Structural and functional characterization of proteins involved in RNA-mediated gene silencing

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of Doctor of Sciences

presented by

ANDREAS LINGEL
Dipl. Natw. ETH

born 08.02.1976
citizen of
Germany

accepted on the recommendation of

Prof. Dr. Ulrike Kutay
Prof. Dr. Frédéric Allain
Dr. Elisa Izaurralde

2006
In Greek mythology, the Argonauts are a group of sailors, named after their ship, the Argo. Under the leadership of Jason, they start an expedition to find and bring back the Golden Fleece, a treasury guarded by a dragon and located in Colchis on the Black Sea. With the help of the gods, they are able to finish their mission successfully and return to Greece.

The illustration shows a detail of an Athenian red-figure clay vase, dated about 425-375 BC. It shows the Argonauts on board their ship Argo, with the NMR ensemble of the Drosophila Argonaute2 PAZ domain in the foreground. The vase is located in the Museo Jatta in Ruvo, Italy.
This thesis describes work carried out in the laboratory of Dr. Elisa Izaurralde at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany under the supervision of Prof. Dr. Ulrike Kutay (Institute of Biochemistry at the Swiss Federal Institute of Technology Zurich). The work was completed between May 2002 and December 2005 and was supported by an EMBL predoctoral fellowship.

The following publications are presented in this thesis:


In addition, a perspective and a review are presented:


Acknowledgements

First of all, I want to thank my supervisor Dr. Elisa Izaurralde for giving me the opportunity to work in her lab. I would like to thank her for excellent scientific guidance and all her enthusiasm, for the great help and for the constant support throughout the last years. This was all invaluable for the success of this Ph.D. thesis. Furthermore, she enabled my long-term collaboration with the laboratory of biomolecular NMR at EMBL.

Especially, I want to thank Dr. Michael Sattler for the possibility to do this collaborative work with his lab. He and Dr. Bernd Simon had essential contributions to my projects and taught me NMR spectroscopy. Without them, this work would not have been possible.

I am very grateful to Prof. Dr. Ulrike Kutay for the supervision of this thesis at ETH Zurich. Furthermore, I thank Prof. Dr. Frédéric Allain for being my ‘co-referent’ and Prof. Dr. Nikolaus Amrhein for his help and for chairing the examination committee at ETH Zurich.

I would like to thank the members of my thesis advisory committee at EMBL for their interest in my work, for fruitful discussions and for their helpful advice. In addition to Dr. Elisa Izaurralde and Dr. Michael Sattler, these were Dr. Andreas Ladurner, and Prof. Dr. Ulrike Kutay.

A special “thank you” goes to Michaela Rode and Gunter Stier. Both of them provided excellent technical support and very useful advice throughout the time of my thesis.

Next, I want to thank the past and present members of the lab for being nice colleagues and friends, and for enjoyable times inside and outside of the lab. These are: Andrea Herold, David Gatfield, Leonie Unterholzner, Daniel Forler, Kerstin Gari, Jan Rehwinkel, Maria Polycarpou-Schwarz, Tamás Orban, Pavel Natalin, Isabelle Behm-Ansmant, Valerie Hilgers, Schu-Fee Yang, Ana Sofia Bregieiro Eulalio, Isabelle Behm-Ansmant, Valerie Hilgers, Schu-Fee Yang, Ana Sofia Bregieiro Eulalio, Suheda Erener, and Foteini Christodoulou. In addition, I would like to thank the members of the NMR lab: Fabian Filipp, Christian Edlich, Cameron Mackereth, Anna Oddone, Lorenzo Corsini, Ana Messias, Malgorzata Duszczzyk, Frank Gabel, and Alexander Gasch.

I also want to thank other people at EMBL who contributed to my work in one way or another. These are Elena Conti, Sébastien Fribourg, Atlanta Cook, Jérome Basquin, all members of the photolab and of other service units.

I am very thankful to Elisa, Sufe, Jan, Pavel, Michael and Bernd for reading this thesis and for helpful comments.

# Table of Contents

## Zusammenfassung

Summary ................................................................................................................................... 5

## 1. Introduction .................................................................................................................. 7

1.1 Regulation of gene expression by RNA silencing ..................................................... 7
  1.1.1 A brief history of the discovery of RNA silencing ............................................ 7
  1.1.2 RNA interference ............................................................................................... 9
  1.1.3 miRNA pathway ............................................................................................... 13
  1.1.4 RNA-mediated transcriptional gene silencing ................................................. 16

1.2 Argonaute proteins ................................................................................................... 18
  1.2.1 Argonaute proteins: a conserved protein family with diverse functions.......... 18
  1.2.2 The Drosophila Ago1 and Ago2 proteins ........................................................ 21
  1.2.3 Drosophila and human Argonaute2 act as Slicer ............................................. 22
  1.2.4 Structures of archaeal Argonaute and Piwi proteins ........................................ 23

1.3 Viral suppression of RNA silencing ......................................................................... 30

1.4 NMR spectroscopy: a tool to understand the function of biomolecules ................. 36
  1.4.1 The basic principle of nuclear magnetic resonance ......................................... 36
  1.4.2 The chemical shift ............................................................................................ 39
  1.4.3 The basic NMR experiment ............................................................................. 40
  1.4.4 The interaction of spins through chemical bonds: scalar coupling .................. 41
  1.4.5 Resonance assignment in proteins .................................................................... 43
  1.4.6 The interaction of spins through space: dipolar coupling and residual dipolar coupling ............................................................................................................ 45
  1.4.7 The nuclear Overhauser effect ......................................................................... 47
  1.4.8 Relaxation and dynamics ................................................................................. 48
  1.4.9 Structure determination with NMR derived constraints .................................. 52
  1.4.10 Interaction studies by NMR spectroscopy ....................................................... 54

## 2. Results .......................................................................................................................... 57

2.1 Structure and nucleic acid binding of the Drosophila Argonaute2 PAZ domain ... 58

2.2 Nucleic acid 3’-end recognition by the Argonaute2 PAZ domain .......................... 71

2.3 NMR assignment of the Drosophila Argonaute2 PAZ domain .............................. 88

2.4 The structure of the FHV B2 protein, a viral suppressor of RNAi, reveals a novel mode of dsRNA recognition ................................................................. 91
3. Discussion .............................................................................................................................. 103
   3.1 The PAZ domain represents a novel nucleic acid binding fold ................................. 103
   3.2 The PAZ domain as a specificity providing module in the RNAi pathway .......... 110
   3.3 The FHV B2 protein binds dsRNA in a novel mode ............................................... 113
   3.4 Viral suppressors of RNAi act at different levels in the RNAi pathway .............. 117

4. Additional Publications ...................................................................................................... 119
   4.1 RNAi: Finding the elusive endonuclease ............................................................... 120
   4.2 Novel modes of protein-RNA recognition in the RNAi pathway .................... 126

5. Abbreviations .................................................................................................................... 136

6. References ....................................................................................................................... 138

7. Curriculum vitae .............................................................................................................. 148
Zusammenfassung


Zu Beginn dieser Arbeit war bereits bekannt, dass die PAZ-Domäne nur in Dicer- und Argonaute-Proteinen vorkommt. Darüber hinaus konnte die Interaktion zwischen Mitgliedern beider Proteinfamilien gezeigt werden und deshalb wurde vorgeschlagen, dass die PAZ-Domäne als Protein-Protein-Bindungsdomäne den direkten Kontakt zwischen den beiden Proteinen vermittelt.

Nukleinsäuren binden können, was ich tatsächlich für mehrere PAZ-Domänen von Argonaute-Proteinen von *Drosophila* und des Menschen zeigen konnte.

Um die beobachtete Bindungsspezifität der PAZ-Domäne besser verstehen zu können, habe ich im weiteren Verlauf der Untersuchungen die Strukturen der PAZ-Domäne des Argonaute2-Proteins im Komplex mit DNA- und RNA-Oligonukleotiden gelöst (Kapitel 2.2). Dadurch wurde deutlich, dass nur die zwei letzten Nukleotide am 3'-Ende gebunden werden und dass die 3’-OH Gruppe des letzten Nukleotids tief innerhalb der Bindungstasche sitzt. Das gleiche Resultat ergab die Charakterisierung der Bindung eines weiteren RNA-Liganden, der wie siRNAs einen doppelsträngigen Teil und einen Überhang am 3'-Ende besaß. Dieses Ergebnis ermöglichte eine Erklärung der beobachteten Bindungsspezifität der PAZ-Domäne und die Definition der molekulare Funktion als die der 3’-Erkennung. Somit kann man schlussfolgern, dass die PAZ-Domäne zur spezifischen Erkennung von siRNAs, die von Dicer erzeugt wurden, beitragen kann und deren Einbau in den RNAi-Effektorkomplex unterstützt.

RNA interference (RNAi) is an evolutionary conserved mechanism that regulates gene expression in response to the presence of double-stranded RNA (dsRNA) in the cell. The dsRNA is first cleaved by the ribonuclease (RNase) III-type enzyme Dicer into 21-23 nucleotide (nt) small interfering RNA duplexes (siRNAs) with 2 nt 3’-overhangs. In a subsequent step, the siRNAs are incorporated into a multimeric RNA-induced silencing complex (RISC), where they guide the selection of a complementary mRNA. Argonaute protein family members are core components of RISCs and characterized by two conserved domains, namely the central PAZ domain and the C-terminal Piwi domain. The Piwi domain has endonucleolytic activity and mediates cleavage of the target mRNA in a region with sequence complementarity to the siRNA guide. The expression of the targeted gene is thereby strongly down-regulated.

