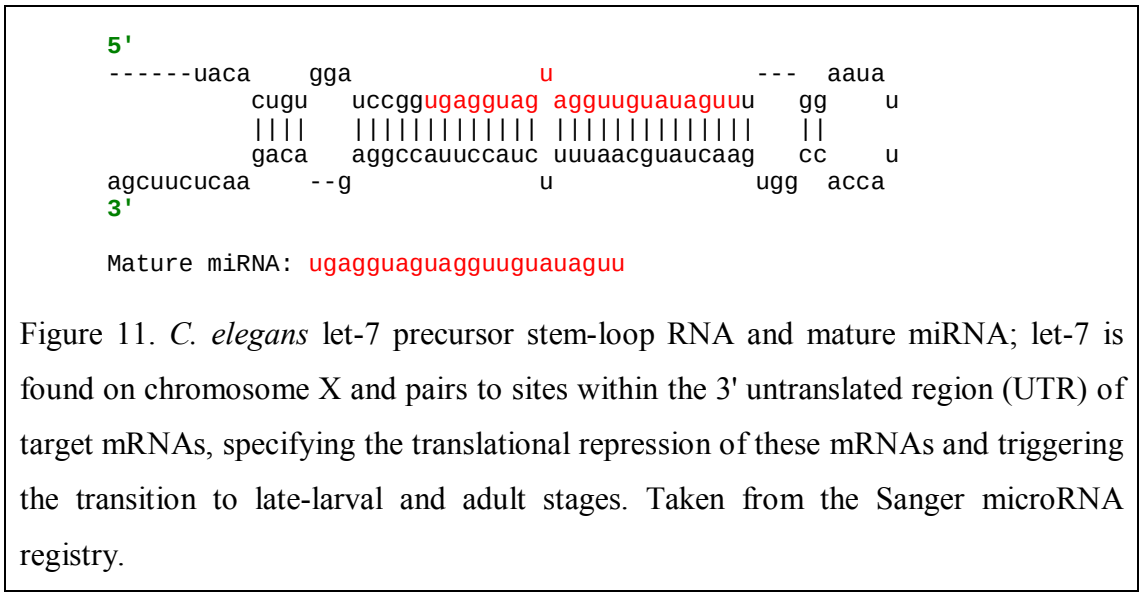

Figure 11. *C. elegans* let-7 precursor stem-loop RNA and mature miRNA; let-7 is found on chromosome X and pairs to sites within the 3' untranslated region (UTR) of target mRNAs, specifying the translational repression of these mRNAs and triggering the transition to late-larval and adult stages. Taken from the Sanger microRNA registry.

The miRNA pre-cursors do not have to be perfectly complementary in sequence to themselves; indeed, it is quite normal for one or more bulges to occur in the stem-loop structure, and typically just one half of the stem is preserved in the mature miRNA. The let-7 mature miRNA and pre-cursor is shown in Figure 11, and can be seen to have characteristic bulges in the stem-loop where nucleotides are mismatched.

Caudy *et al*[31] have noted that either siRNAs or miRNAs can bind to the RISC complex. In constrast to the action of siRNA, miRNAs cause translation to be repressed, rather than the mRNA to be cleaved and degraded. Voinnet[50] postulates that imperfectly matched miRNAs could affect other cellular processes, such as mRNA splicing, localisation or stability, and notes that many miRNAs found in *Drosophila* have been found to complement motifs which can alter both translational efficiency and the stability of transcripts.

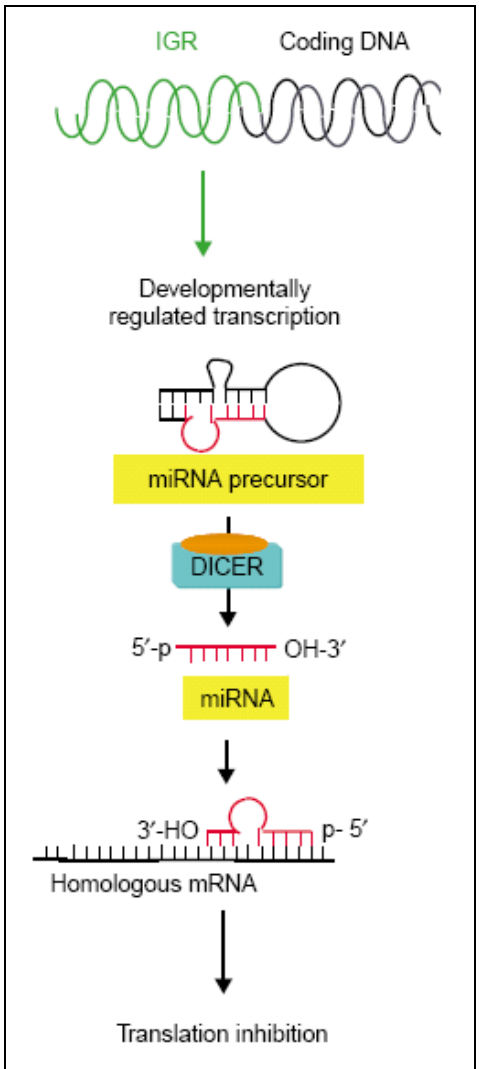

Figure 12. Processing of miRNA pre-cursors from stem-loop pre-cursors by the DICER enzyme. Taken from Voinnet, O.