At the time when this study was initiated, the PAZ domain had been described to be present exclusively in two protein families, namely the Dicer and Argonaute protein family. As members of these families had been shown to interact with each other, it was proposed that the PAZ domain might serve as a protein-protein interaction domain, mediating physical contact between Dicer and Argonaute proteins.

The main focus of the work described in this thesis was to reveal the function of the PAZ domain on a molecular level. To achieve this, I first solved the three dimensional nuclear magnetic resonance (NMR) structure of the free *Drosophila melanogaster* Ago2 PAZ domain (chapter 2.1). This showed that the PAZ domain adopts a novel fold composed of a central five-stranded $\beta$-barrel flanked by two $\alpha$-helices and a conserved $\beta$-hairpin/$\alpha$-helix insertion. Based on the structure, I performed further experiments that showed that the PAZ domain binds nucleic acids *in vitro*, which was an unexpected result. I could demonstrate that the PAZ domain binds single-stranded RNAs and double-stranded RNAs with 2 nt single-stranded 3’-overhangs, but has reduced affinity for blunt-ended double-stranded RNA molecules. This pointed towards a specificity of the PAZ domain for single-stranded nucleic acids. The conservation of residues located in the binding site suggested a similar function for all PAZ domains. Indeed, I could show nucleic acid binding for multiple PAZ domains of *Drosophila* and human Argonaute proteins.
Summary

Having established the function of the PAZ domain in nucleic acid binding, the next goal was to understand the apparent specificity for single-stranded nucleic acids. For this, I defined short oligonucleotides as effective ligands of the PAZ domain and determined two complex structures of the Drosophila Ago2 PAZ domain bound to RNA and DNA oligomers, respectively (chapter 2.2). This revealed that only the two 3’-terminal nucleotides are bound by the PAZ domain, whereas the other nucleotides showed no stable interaction. A very similar binding mode was observed when I studied the binding of an siRNA mimic, which had a double-stranded stem and a 2 nt 3’-overhang. These observations provided an explanation for the characterized specificity of the PAZ domain and demonstrated that the PAZ domain is a 3’-end recognition domain. Together, my results indicate that the PAZ domain contributes to the recognition and incorporation of Dicer-processed siRNAs into RISC and thus serves as a specificity providing module in the RNAi pathway.

In the second part of my thesis, I studied the structure and function of a viral suppressor of RNA interference. RNAi is thought to originate from an ancient endogenous defense mechanism against viral and other heterologous dsRNAs. It had been shown before that many viruses have evolved proteins that suppress RNA silencing and thereby counteract the cellular response to the presence of double-stranded viral RNA. Although a large number of viral suppressors had been identified, little is known about their molecular mechanisms. One of these previously described suppressors of RNAi is the Flock House virus (FHV) B2 protein, which was shown to inhibit an anti-viral response in insect cells, plants and Caenorhabditis elegans. To understand the molecular basis of its suppression activity, I determined the solution structure of the FHV B2 protein (chapter 2.4). This study demonstrated that B2 is an elongated dimer in solution, with each monomer composed of three α-helices. NMR experiments demonstrated binding of the B2 dimer to a double-stranded siRNA with high affinity. The binding site was mapped to a surface which is comprised of one complete side of the dimer and composed of two anti-parallel helices (one of each monomer). The distribution of positively charged side chains located in these helices suggested that the dsRNA is contacted mainly via the sugar phosphate backbone, which explains the lack of sequence specificity. This binding mode represents a novel way of dsRNA recognition. As both siRNAs and long dsRNAs are bound, a dual mode of suppression of RNAi is suggested: the B2 dimer could prevent the incorporation of viral RNA into RISC by inhibition of Dicer-mediated cleavage of viral dsRNA and by binding directly to siRNAs.
1. Introduction

1.1 Regulation of gene expression by RNA silencing

The basic flow of information in gene expression is from DNA to RNA to protein. This is also known as the ‘central dogma of molecular biology’ that was established by Francis Crick in the year 1958 (Crick, 1958). In order to ensure the right correlation between genes and the amount of protein being produced, the expression of genes has to be tightly controlled by complex regulatory mechanisms in both temporal and spatial manner. Recently, several post-transcriptional mechanisms that regulate gene expression on the level of the messenger RNA (mRNA) were discovered to be mediated by RNA. All these pathways, now commonly referred to as RNA silencing pathways, change mRNA levels and/or influence the level of the corresponding protein in response to the presence of dsRNA.

1.1.1 A brief history of the discovery of RNA silencing

Before RNA silencing was discovered in animals, two phenomena of double-stranded RNA-induced gene silencing had been already known for some years in plant systems, namely virus-induced gene silencing (VIGS) and co-suppression. In the latter case it was observed that after introducing exogenous transgenes into petunias with the goal of altering pigmentation, the flowers did not deepen flower colour as expected. Instead, the flowers showed variegated pigmentation, with some lacking pigment completely (Jorgensen, 1990). This phenomenon, then called co-suppression, indicated that not only the transgenes themselves were inactive, but also that the added DNA sequences somehow affected the expression of the endogenous loci. Also, several laboratories found that plants responded to RNA viruses by targeting viral RNAs for destruction (Ruiz et al., 1998 and chapter 1.3). Nowadays, it is established that both complex transgene arrays and replicating RNA viruses generate dsRNA, which triggers silencing as biological response.

A related phenomenon in animals was first discovered in the nematode worm Caenorhabditis elegans as a response to double-stranded RNA, which resulted in sequence-specific gene silencing. It was found that in an antisense RNA approach to inhibit gene expression, sense RNA, that was used as a control, was as effective as antisense RNA for specific silencing of the target gene (Guo & Kemphues, 1995). In a subsequent study, Fire et al. reasoned that the
in vitro transcribed RNA preparations used by Guo et al. in their antisense studies were not purely single-stranded RNA and that dsRNA in these preparations might be the key trigger of silencing (Fire et al., 1998). Their breakthrough was to test the synergy of sense and antisense RNAs. They could show that the dsRNA mixture was at least 10-fold more potent as a silencing trigger than the sense or antisense RNAs alone. This phenomenon was named RNA interference (RNAi), distinguishing it mechanistically from classical antisense-mediated suppression. Silencing by dsRNAs had a number of remarkable properties: RNAi could be induced by injection of dsRNA into the C. elegans gonad or by introduction of dsRNA through either soaking of worms in dsRNA solution or feeding them with bacteria engineered to express it (Timmons & Fire, 1998). Genetic and biochemical studies have now confirmed that RNAi, co-suppression and virus-induced gene silencing share mechanistic similarities, and that the biological pathways underlying dsRNA-induced gene silencing exist in many, if not most, eukaryotic organisms (Hannon, 2002; Denli & Hannon, 2003; Meister & Tuschl, 2004; Novina & Sharp, 2004).

The initial observations in C. elegans were consistent with dsRNA-induced gene silencing operating at the post-transcriptional level. Exposure to dsRNAs resulted in loss of the complementary mRNAs, and promoter and intronic sequences were largely ineffective as silencing triggers (Fire et al., 1998). A post-transcriptional mode was also consistent with data from plant systems in which exposure to dsRNA, for example in the form of an RNA virus, triggered depletion of mRNA sequences without an apparent effect on the rate of transcription (Jones et al., 2001). Indeed, viral transcripts themselves were targeted, despite the fact that they were synthesized cytoplasmically by transcription of RNA genomes (Ruiz et al., 1998). These studies led to the notion that RNAi induces degradation of complementary mRNAs.

In parallel, another mechanism was discovered that operates mainly at the level of protein synthesis and is now known as miRNA-mediated gene silencing. One of the first observations was that lin-4, a gene known to control the timing of C. elegans larval development, does not code for a protein but instead produces a pair of small RNAs (Lee et al., 1993). It was noticed that these lin-4 RNAs had antisense complementarity to multiple sites in the 3’-UTR of the lin-14 gene (Lee et al., 1993; Wightman et al., 1993), which were shown to be important for regulation of lin-14 by lin-4. The fact that this regulation substantially reduced the amount of Lin-14 protein without a noticeable change in lin-14 mRNA levels, supported a model in which the lin-4 RNAs pair to the lin-14 3’-UTR to specify translational repression of the lin-14 message. The shorter lin-4 RNA is now recognized as the founding member of an abundant class of tiny regulatory RNAs called microRNAs (miRNAs). The importance of
miRNA-directed gene regulation became evident as more miRNAs and their regulatory targets were discovered (reviewed by Bartel, 2004; He & Hannon, 2004; Kim, 2005). The list of biological processes regulated by miRNAs is already long and is expected to grow in the future.

A third mechanism that affects gene expression in response to the presence of dsRNA acts on the transcriptional level, i.e. on the chromatin structure (for review, see Bayne & Allshire, 2005). The initial observation of dsRNA-induced changes of chromatin was DNA methylation of endogenous sequences that shared homology with viroids in infected plants (Wassenegger et al., 1994). This observation was later followed by the finding that dsRNA sharing sequence homology with promoter regions was able to induce gene silencing which correlated with de novo methylation of promoter sequences (Mette et al., 2000). Later, studies in fission yeast and flies showed that RNA-directed transcriptional silencing is not confined to plants (see chapter 1.1.4).

After this brief historical overview, the different types of RNA-mediated gene silencing mechanisms will be described in more detail in the following three subchapters.

1.1.2 RNA interference

In the course of RNA interference, a longer dsRNA precursor is cleaved into smaller RNAs that act as guides for the silencing machinery. As a result, the expression of the corresponding gene is down-regulated due to the degradation of the complementary mRNA (see Figure 1.1A).