Doench, Petersen and Sharp[51] found that a siRNA could also function as a miRNA, repressing expession of a target mRNA. By including a "bulge" in match of the siRNA to the mRNA, they found that this precluded mRNA cleavage by RISC, but translation into protein was repressed.

Recent papers suggest that RNA interference may protect the genome from the effects of mobile transposons and other repetitive sequences. These include defending cells against RNA viral infection (exogenous threats), suppressing mobilization of transposons (endogenous transposon threats), and regulating expression of endogenous genes in development.

Protecting against viral infection is appears to be a particularly important application of post-transcriptional gene silencing (PTGS), in plants[52] (which is very similar to RNAi) since they do not possess an immune system that uses antibodies to defend against threats. Plant viruses have ssRNA genomes in more than 90% of cases, and these are replicated by an RNA-dependent RNA polymerase (RdRP). In plants that exhibit co-suppression, virus-induced gene silencing (VIGS) and virus resistance, but not in control plants, ~25nt sense and antisense RNAs with homology to the gene being silenced have been found. This indicates strongly that PTGS is involved in protecting the plant from virus threats.

Sijen and Plasterk[53] have observed transposon silencing in *C. elegans* by RNAi. siRNAs appear to be involved in a pathway that methylates specific genes, and this may related to the transposon silencing. Ketting *et al*[54] found evidence that RNAi defends against transposons. In *C. elegans*, out of 30 mutants which allow transposons to become active, 22 of these mutants also cause defects in the RNAi process.

To further complicate the picture, Hamilton *et al*[55] have found two classes of siRNA that differ slightly in size (short siRNAs being 21-22nt and long siRNAs being 24-26nt), and it appears that only long siRNAs may be involved in methylating retrotransposons, preventing them from being expressed and jumping between locations on the chromosome.

Conclusion

What started out as a simple observation in petunias, of altered pigmentation due to the degradation of mRNA transcripts, has been expanded into the discovery of a number of families of small, temporal RNAs (siRNA, miRNAs and stRNAs) and the sophisticated enzyme machinery to of DICER and RISC produce and process these RNAs, which in turn regulate and repress mRNAs of both endogenous and exogenous genes.

The RNAi pathway is required for the correct development of *Arabidopsis thaliana* and *C. elegans*, and allows plants and organisms to guard against viral threats and transposable elements in the chromosome. Harnessing the power of RNAi to knock down one or more desired genes in any organism, whilst theoretically simple, has proved more challenging in mammalian cells, and even in *C. elegans*, where the nematode can easily ingest dsRNA, subtleties in the design of siRNAs can affect the efficacy of the RNAi process, and potentially affect gene expression non-specifically at higher concentrations.

Only when these factors and the mechanisms underlying RNAi are more fully understood, through a combination of genetic and biochemical experimental approaches, can RNAi become a panacea for gene knockdown. However, in the meantime, it continues to hold great potential and remains a useful tool to create and study mutant phenotypes, although it may take many years to elucidate the function and structure of the proteins involved in the DICER, RISC and other complexes that are involved in RNAi.

Table of Figures

References

[1] Fire, A. *et al*. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis Elegans*. Nature **391**, 806-811.

[2] Zamecnik P. and Stephenson M. (1978) Inhibition of rous sarcoma replication and cell transformation by a specific oligodeoxynucleotide. Proc. Natl Acad. Sci. USA **75**, 280-284

[3] Jorgensen, R. Altered gene expression in plants due to *trans* interactions between homologous genes. Trends Biotechnol. **8**:340-344 (1990)

[4] Cogoni, C. and Macino, G. (1997). Isolation of quelling-defective (qde) mutants impaired in post-transcriptional transgene-induced gene silencing in *Neurospora crassa*. Proc. Natl. Acad. Sci. USA **94:**10223-10238.

[5] Misquitta, L. and Paterson, B.M. (1999). Targetted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. Proc. Natl. Acad. Sci. USA **96**, 1451-1456.

[6] Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M. and Ahringer, J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature*. **408**, 325-330.

[7] Ngo, H. Tschudi, C., Gull, K. and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA **95**, 14687-14692.

[8] Lohmann, J.U., Endl. I. and Bosch, T.C. (1999). Silencing of Developmental Genes in *Hydra*. Dev. Biol. **214**, 211-214.

[9] Sanchez-Alvarado, A. and Newmark, P.A. (1999). Double-strnaded RNA specifically disrupts gene expression during planarian regeneration. Proc. Natl. Acad. Sci. USA **96**, 5049-5054.

[10] Li, Y.X., Farrell, M.J., Liu, R., Mohanty, N. and Kirby, M.L. (2000). Double-stranded **RNA** injection
produces null phenotypes in zebrafish. *Developmental Biology*. **217**, 394-405

[11] 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.

[12] Jacque, J.M, Triques, K. and Stevenson, M. (2002). Modulation of HIV-1 replication by RNA inteference. Nature **418**:435-438.