The presence of these small RNAs was described for the first time in a study on transgene- and virus-induced post-transcriptional gene silencing in plants. Hamilton et al. observed the formation of discrete, small RNAs of around 25 nt that were complementary to the target of silencing (Hamilton & Baulcombe, 1999). This result initiated biochemical studies to reveal the relationship between the dsRNA trigger and the resulting small RNAs. For this, Tuschl et al. developed a cell free Drosophila extract system that recapitulates many of the features of RNAi (Tuschl et al., 1999). They observed sequence-specific mRNA degradation that is promoted by dsRNA. Using the same system, it was shown that the dsRNA silencing trigger is processed to RNA segments of 21-23 nt in length (Zamore et al., 2000). The authors termed these RNAs small interfering RNAs (siRNAs). Simultaneously, it was found that extracts of Drosophila cells that were transfected with dsRNA contain a nuclease activity that
specifically degrades mRNAs homologous to this dsRNA (Hammond et al., 2000). The siRNAs were shown to be associated with this nuclease, named RNA-induced silencing complex (RISC, Figure 1.1A). Analysis of the chemical structure of siRNAs showed that they were double-stranded and contained 5′-phosphorylated termini and 2 nt 3′-overhangs (Elbashir et al., 2001c). This anatomy pointed towards a role of an RNase III ribonuclease in generating siRNAs, because this family of nucleases displays specificity for dsRNAs and generates such termini. The identity of the nuclease that cleaves the dsRNA into siRNAs in the initiation step was revealed by a biochemical approach to be indeed a member of the RNase III family of enzymes and was termed Dicer (Bernstein et al., 2001). The family of Dicer enzymes is evolutionarily conserved, and proteins from many organisms besides Drosophila, including Arabidopsis, tobacco, C. elegans, mammals and Neurospora have all been shown to recognize and process dsRNA into siRNAs of a characteristic size for the relevant species (Bernstein et al., 2001; Ketting et al., 2001). In general, RNase III enzymes exhibit specificity for dsRNA and are involved in RNA metabolism in organisms ranging from phages to animals (reviewed by Nicholson, 2003).

![Figure 1.1](image)

**Figure 1.1** siRNA- and miRNA-mediated gene silencing pathways (A) Pathways of RNA silencing originating from dsRNA or endogenous hairpin pre-miRNAs. Current models of the structural and topological features of the protein–RNA complexes involved in RNAi are shown. The color coding of the different domains is the same as in B. Red arrows indicate endonucleolytic cleavage. RISC and miRNP (micro ribonucleoprotein particle) are the effector complexes of the siRNA and miRNA pathways, respectively. (B) Typical domain structure of Dicer and Argonaute proteins. Dicer comprises a DEXH helicase, a dsRBD, a PAZ domain and a domain of unknown function (DUF283), in addition to two RNase III domains. Argonaute proteins share a PAZ domain and a Piwi domain. This figure is taken from the review that is reprinted in chapter 4.2 (Lingel & Sattler, 2005).
Three structural classes of RNase III proteins have been described. The first class is represented by *Escherichia coli* RNase III, the second by Drosha and the third by Dicer. The first class comprises the simplest RNase III proteins, each of which contains one catalytic endonuclease domain (RNase III domain) and a dsRNA binding domain (dsRBD). Members of the second class of RNase III proteins comprise Drosha and homologs and contain two RNase III domains, a dsRBD, and a long N-terminal segment, which is believed to be involved in protein-protein interaction. Drosha was shown to have a function in miRNA maturation (see next subchapter). The third class of RNase III enzymes, comprised of Dicer-like proteins, are ~200 kDa multidomain proteins. Typically, Dicers contain an N-terminal DEXH-box RNA helicase domain, a domain of unknown function (DUF283), a PAZ domain, two RNase III domains, and a dsRBD (Figure 1.1B).

Insight into the molecular details of the RNase III activity of Dicer has been derived from the crystal structure of the endonuclease domain of the RNase III homolog from *Aquifex aeolicus* (Blaszczyk et al., 2001). This structure, together with biochemical studies, has established that the catalytically active enzyme in prokaryotes comprises a homodimer of two RNase III domains. As the Drosha and Dicer enzymes each contain two RNase III domains (RNase IIIa and RNase IIIb), it was proposed that they could dimerize intra-molecularly to form an enzymatically active unit that resembles the RNase III homodimer of *A. aeolicus*. For Dicer, it was shown that the RNase IIIa and RNase IIIb domains of monomeric human Dicer do indeed constitute the active enzyme that contains only a single center for the processing of dsRNA (Zhang et al., 2004). This single processing center comprises two catalytic sites (one from each RNase III domain) and cleaves both strands of the dsRNA substrate at one location, leaving a 3’-overhang of 2 nt.

Several organisms contain more than one Dicer gene, with each Dicer preferentially processing dsRNAs originating from a specific source. *Drosophila* has two paralogues: Dicer-1 preferentially processes miRNA precursors (Lee et al., 2004b), and Dicer-2 is required for long dsRNA processing (Pham et al., 2004). Dicer-2 is stably associated with the dsRBD-containing protein R2D2 (Liu et al., 2003).

After the generation of siRNAs, the ribonucleoprotein particles (RNPs) are subsequently rearranged into the RISC, which was introduced above as being the effector complex of RNAi. For each siRNA duplex, only one strand, the so-called guide strand, is present in the active, mature RISC, whereas the other strand, named the passenger strand, is destroyed. The single-stranded siRNA in RISC guides the sequence-specific degradation of complementary or near-complementary target mRNAs. The mRNA is cleaved at a position 10 nt from the 5’-
end of the guiding siRNA (Elbashir et al., 2001b; Nykanen et al., 2001; Martinez et al., 2002). Subsequently, the 5’-mRNA fragment generated by RISC cleavage is rapidly degraded from its 3’-end by the exosome, whereas the 3’-fragment is degraded from its 5’-end by XRN1 (Orban & Izaurralde, 2005).

Like for Dicer, the identification of RISC components started with the analyses of the associated activity in extracts of Drosophila embryos and cultured S2 cells. The first subunit of RISC to be identified was the siRNA, which recognizes the substrate through Watson-Crick base pairing (Hammond et al., 2000). Nykanen et al. showed that RISC is formed in embryo extracts as a precursor complex of ~ 250 kDa which becomes activated upon addition of ATP to form a ~100 kDa complex that can cleave substrate mRNAs (Nykanen et al., 2001), whereas RISC from Drosophila S2 cells was purified as a ~500 kDa ribonucleoprotein (Hammond et al., 2000; Hammond et al., 2001). The first protein component to be identified as part of RISC was Argonaute2 (Hammond et al., 2001), a homolog of C. elegans Rde-1. Rde-1 had previously been shown to be required for RNAi in C. elegans (Tabara et al., 1999) and belongs to the Argonaute family of proteins (see chapter 1.2). In analogy to Argonaute2 in Drosophila, the human homologs eIF2C1 and eIF2C2 were isolated from HeLa cells in an attempt to purify RISC activity (Martinez et al., 2002). In two subsequent studies on the composition of RISC, three putative RNA binding proteins, the Drosophila homolog of the fragile X mental retardation protein (FMRP), dFXR, VIG (Vasa intronic gene) and a Drosophila homolog of the p68 RNA helicase (Dmp68), were shown to associate with a tagged version of Ago2 (Caudy et al., 2002; Ishizuka et al., 2002). Tudor-SN, a 100 kDa protein with 4 N-terminal Staphylococcus nuclease (SN) domains and a hybrid tudor/SN domain, was another factor demonstrated to be bound to RISC (Caudy et al., 2003). It was speculated that this protein might be part of the nucleolytic activity of RISC, but later studies revealed that the actual endonuclease in RISC is a Mg$^{2+}$ dependent nuclease (Martinez & Tuschl, 2004; Schwarz et al., 2004), which ruled out Tudor-SN as it is Ca$^{2+}$ dependent. A set of structural and biochemical studies showed then that Ago2 itself is the protein responsible for the endonucleolytic activity that cleaves the mRNA in RISC, now referred to as Slicer (see Figure 1.1A and chapter 1.2). In addition to cleaving the target mRNA, it was recently demonstrated that Ago2 also cleaves the passenger strand (Matranga et al., 2005; Rand et al., 2005), thereby liberating the single-stranded guide strand. These results indicate that Ago2 receives directly the double-stranded siRNA duplex from the RISC assembling machinery instead of binding a single-stranded siRNA after its separation by a helicase, as it was proposed before (Tomari et al., 2004a).
The possibility of specific down-regulation of target genes by RNAi provides a novel and now frequently used tool to study gene function. In contrast to other functional inhibition approaches like the use of antibodies or small molecule antagonists, RNAi has much higher specificity and a less restricted applicability. For example, the transfection of mammalian cells with siRNAs results in the potent, long-lasting post-transcriptional silencing of the specific target genes (Caplen et al., 2001; Elbashir et al., 2001a). Beyond their value for validation of gene function, siRNAs also hold great potential as gene-specific therapeutic agents.

1.1.3 miRNA pathway

MicroRNAs (miRNAs) are single-stranded RNAs of 19-25 nucleotides in length and are generated from endogenous hairpin-shaped transcripts by the same RNase III-type enzyme Dicer (Figure 1.1A, Ambros et al., 2003; Bartel, 2004). miRNAs function as guide molecules in post-transcriptional gene silencing by base pairing with target mRNAs, which leads to mRNA cleavage or translational repression (Figure 1.1A). miRNAs have been shown to play key roles in diverse regulatory pathways, including control of developmental timing, haematopoietic cell differentiation, apoptosis, cell proliferation and organ development (reviewed by Ambros, 2004; Bartel, 2004; He & Hannon, 2004; Kim, 2005).

miRNA genes can be clustered in the genome as it is observed for Drosophila (Lagos-Quintana et al., 2001). In contrast to that, the majority of worm and human miRNA genes are isolated and not clustered (Lim et al., 2003a; Lim et al., 2003b). Transcription of miRNA genes is mediated by RNA polymerase II, and the transcripts, called primary miRNAs (pri-miRNAs) were shown to contain both cap-structures and poly(A)-tails (Cai et al., 2004; Lee et al., 2004a). The pri-miRNAs are usually several kilobases long and contain one or several local hairpin structures. The stem loop structure is cleaved by the nuclear RNase III Drosha to release the precursor miRNA (pre-miRNA) (Lee et al., 2003, Figure 1.1). The hairpin precursors are usually ~60-80 nt in animals, but the lengths are more variable in plants. The conserved family of Drosha ribonucleases was introduced in the previous chapter. Drosha forms a large complex of ~500 kDa in D. melanogaster (Denli et al., 2004) or ~650 kDa in humans (Gregory et al., 2004; Han et al., 2004). In this complex, called microprocessor, Drosha interacts with its cofactor, the two dsRBDs containing DiGeorge syndrome critical region gene 8 (DGCR8) protein in humans, which is also known as Pasha in D. melanogaster.
and *C. elegans* (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). The tertiary structure of pri-miRNAs seems to be the primary determinant for substrate specificity. Apparently, the Drosophila complex can measure the length of the stem, because the cleavage site is located approximately two helical turns (~22 nucleotides) from the terminal loop (Zeng et al., 2005). When Drosophila excises the hairpin miRNA precursor, a 5′-phosphate and a 2 nt 3′-overhang remain at the base of the stem (Lee et al., 2003).

After being processed by Drosha, the pre-miRNA is exported out of the nucleus into the cytoplasm by the nuclear export factor Exportin5 (Exp5) (Yi et al., 2003; Bohnsack et al., 2004; Lund et al., 2004). Exp5 is a member of the RanGTP-binding transport receptors and was initially identified as a factor exporting the eukaryotic elongation factor 1A together with tRNAs (Calado et al., 2002). It was shown that the interaction between Exp5 and the RNA is direct and involves contacts to the tRNA or miRNA double-stranded stem (Lund et al., 2004). Following their export from the nucleus, pre-miRNAs are subsequently processed into ~22 nucleotide, usually imperfect miRNA duplexes by the cytoplasmic RNase III Dicer (Grishok et al., 2001; Hutvagner et al., 2001; Knight & Bass, 2001). In *Drosophila*, the pre-miRNA hairpins are cleaved by Dicer-1, which was shown to act in concert with the double-stranded RNA binding domain protein Loquacious (Forstemann et al., 2005; Saito et al., 2005). As a result of being cleaved in two subsequent steps by an RNase III enzyme, both ends of the miRNA duplex have 2 nt 3′-end overhangs and 5′-phosphates. Like in the course of RNAi, the double-stranded miRNA is incorporated into an Argonaute protein containing complex, which is referred to as miRNP (Figure 1.1A). For example, the first described human miRNP showed that miRNAs are present in a complex with the Argonaute eIF2C2, together with the helicase Gemin3, and Gemin4 (Mourelatos et al., 2002).

MicroRNA loaded miRNPs can direct the downregulation of gene expression by either of two post-transcriptional mechanisms: mRNA cleavage or translational repression. According to the current model, the choice of the post-transcriptional mechanisms is not determined by whether the small silencing RNA originated as an siRNA or a miRNA, but by the degree of complementarity with the target. Once incorporated into a cytoplasmic RISC, the miRNA will specify cleavage, if the mRNA has sufficient complementarity to the miRNA. However, it will repress productive translation if the mRNA does not have sufficient complementarity to be cleaved but does have a suitable constellation of miRNA complementary sites (Doench et al., 2003; Zeng et al., 2003). Another determinant is the nature of the Argonaute protein that is present in the particular miRNP, as it was demonstrated that not all Argonaute proteins are competent for substrate cleavage (see chapter 1.2). In plants, miRNAs base pair with
messenger RNA targets by precise or nearly precise complementarity and direct cleavage and destruction of the target mRNA (Rhoades et al., 2002). When a miRNA guides cleavage, the cut is at precisely the same site as that seen for siRNA-guided cleavage, i.e., between the nucleotides pairing to residues 10 and 11 of the miRNA (Llave et al., 2002). In contrast, most animal miRNAs are imprecisely complementary to their mRNA targets, with targets sites being generally in the 3’-UTRs of the target mRNA. It was shown that not the complete length of the miRNA is equally important for this interaction. Positions 2-7 form the critical “seed”, and base pairing in this region is most important for target recognition. This fact is also employed for the prediction of miRNA target sites (Enright et al., 2003; Stark et al., 2003; John et al., 2004; Brennecke et al., 2005). Initially, it was believed that animal miRNAs act only by inhibiting protein synthesis without affecting mRNA levels. More recently, it was established that a miRNA can also act to guide the miRNP for cleavage of the cognate mRNA (Hutvagner & Zamore, 2002), resulting in a change of mRNA levels (Lim et al., 2005). Thus, it seems that miRNA-guided translational regulation and targeted mRNA degradation are used simultaneously as natural regulatory mechanisms.

Recent observations concerning the localization of human Argonaute proteins provide a possible and attractive explanation of the apparent repression of protein synthesis mediated by miRNAs. It was shown that miRNA loaded Argonaute proteins and their target mRNAs localize to mammalian processing bodies (P-bodies) (Liu et al., 2005b; Sen & Blau, 2005), which were found previously to contain untranslated mRNAs and to serve as sites of mRNA degradation (Ingelfinger et al., 2002; Eystathioy et al., 2003; Sheth & Parker, 2003). Such localization could potentially explain in part or completely the translational repression and degradation. Interestingly, one of the marker proteins of P-bodies, GW182 (Eystathioy et al., 2002; Ingelfinger et al., 2002; Sheth & Parker, 2003), was found in a subsequent study to be required for miRNA-mediated gene silencing (Rehwinkel et al., 2005). This emphasized the role of P-body components in the miRNA pathway. Now, Argonaute proteins were shown to interact physically with GW182 and it was further demonstrated that the structural P-body integrity is crucial for miRNA-mediated silencing (Liu et al., 2005a).
1.1.4 RNA-mediated transcriptional gene silencing

As outlined in chapter 1.1.1, RNA silencing mechanisms have been shown to be involved in heterochromatin formation and transcriptional gene silencing (TGS). RNA-mediated heterochromatin formation appears to be a natural epigenetic gene regulation mechanism. This mechanism is believed to be active in most eukaryotes in response to environmental changes and controls heritable changes in gene expression that are not caused by mutations (for reviews, see Baulcombe, 2004; Lippman & Martienssen, 2004; Matzke & Birchler, 2005; Wassenegger, 2005).

Most heterochromatin is found near centromeres and telomeres, and consists of tandem and satellite repeats, which are sometimes interrupted by transposable elements. The expression of heterochromatin is silenced by conserved epigenetic modifications of histones (methylation, acetylation, phosphorylation and ubiquitination) and DNA (methylation) (Jaenisch & Bird, 2003). DNA methylation is absent, or nearly absent, in yeast, flies and nematodes, but a link between DNA methylation and histone methylation is well established in fungi apart from yeast, and in animals and plants.

The main fraction of endogenous siRNAs in Arabidopsis corresponds to transposons and repeats whose histones and DNA are heavily methylated. The initiator dsRNA may be produced by bidirectional transcription or transcription of inverted repeats. Another source of dsRNA can be the activity of an RNA-dependent RNA polymerase (RdRP), which makes dsRNA from a single-stranded RNA template. RNA-dependent RNA polymerases are found in many organisms including C. elegans and plants, where they were shown to be required for the amplification and systemic spreading of RNA-mediated silencing (Dalmay et al., 2000; Sijen et al., 2001). However, RdRP genes have not been identified in insects and mammals.

RNA-directed DNA methylation (RdDM) was the first RNA-guided epigenetic modification of the genome to be discovered. Originally detected in viroid-infected tobacco plants (Wassenegger et al., 1994), RdDM was subsequently shown to require a dsRNA that is processed into small RNAs of 21-24 nt, therefore reinforcing a link with RNAi (Mette et al., 2000). dsRNAs which contain sequences that are homologous to promoter regions can then trigger promoter methylation and transcriptional gene silencing. The Arabidopsis Dicer-like 3 protein and the Argonaute proteins Ago4 and Ago1 were shown to be involved in de novo methylation or maintenance of methylation (Hamilton et al., 2002; Zilberman et al., 2003), establishing the role of RNAi components in transcriptional gene silencing. The existence of RdDM in other organisms is not clarified yet.
In *Schizosaccharomyces pombe*, heterochromatin consists of simple transposon-derived tandem arrays, which surround the central-core centromeric region of each chromosome. The derepression of the centromeric outer-transposon repeats in mutants deficient for components of the RNAi machinery led to the proposal that small RNAs function as guides to target the chromatin modifications that are typical of heterochromatin (Volpe et al., 2002; reviewed by Grewal & Rice, 2004). Small interfering RNAs (siRNAs) are generated from centromeric dsRNAs by the RNase III-type endonuclease Dicer. It was demonstrated later that a nuclear effector complex known as RITS (RNA-induced initiation of transcriptional gene silencing) complex is involved in targeting chromatin modifications. RITS contains siRNAs that originate from heterochromatic regions, such as centromeres, and three identified proteins: Ago1, Chp1 (a centromere-associated chromodomain protein), and Tas3 (a serine-rich protein that is specific to fission yeast) (Verdel et al., 2004). Binding of the complex enables recruitment of chromatin modifying proteins. Argonaute proteins were also shown to be involved in heterochromatin formation in *D. melanogaster* (see next chapter).
1.2 Argonaute proteins

As described in the previous chapter, Argonaute proteins constitute core components of RNA-mediated gene silencing mechanisms. They form a highly conserved family of proteins with members found to be involved in various biological pathways ranging from RNAi to development and stem cell determination and maintenance (reviewed by Carmell et al., 2002). In the following subchapters, the functional diversity of Argonaute proteins will be introduced briefly, followed by a more detailed description of the *Drosophila* Argonaute1 and Argonaute2 proteins, with an emphasis on the role of Argonaute2 as Slicer. At the end, recent structural insights into archaeal Argonaute and Piwi proteins will be provided. As the structural and functional characterization of the PAZ domain was a main objective of this thesis, this domain will only be introduced briefly without giving further details (for this, see chapters 2.1, 2.2 and the discussion).