[13] Hannon GJ. (2002). RNA interference. Nature **418**:244-51

[14] Carmichael, G.G. (2002), Silencing viruses with RNA. Nature **418**:379-380.

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

[16] Tabara. H. *et al.* (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. Cell **99**, 123-132.

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

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

[19] 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.

[20] Provost, P. *et al* (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J. **21** 5864-5874

[21] Tang, G. et al (2003). A biochemical framework for RNA silencing in plants. Genes and Development **17**:49-63

[22] Ketting, R.F. *et al* (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes and Development **15**:2654-2659.

[23] Grishok, A. *et al* (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

[24] 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.

[25] Blaszczyk, J., *et al* (2001). Crystallographic and Modeling Studies of RNase III Suggest a Mechanism for Double-Stranded RNA Cleavage. Structure, Vol **9**, 1225-1236.

[26] Song *et al*. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nature Structural Biology Vol. **10** No. 12:1026-1032.

[27] Yan, K.S. et al (2003). Structure and conserved RNA binding of the PAZ domain. Nature **426**:469-474.

[28] Hammond S.M., Boettcher S., Caudy A.A., Kobayashi R., Hannon G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science. **293**:1146-1150.

[29] 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 and Development **16**:2733-2742.

[30] Caudy A.A., Myers M., Hannon G.J., Hammond S.M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev. **16**:2491-2496.

[31] Caudy, A.A. *et al*., (2003). A micrococcal nuclease homologue in RNAi effector complexes, Nature, **425**:411-415

[32] Denli, A.M. and Hannon, G.J. (2003). RNAi: an ever-growing puzzle. Trends in Biochem. Sci. **28**:196-200

[33] 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.

[34] Elbashir, S.M. *et al*. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J. **20**:6877-6888.

[35] Liang, X.-H., Liu, Q. and Michaeli, S. (2003). Small nucleolar RNA interference induced by antisense or double-stranded RNA in trypanosomatids. Proc. Natl. Acad. Sci. USA **100**, 7521-7526.

[36] Baglioni, C. and Niulsen, T.W. (1983) Mechanisms of antiviral action of interferon. Interferon **5**, 23-42.

[37] Gil, J. and Esteban, M. (2000). Induction of apoptosis by the dsRNA-dependent protein kinase (PKR): mechanism of action. Apoptosis **5**, 107-114.

[38] Wang, Z., Morris, J.C., Drew, M.E. and Englund, P.T. (2000). Interference of Trypanosoma brucei gene expression by RNA interference using an integratable vector with opposing T7 promotors. J. Biol. Chem. **275**, 40174-40179.

[39] Reinhart, B.J. *et al* (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *C. elegans*. Nature **403**:901-906.

[40] Gupta, S. *et al* (2004). Inducible, reversible, and stable RNA interference in mammalian cells. Proc. Natl. Acad. Sci. USA **101**:1927-1932.

[41] Semizarov, D *et al*. (2003). Specificity of short interfering RNA determined through gene expression signatures. Proc. Nat. Acad. Sci. USA vol. **100** no. 11, 6347-6352

[42] Jackson, A.L. *et al*. (2003) Expression profiling reveals off-target gene regulation by RNAi. Nature Biotechnology, Vol **21** No. 6, 636-637.

[43] Persengiev SP, Zhu X, and Green MR. (2004). "Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs)". RNA, vol. **10**, no. 1, 12-18.

[44] Hammond, S.M., Caudy, A.A, Hannon, G.J. (2001). Post-transcriptional gene silencing by double-stranded RNA. Nature Rev. Genetics **2**:110-119.

[45] Pasquellini, A.E. (2002). MicroRNAs: deviants no longer. Trends in Genetics **18**:171-173.

[46] Lee, R.C. and Ambros, V. (2001). An Extensive Class of Small RNAs in Caenorhabditis elegans. Science **294**:862-864

[47] Pederson, T. (2004). RNA Interference and mRNA silencing, 2004: How Far will they reach? Mol. Biol. Cell. 15, 407-410.

[48] Griffiths-Jones S. (2004). The microRNA Registry. Nucleic Acids Research, **32**, Database Issue, D109-D111. http://www.sanger.ac.uk/Software/Rfam/mirna/

[49] Lee, Y. *et al*. (2003). The nuclear RNase III Drosha initiates miRNA processing. Nature **425**:415-8.

[50] Voinnet O. (2002). RNA silencing: small RNAs as ubiquitous regulators of gene expression. Curr. Opin. in Plant Biology, **5**:444-451

[51] Doench, J.G, Petersen, C.P. and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes and Development, **17**:438-442.

[52] Waterhouse, P.M., Wang, M. and Lough, T. (2001) Gene silencing as an adaptive defence against viruses. Nature **411**, 834-842

[53] Sijen, T. and Plasterk, R.H.A (2003). Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature **426**:310-314.

[54] Ketting, R.F., Haverkamp, T.H.A., van Luenen, H.G.A. and Plasterk, R.H. (1999). mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Warner Syndrome helicase and RNase D. Cell **99**:133-141.

[55] Hamilton A. *et al* (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. **21** No 17: 4671-4679.