1.2.1 Argonaute proteins: a conserved protein family with diverse functions

The founding member of the Argonaute protein family is the Argonaute1 protein of *Arabidopsis thaliana*. It was identified during an examination of ethyl methanesulfonate (EMS) mutagenized *A. thaliana* populations for abnormal leaf morphology (Bohmert et al., 1998). In this study, it was found that *AtAgo1* mutation results in pleiotropical effects on general plant architecture. Because of very narrow rosette leaves that looked similar to squids, the authors called the identified genetic locus *argonaute*.

Argonaute family members constitute ~100 kDa highly basic proteins that contain two signature domains, namely the PAZ and the Piwi domain (Figure 1.1B). Conservation of amino acids in the Piwi domain of Argonaute proteins was first mentioned by Cox et al. when analyzing the *Drosophila* Piwi (P-element induced wimpy testis) protein (Cox et al., 1998). They showed that the central and C-terminal region of these proteins are well conserved and defined a ~40 amino acid long, highly conserved C-terminal region, which they called the Piwi box. In a subsequent bioinformatics study it was found that the Piwi box is part of a larger ~300 amino acid domain which was termed Piwi domain (Cerutti et al., 2000). The authors also demonstrated that the Piwi domain is not restricted to eukaryotes, but is also
found in prokaryotes. In this initial study, no function could be proposed for the conserved Piwi domain on the basis of sequence homology.

A central region of the *Drosophila* Piwi protein was identified to show a high level of homology with other Argonaute proteins and with a region found in Dicer proteins (Cerutti et al., 2000; Bernstein et al., 2001). Cerutti et al. named this novel, ~120 amino acid region PAZ domain, after three proteins containing this domain: Piwi, Argonaute and Zwille/Pinhead. The alignment indicated that PAZ domains can be grouped in two subfamilies, namely the Argonaute and the Dicer family. The PAZ domains of the latter one are characterized by an approximately 30 residue long conserved insert (see Figure S1 in chapter 2.1). Importantly, the Dicer and Argonaute protein families are the only ones containing this domain. Based on this fact and on co-immunoprecipitation experiments showing that both *Drosophila* Ago1 and Ago2 proteins interact with Dicer (Hammond et al., 2001; Caudy et al., 2002), the function of the PAZ domain was proposed to be protein-protein interaction, potentially mediating both homo- and hetero-dimerization (Cerutti et al., 2000). As the PAZ domain was described to be present only in eukaryotes, it was initially thought that also Argonaute proteins exist only in eukaryotes. However, recent structural studies on Piwi domain containing proteins of archaeabacteria revealed that Argonaute proteins are also present in prokaryotes.

At the time of this initial bioinformatics analysis of eukaryotic Argonaute protein domains, no structural information was available for any of the family members or of fragments of them. Now, after intensive structural studies on the PAZ domain and on Argonaute proteins of archaeabacteria and biochemical investigations into eukaryotic Argonaute proteins, the molecular function of both conserved domains is established and additional conserved domains have been identified and annotated, namely the N- and Mid domain (see below).

In addition to the initial report on the developmental phenotype caused by a mutation in the *ago1* gene of *A. thaliana*, Argonaute family genes have been isolated from several organisms in genetic screens for mutants that are deficient in RNAi and related phenomena, including post-transcriptional gene silencing in plants and quelling in fungi. These include the Qde2 protein of *Neurospora* (Cogoni & Macino, 1997) and the *Caenorhabditis elegans* Rde-1 protein, of which mutants are strongly resistant to RNAi but are developmentally normal (Tabara et al., 1999). In contrast, following the discovery of its pleiotropic effect on plant architecture, *Arabidopsis* Ago1 was also shown to be necessary for post-transcriptional gene silencing of transgenes as well as for co-suppression of transgenes and their corresponding homologous endogenous genes (Fagard et al., 2000). Thereby, it was established that
Argonaute proteins have functions in both development and RNA-mediated gene silencing. This participation in two classes of biological functions is reflected by the fact that Argonaute proteins can be separated according to their sequences into two subclasses: those that resemble *Arabidopsis* Ago1, and those that more closely resemble *Drosophila* Piwi (see Figure 1.2). Proteins of the Piwi subfamily are believed to be involved primarily in developmental processes, whereas the Argonaute subfamily constitutes members that are mainly implicated in RNA silencing mechanisms.

![Figure 1.2 Phylogenetic tree of Argonaute proteins](image)

The diversity of functions of Argonaute proteins in development and RNA silencing can be demonstrated by looking at the *Drosophila* Argonaute proteins. *Drosophila* contains four characterized Argonaute proteins, namely Ago1, Ago2, Piwi and Aubergine, plus one predicted from genomic DNA (Ago3). The first two (and the putative Ago3) belong to the Argonaute subfamily and the latter two are included in the Piwi subfamily. *DmAgo1* and *DmAgo2* function mainly in RNA silencing and will be discussed below. The *piwi* gene encodes a nucleoplasmic protein known to be required for self-renewal of stem cells in the male and female germline and the regulation of their division (Cox et al., 1998; Cox et al., 2000). Similar to Piwi, Aubergine (also called Sting) is expressed embryonically in the
presumptive gonad and is characterized by mutations that affect germline development (Wilson et al., 1996; Schmidt et al., 1999). Aubergine is also responsible for acting post-transcriptionally to maintain silencing of the X-linked repetitive Stellate locus that is necessary for male fertility. The Stellate locus is silenced through a homology-dependent mechanism mediated by an RNA transcribed from paralogous repeats called Suppressor of Stellate, Su(Ste) (Aravin et al., 2001). Sense and antisense transcripts of Su(Ste) can be detected and it was suggested that they form double-stranded RNA. Mutations in the aubergine gene impair silencing by eliminating the short Su(Ste) RNA (Aravin et al., 2004).

In another study, Piwi and Aubergine were also shown to have a role in heterochromatin silencing and HP1 localization in Drosophila (Pal-Bhadra et al., 2004). The authors demonstrated loss of silencing as a result of mutations in piwi, aubergine, or spindle-E (homeless), using tandem mini-white arrays and white transgenes in heterochromatin.

### 1.2.2 The Drosophila Ago1 and Ago2 proteins

*DmAgo1* was the first identified *Drosophila* homologue of *Arabidopsis* Ago1, isolated by a genetic approach to search for regulators of Wingless signal transduction. *DmAgo1* maternal and zygotic mutant embryos showed developmental defects, with malformation of the nervous system being the most prominent (Kataoka et al., 2001). Later, Ago1 was shown to be required for efficient RNAi in *Drosophila* embryos, with its particular function lying downstream of Dicer (Williams & Rubin, 2002).

Argonaute2 was identified by a biochemical approach to purify the RNAi effector nuclease from cultured *Drosophila* cells (Hammond et al., 2001). After several steps of purification, the remaining active fraction contained a ribonucleoprotein complex of around 500 kDa. The identity of Argonaute2 was then revealed by mass spectrometry sequencing. It was also shown that down-regulation of Ago2 by RNAi could suppress the silencing of a reporter, indicating that Ago2 plays an important role in the silencing pathway. Soon after that, the structure of the Ago2 PAZ domain was solved by NMR spectroscopy (see chapter 2.1), providing the first molecular insight into the function of Argonaute proteins. At the same time, the solution structure of the *DmAgo1* PAZ domain was determined (Yan et al., 2003), and later, the *DmAgo2* PAZ domain was also solved by x-ray crystallography (Song et al., 2003). Subsequent structural studies on the interaction of PAZ domains with nucleic acids revealed the basis for RNA binding and specificity of the PAZ domain (chapter 2.2, and Ma et
al., 2004), establishing a detailed characterization of the molecular function of this conserved domain.

In an attempt to distinguish the functions performed by Ago1- and Ago2-associated RISCs in *Drosophila*, Okamura et al. did a detailed study of RNA silencing pathways in embryos lacking Ago1 or Ago2, complemented with experiments in cultured cells (Okamura et al., 2004). They showed that Ago2 is an essential component for the siRNA-directed RNA interference response and is required for the unwinding of the siRNA duplex and in consequence for the assembly of the siRNA into RISC in *Drosophila* embryos. However, *Drosophila* embryos lacking Ago2, which were siRNA-directed RNAi-defective, were still capable of miRNA-directed target RNA cleavage. In contrast, Ago1, which is dispensable for siRNA-directed target RNA cleavage, was shown to be required for mature miRNA production that results in miRNA-directed RNA cleavage. This pointed already into the direction that Ago2 might be very closely associated with the mRNA cleaving activity in RISC and that Ago1 is more implicated in miRNA-mediated pathways.

### 1.2.3 *Drosophila* and human Argonaute2 act as Slicer

To figure out the relationship between *Dm*Ago2 and the endonucleolytic activity, a stringent biochemical purification of RISC activity from *Drosophila* Schneider cells was performed, and Ago2 was found the be the only protein component present in the functional RISC that was purified to homogeneity (Rand et al., 2004).

These results on *Dm*Ago2 were in agreement with simultaneously published key studies which presented compelling evidence that human Argonaute2 (*Hs*Ago2) is the previously unidentified endonuclease in human cells (Liu et al., 2004; Meister et al., 2004; Song et al., 2004). Human Ago2 (also known as eIF2C2) has been shown before to be part of purified RISC activity in HeLa cells, together with single-stranded siRNAs (Martinez et al., 2002) and to be associated with miRNA containing ribonucleoprotein complexes (Mourelatos et al., 2002). To get further insight into the specific functions of human Argonaute proteins in RNAi, Liu *et al.* investigated their roles concerning RISC activity (Liu *et al.*, 2004). By transfection and immunoprecipitation experiments followed by an assay for siRNA-guided cleavage of a complementary synthetic mRNA, they found that only *Hs*Ago2-associated RISC was able to catalyze the cleavage. This was true even though all Argonaute proteins bound the co-transfected siRNA. The authors could also reconstitute RISC-mediated cleavage
activity *in vitro* by adding an siRNA to immunopurified tagged *HsAgo2*, which was isolated from cells that were not co-transfected with siRNAs. Mutational analysis, based on the accompanying study describing the crystal structure of an Argonaute protein from the archaea *Pyrococcus furiosus* (Song et al., 2004), provided strong evidence that the cleavage activity was an intrinsic feature of Ago2, making it unlikely that the endonucleolytic activity was just contributed by a co-purified Ago2-associated factor. Similar results providing additional support for a role of *HsAgo2* in cleavage were obtained by Tuschl and coworkers. They demonstrated that purified Ago2 complexes, but not Ago1, Ago3 or Ago4 complexes, had RISC activity (Meister *et al*., 2004).

The final biochemical evidence that Ago2 constitutes the Slicer activity was provided by a recent reconstitution experiment (Rivas et al., 2005). The authors could show that human Argonaute2, expressed recombinantly in *E. coli* and purified to homogeneity, can combine with a small interfering RNA to form a minimal RISC that accurately cleaves substrate RNAs.

### 1.2.4 Structures of archaeal Argonaute and Piwi proteins

As described above, first structural insights into the molecular function of Argonaute proteins were obtained from studies on the PAZ domain, which could be expressed recombinantly and yielded soluble protein. Attempts of many laboratories to express full length eukaryotic Argonaute proteins or other fragments besides the PAZ domain in amounts necessary for structure determination were not successful up to this day. Being aware of these difficulties, several labs turned successfully to prokaryotic homologs with unknown function (reviewed by Hall, 2005).

**The structure of the Argonaute protein of *Pyrococcus furiosus* and the molecular basis of Slicer activity**

To gain insight into the possible function of the larger signature domain of Argonaute proteins, namely the Piwi domain, Song et al. performed a structural analysis of a multidomain protein from the archaeabacterium *Pyrococcus furiosus* (Song et al., 2004). It has been predicted by sequence homology that this protein contains a Piwi domain. Remarkably, the crystal structure of the protein revealed that it also comprises an unanticipated PAZ domain (Figure 1.3). Thus, containing a Piwi and a PAZ domain, the protein represents a *bona fide* Argonaute protein (now known as *PfAgo*), indicating that this protein family is not
restricted to eukaryotes. *PfAgo* consists of four distinct domains held together by an interdomain connector: an N-terminal domain, a PAZ domain, a middle (Mid) domain structurally related to the Lac repressor (Friedman et al., 1995), and a C-terminal Piwi domain (Figure 1.3). The N-terminal, middle and Piwi domains form a crescent-shaped structure, with the Piwi domain sitting in the center of the crescent (Figure 1.3B). The N-terminal domain forms a “stalk” that holds the PAZ domain on top of the concave face of the crescent, placing it in opposition to the Piwi domain. Overall, the structure is quite compact, considering the length of the polypeptide chain (i.e. 770 residues).

![Figure 1.3 Crystal structure of the *Pyrococcus furiosus* Argonaute protein. (A) Schematic diagram of the domain organisation. Domain borders are indicated by the respective amino acid numbers. (B) Ribbon representation of *PfAgo* showing the N-terminal domain (gray), the PAZ domain (blue), the middle domain (red) and the Piwi domain (green). Putative active site residues in the Piwi domain and conserved residues in the PAZ domain are drawn in stick representation (see also Figure 1.4 and Figure 3.3). This figure and all other figures displaying molecular structures were prepared with MOLMOL (Koradi et al., 1996).](image)

Despite very low sequence identity (<10%), the PAZ-like domain of *PfAgo* protein superimposes well with the known structures of eukaryotic PAZ domains. The similarities and differences will be presented in the discussion (see chapter 3.1).

The structure of the full length *PfAgo* also provided the first structural view on a Piwi domain. The Piwi domain of *PfAgo* shows a striking structural similarity and conserved secondary structure topology with the family of RNase H enzymes. This class of enzymes is known to cleave the RNA strand of an RNA:DNA hybrid (Yang & Steitz, 1995). The RNase H fold consists of a five-stranded mixed β-sheet surrounded by α-helices on both sides (see Figure 1.4).
The Piwi domain of PfAgo resembles the catalytic domain of RNase H endonucleases. (A) Crystal structure of RNase H1 from E. coli, showing side chains of the catalytic DDE triad residues and a bound Mg\textsuperscript{2+} ion. Conserved and unconserved secondary structure elements are colored green and yellow, respectively. (B) The Piwi domain in the crystal structure of PfAgo. Side chains of putative residues of a catalytic DDE triad and Arg627 are shown. Structural elements are colored as in (A). This figure is taken from the review reprinted in chapter 4.2 (Lingel & Sattler, 2005).

The active site of RNase H enzymes is comprised of a so-called DDE motif, three acidic amino acids with the side chain carboxylates positioned to catalyze the cleavage reaction. The reaction is known to require the presence of divalent cations, such as Mg\textsuperscript{2+} or Mn\textsuperscript{2+} (Steitz & Steitz, 1993; Kanaya & Ikehara, 1995; Chapados et al., 2001). The reaction mechanism differs from most ribonucleases but resembles deoxyribonucleases, leaving 3'-OH and 5'-phosphate termini (Wintersberger, 1990). Notably, Slicer activity results in the same products (Martinez & Tuschl, 2004; Schwarz et al., 2004). In the Piwi domain of the PfAgo protein, two of the three potentially catalytic carboxylates (D558 and D628) are in equivalent secondary structure positions as those found in the RNase H fold (Figure 1.4). Mutagenesis of the equivalent residues in HsAgo2 to alanine eliminated mRNA cleavage activity, as described above (Liu et al., 2004). Song et al. suggested that the active site may be completed by a third carboxylate from a conserved glutamate (E635), even though this residue is located in a different sequence position in bacterial proteins (e.g. E48 in E. coli RNase H1). The identity of the third carboxylate was revealed in a subsequent study on PfAgo, which made use of PfAgo crystals that were soaked in with Mn\textsuperscript{2+} solution. The Mn\textsuperscript{2+} ion can replace the
Mg\(^{2+}\) expected to be coordinated at the active site but is easier to detect crystallographically. This experiment indicated that the third coordinating side chain is H745 (Figure 1.5). In addition, the equivalent residue in an Argonaute protein of *Aquifex aeolicus* (*AaAgo*), D683, was coordinated by a Ca\(^{2+}\) ion along with the two other conserved acidic residues in the crystal structure of this protein (Yuan et al., 2005, see below). Mutation of the corresponding residue in *HsAgo*2, H807, also eliminated mRNA cleavage activity, confirming its importance for catalytic activity (Rivas et al., 2005).

Notably, not all Ago proteins are active as endonucleases, and the identification of the active site residues of this family of proteins explains why some Ago proteins are not cleavage competent. For example, in cleavage incompetent *HsAgo*1, the catalytic histidine is replaced by arginine and mutation of the equivalent histidine in cleavage competent *HsAgo*2 to arginine inactivates it (Liu et al., 2004; Rivas et al., 2005). Similarly, in cleavage incompetent *HsAgo* 4, one of the catalytic aspartates is replaced by glycine.

![Figure 1.5](image)

**Figure 1.5** Superimposition of the active sites of *PfAgo* (blue) and *Bh* -RNase HC (green). The RNA strand from the DNA:RNA hybrid in the *Bh*-RNase HC structure is shown in pink. Water molecules from the *PfAgo* (red) and *Bh*-RNase HC (dark pink) are shown as spheres. Gray dashed lines indicate coordination of metals A and B in *Bh*-RNase HC. Adapted from Hall, 2005.

Additional insight into the catalytic mechanism of Ago proteins was provided by the recent crystal structures of the RNase H of *Bacillus halodurans* (*Bh*-RNase HC) with its DNA:RNA hybrid substrate (Nowotny et al., 2005). The authors found that the DNA:RNA hybrid binds at the active site of *Bh*-RNase HC via two divalent cations. One is equivalent to the Mn\(^{2+}\) coordinated by two aspartates and the histidine in Ago proteins. This metal is also coordinated by two water molecules, one that is positioned ~3 Å from the scissile phosphate group and is predicted to be the nucleophile (metal A in Figure 1.5). Equivalent water molecules are found coordinating the Mn\(^{2+}\) ion in the *PfAgo* structure. A third water molecule is coordinated to the
Mn$^{2+}$ and this corresponds to a non-bridging oxygen atom in the phosphate group preceding the scissile phosphate.

A second divalent cation (metal B in Figure 1.5) is coordinated by an aspartate that coordinates both metals, a glutamate, an asparagine, a water molecule, and the RNA. It is more difficult to predict where this second metal binds in the PfAgo structure. Two possible metal ligands are E635 and R627, although they are more distant from the first metal ion than the ligands for the second metal in the Bh-RNase HC structure (Nowotny et al., 2005). The two-metal ion catalytic mechanism is likely to be conserved among RNase H family enzymes, but identification of the ligands for the second metal ion site may require a crystal structure of an Ago protein with RNA bound at the active site.

**The structures of the Piwi protein of *Archaeoglobus fulgidus* and the Argonaute protein of *Aquifex aeolicus*: insights into 5'-end recognition**

The crystal structure of a simpler archaeal, Piwi domain containing protein from *Archaeoglobus fulgidus* was solved by Parker et al. They described first the structure of the free protein together with *in vitro* binding data (Parker et al., 2004). Because of the lack of a PAZ domain, the protein is not considered to be an Argonaute protein and was called AfPiwi. In comparison to PfAgo, AfPiwi lacks also the N-terminal domain and thereby represents an N-terminally truncated version of the PfAgo. The structure confirmed the similarity of the Piwi domain fold with the RNase H fold and showed the presence of carboxylate side chains in the right position in a putative active site. Most importantly, the structure revealed that the side chains of the C-terminal residues contribute to the hydrophobic core of the protein. Interestingly, the C-terminal four residues of Ago and Piwi proteins are highly conserved as aliphatic and aromatic residues. Moreover, the C-terminal carboxylate group projects onto the molecular surface, lying at the interface between the middle and the Piwi domain. There, the carboxylate is bound to a well-ordered metal ion, which is coordinated also by other side chains of highly conserved amino acids. Binding data revealed that AfPiwi forms a distinct complex with an siRNA-like duplex. A mutation in the conserved region at the C-terminus reduced the binding affinity significantly, suggesting an important role for this region in RNA binding.

This was confirmed by two subsequent structural studies of the AfPiwi protein in complex with a double-stranded RNA solved simultaneously (Ma et al., 2005; Parker et al., 2005). Both structures revealed the molecular basis of RNA binding and of 5'-end recognition. For several reasons, the 5’-ends of siRNAs and miRNAs are important for mRNA target
recognition and definition of the site of RNA cleavage. The main reasons are that the positions 2-7 of miRNAs are the most critical region for target recognition (see chapter 1.1.3) and that a phosphate group at the 5’-end of the siRNA strand that forms the guide for mRNA target recognition has been shown to be required for efficient RNAi and for proper cleavage site selection (Nykanen et al., 2001; Schwarz et al., 2002). The crystal structures of ApPiwi with siRNA-like duplexes could provide a structural perspective on the importance of the 5’-end. In both structures, the 5’-nucleotide of the guide siRNA is unpaired and bound in a pocket made up primarily of residues in the middle domain. The first base, U1 or A1, forms a stacking interaction with Y123, a conserved aromatic position in Ago proteins (Figure 1.6). The unpairing of the first base explains its relative unimportance for specifying miRNA targets. The structures contain a highly conserved metal binding site that anchors the 5’-nucleotide of the guide RNA. The 5’-phosphate group of the guide siRNA is bound directly by side chains of four residues that are invariant in Ago proteins (Y123, K127, Q137, and K163) and by the main chain nitrogen of F138 (Figure 1.6).

The phosphate group is also bound by a divalent cation that is coordinated by Q159, the C-terminal carboxylate, a water molecule, and the third phosphate group of the guide RNA, confirming the previously observed importance of the C-terminus (see above). Mutation of the corresponding amino acids that contact the 5’-phosphate in human Ago2 resulted in attenuated mRNA cleaving activity, suggesting that 5’-end binding is a conserved function
(Ma et al., 2005). Thus, the binding of the 5’-phosphate group observed in the AfPiwi structures appears to represent a good model for the RNAi effector complex.

As mentioned above, a crystal structure of a full length Argonaute protein from *Aquifex aeolicus* (*Aa*Ago) in complex with a single-stranded rU₈ oligonucleotide was determined recently and reported together with binding and cleavage studies (Yuan et al., 2005). Although the rU₈ was required for successful crystallization, it is disordered and therefore not visible in the electron density map. The structure confirms the domain organisation of the *Pf*Ago protein. It shows a bilobal architecture, with one lobe comprising the N-terminal and PAZ domain and the other one comprising the middle and Piwi domain. The PAZ domain observed in this structure superimposes with the structures of eukaryotic PAZ domains, but is most similar to the PAZ domain of *Pf*Ago (see chapter 3.1 for discussion). The Piwi core fold of *Aa*Ago is closely related to the RNase HII domain (Lai et al., 2000). A deep basic surface pocket is lined by highly conserved residues, including an aromatic residue, a patch of basic residues and polar residues from the Mid domain, supplemented by the C-terminal carboxylate and conserved aromatic residue, which are contributed to the Mid-Piwi interface by the Piwi domain.

This highly conserved basic pocket is localized to the same region that was identified in the crystal structures of *Af*Piwi-RNA complexes to anchor the phosphorylated 5’-end of the guide RNA. This implies that this pocket in *Aa*Ago is also likely to be the site for 5’-end recognition of the guide strand in all Argonaute proteins. In addition to describing the crystal structure, the authors found that *Aa*Ago could direct the cleavage of the RNA target when pre-incubated with DNA or RNA, showing for the first time a ribonuclease activity for an archaeal Argonaute protein. Whether this demonstration of DNA-directed cleavage of RNA is reflecting the function of prokaryotic Argonaute proteins *in vivo* remains to be established. Up to now, no regulatory mechanism similar to eukaryotic RNA silencing was reported for prokaryotes.
1.3 Viral suppression of RNA silencing

RNA interference is a post-transcriptional gene regulation pathway (see chapter 1.1) that is thought to have originated from an ancient cellular defense mechanism against viral and other endogenous double-stranded RNAs in the cell, like transposons with dsRNA intermediates (Plasterk, 2002). Early studies in plants and the concomitant discovery of plant viral suppressors of RNA silencing already pointed towards a role of RNAi in antiviral defense. Most plant viruses have positive-sense, single-stranded RNA genomes (Hull, 2002). During their replication cycle, there are dsRNA intermediates present which are strong inducers of RNA silencing and are believed to be the precursors of viral siRNAs (Ahlquist, 2002). Upon incorporation of these viral-derived siRNAs into RISC, the viral RNA is specifically targeted for degradation (Baulcombe, 1999; Waterhouse et al., 1999), allowing the plant to recover from viral infection. Thus, it is not surprising that many plant viruses encode suppressors of RNA silencing (reviewed by Li & Ding, 2001; Voinnet, 2001; Ding et al., 2004; Roth et al., 2004; Voinnet, 2005), which can serve as an example of co-evolution. The expression of these suppressors is believed to be a means for viruses to overcome the host antiviral silencing defense.

The first viral suppressors of RNA silencing reported were the helper component proteinase HC-Pro and the 2b protein encoded by potyviruses and cucumoviruses, respectively (Anandalakshmi et al., 1998; Brigneti et al., 1998; Kasschau & Carrington, 1998; Li et al., 1999), which were both identified as pathogenicity determinants of synergistic viral diseases (Ding et al., 1996; Pruss et al., 1997). This provided a rationale for identifying novel viral-encoded suppressors, and the re-investigation of other factors of this type revealed that several effectively inhibit RNA-mediated silencing (Voinnet et al., 1999).

For the identification of viral suppressors, a transient expression assay using Agrobacterium co-infiltration is used, as it provides a rapid and easy test of suppressor activity (Llave et al., 2000; Voinnet et al., 2000). Agrobacterium tumefaciens is a commonly used bacterial pathogen of plants. In the assay, two different strains are used: one strain is used to induce RNA silencing of a reporter gene (usually green fluorescent protein (GFP)), and another strain is used to express the candidate suppressor. Briefly, the strategy is to co-infiltrate mixtures of the two bacterial strains into a plant leaf (usually tobacco leaves) and then examine the infiltrated patch over time for silencing of the reporter. Most of the known suppressors were identified by this simple assay (Table 1.1).
### Introduction

<table>
<thead>
<tr>
<th>Viral family</th>
<th>Virus</th>
<th>Suppressors</th>
<th>Other functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive-strand RNA viruses in plants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carmovirus</td>
<td>Turnip Crinkle virus</td>
<td>p38</td>
<td>Coat protein</td>
</tr>
<tr>
<td>Cucumovirus</td>
<td>Cucumber Mosaic virus</td>
<td>2b</td>
<td>Host-specific movement</td>
</tr>
<tr>
<td></td>
<td>Tomato Aspermy virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closterovirus</td>
<td>Beet Yellows virus</td>
<td>p21</td>
<td>Replication enhancer</td>
</tr>
<tr>
<td></td>
<td>Citrus Tristeza virus</td>
<td>p20</td>
<td>Replication enhancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p23</td>
<td>Nucleic acid binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td>Coat protein</td>
</tr>
<tr>
<td>Comovirus</td>
<td>Cowpea Mosaic virus</td>
<td>S protein</td>
<td>Small coat protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication enhancer, movement, seed transmission, pathogenicity determinant</td>
</tr>
<tr>
<td>Hordeivirus</td>
<td>Barley Yellow Mosaic virus</td>
<td>γb</td>
<td></td>
</tr>
<tr>
<td>Pecluvirus</td>
<td>Peanut Clump virus</td>
<td>p15</td>
<td>Movement</td>
</tr>
<tr>
<td>Polerovirus</td>
<td>Beet Western Yellows virus</td>
<td>P0</td>
<td>Pathogenicity determinant</td>
</tr>
<tr>
<td></td>
<td>Cucurbit Aphid-borne Yellows virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potexvirus</td>
<td>Potato virus X</td>
<td>p25</td>
<td>Movement</td>
</tr>
<tr>
<td>Potyvirus</td>
<td>Potato virus Y</td>
<td>HC-Pro</td>
<td>Movement, polyprotein processing, aphid transmission, pathogenicity determinant</td>
</tr>
<tr>
<td></td>
<td>Tobacco Etch virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Turnip Yellow virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sobernovirus</td>
<td>Rice Yellow Mottle virus</td>
<td>P1</td>
<td>Movement, pathogenicity determinant</td>
</tr>
<tr>
<td>Tombusvirus</td>
<td>Tomato Bushy Stunt virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cymbidium Ringspot virus</td>
<td>p19</td>
<td>Movement, pathogenicity determinant</td>
</tr>
<tr>
<td></td>
<td>Carnation Italian Singspot virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tobamovirus</td>
<td>Tobacco Mosaic virus</td>
<td>p30</td>
<td>Replication</td>
</tr>
<tr>
<td></td>
<td>Tomato Mosaic virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tymovirus</td>
<td>Turnip Yellow Mosaic virus</td>
<td>p69</td>
<td>Movement, pathogenicity determinant</td>
</tr>
<tr>
<td><strong>Negative-strand RNA viruses in plants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tospovirus</td>
<td>Tomato Spotted Wilt virus</td>
<td>NSs</td>
<td>Pathogenicity determinant</td>
</tr>
<tr>
<td>Tenuivirus</td>
<td>Rice Hoja Blanca virus</td>
<td>NS3</td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>DNA viruses in plants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geminivirus</td>
<td>African Cassava Mosaic virus</td>
<td>AC2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tomato Yellow Leaf Curl virus</td>
<td>C2</td>
<td>Transcriptional activator protein (TrAP)</td>
</tr>
<tr>
<td></td>
<td>Mungbean Yellow Mosaic virus</td>
<td>C2</td>
<td></td>
</tr>
<tr>
<td><strong>Positive-strand RNA viruses in animals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodavirus</td>
<td>Flock House virus</td>
<td>B2</td>
<td>Plaque formation</td>
</tr>
<tr>
<td></td>
<td>Nodamura virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Negative-strand RNA viruses in animals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthomyxovirus</td>
<td>Influenza virus A*</td>
<td>NS1</td>
<td>Poly(A) binding, inhibitor of mRNA export, PKR inhibitor</td>
</tr>
<tr>
<td><strong>DNA viruses in animals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenovirus</td>
<td>Adenovirus</td>
<td>VA1 RNA</td>
<td>PKR inhibitor</td>
</tr>
<tr>
<td>Poxvirus</td>
<td>Vaccinia virus*</td>
<td>E3L</td>
<td>PKR inhibitor</td>
</tr>
</tbody>
</table>

**Table 1.1** RNA silencing suppressors encoded by plant, insect and vertebrate viruses.

* These have only been demonstrated in heterologous systems (insects and plants). HC-Pro, helper component protease; PKR, RNA-dependent protein kinase. Adapted from Voinnet et al., 2005.

Many unrelated viral proteins have evolved silencing suppressor activities in addition to their other functions, contributing to the remarkable diversity of these factors, which have now
been identified for almost all types of plant viruses (see Table 1.1). It would not be surprising if all plant viruses are found to encode at least one suppressor of RNA silencing, provided that each of the viral proteins is assayed for possible suppression of both intra- and inter-cellular (systemic) RNA silencing.

Silencing suppression has also been documented in insect cells and was discovered through the related expression strategies and functional similarities of the Cucumber Mosaic virus (CMV) suppressor 2b and the B2 protein of Flock House virus (FHV) (Table 1.1, Li et al., 2002). FHV is an animal virus and belongs to the *Nodaviridae* family of non-enveloped icosahedral viruses. FHV infects mainly insect cells, but replicates also in yeast, plant and mammalian cells. All members of the *Nodaviridae* family have a bipartite genome comprised of a positive-sense, single-stranded RNA (Johnson et al., 2001). The first genome segment of FHV, RNA1, encodes the entire viral contribution to the viral RNA-dependent RNA polymerase and replicates autonomously in the absence of RNA2, which depends on RNA1 for replication (Ball & Johnson, 1999). During replication, each RNA1 synthesizes a subgenomic RNA3 that corresponds to the 3’-end of RNA1 and encodes two open reading frames (ORFs, B1 and B2, see Figure 1.7).

![Figure 1.7](image)

Figure 1.7 Schematic representation of the nodavirus genome organization showing the two segments of the bipartite RNA genome (RNA 1 and 2) and the subgenomic RNA3. ORFs are shown in white and gray boxes.

Deletion of the 106 amino acids encoding B2 ORF from FHV results in a drastic loss of virus accumulation in *Drosophila* S2 cells, which can be rescued by decreasing the cellular content of Ago2 (Li et al., 2002). Therefore, B2 suppresses the effect of the Ago2-dependent silencing response that normally restricts FHV accumulation. A subsequent study showed complete replication of the FHV RNA genome in *C. elegans* (Lu et al., 2005). Similarly to the Ago2 dependence in *Drosophila* cells, replication of FHV in *C. elegans* triggered potent antiviral silencing that required the Argonaute protein Rde-1, which is essential for RNAi mediated by small interfering RNAs (siRNAs) but not by microRNAs. *C. elegans* was capable of rapid virus clearance in the absence of FHV B2 protein but not when B2 was present. Thus, B2 acts as a broad-spectrum RNAi inhibitor.
The B2 protein of a related Nodavirus family member, Nodamura virus (NoV), which is infectious for insect and mammalian hosts, has been shown to block RNAi in mammalian cells (Sullivan & Ganem, 2005). Both the FHV and NoV B2 proteins bind to siRNAs and longer dsRNAs, and therefore it was proposed that they may sequester siRNAs and prevent the processing of long dsRNAs into siRNAs by the host RNaseIII-like enzyme Dicer (Lu et al., 2005; Sullivan & Ganem, 2005).

Silencing suppression clearly plays a role in establishing virus infections as most of these suppressors were previously shown to be essential for systemic virus spread in their hosts. The molecular mechanism is unknown for most of the identified suppressors, with the exception of p19 and p21, which will be introduced below. Structural and biochemical studies have been initiated to reveal the molecular function of these interesting proteins, including FHV B2 (see chapter 2.4).

Size selective recognition of siRNAs by the tombusvirus p19 protein

The linear, positive-sense ssRNA genome of tombusviruses encodes a 19 kDa protein that was shown to be a suppressor of RNA silencing (Voinnet et al., 1999; Qiu et al., 2002; Qu & Morris, 2002; Silhavy et al., 2002) and to bind RNA silencing-generated and synthetic 21 nt siRNAs \textit{in vitro} under stoichiometric conditions (Silhavy et al., 2002). Therefore, it was suggested that p19 binds siRNAs and sequesters them, thereby preventing their incorporation into RISC. This was confirmed by crystal structures of the p19 protein bound to siRNAs (Vargason et al., 2003; Ye et al., 2003), which will be described in more detail.

The p19 homodimer bound to siRNAs revealed a C-terminal domain comprised of a four-stranded $\beta$-sheet flanked on one side by three $\alpha$-helices (Figure 1.8). Two N-terminal helices are docked onto the C-terminal domain by salt bridges. The secondary structure topology of the C-terminal domain corresponds to a circular permutation of the ribosomal L1 protein, which binds dsRNA as a monomer with its $\beta$-sheet surface (Nikulin et al., 2003). Homodimeric p19 forms an extended concave anti-parallel $\beta$-sheet surface, which mediates multiple interactions with the sugar phosphate backbone of the double-stranded helical region of the siRNA (Figure 1.8). Recognition of the siRNA stem by the $\beta$-sheet surface is distinct from the mode of binding of dsRBDs, in which mainly loop regions interact with the stem of the RNA duplex (see chapter 3.3). The p19 homodimer binds the siRNA with very high affinity ($K_D \sim 0.17$ nM) (Vargason et al., 2003). The siRNA recognition is sequence-unspecific and employs hydrogen bonds, electrostatic contacts (involving phosphate and 2’-
hydroxyl groups), and aromatic-base stacking interactions. A number of conserved serine and threonine residues mediate key interactions with 2’-OH groups in the minor groove of the RNA duplex, with the axis of the RNA helix being bent ~40° towards the protein.

Figure 1.8 Crystal structure of the p19 homodimer bound to a 21 nt siRNA (PDB code 1RPU). The two N-terminal helices of p19, which mediate key interactions with the terminal base pairs and contain two important tryptophan residues, are colored orange. The inset shows essential interactions that contribute to the size-selective RNA binding by the p19 homodimer; these interactions occur symmetrically at both ends of the 19 bp RNA stem. This figure is taken from the review reprinted in chapter 4.2 (Lingel & Sattler, 2005).

Size selectivity for an RNA helix of 19 bp is defined by symmetric interactions of the highly conserved residue W42 and the more variable residue W39 with the 5’- and 3’-terminal bases, respectively, at both ends of the RNA duplex region of a 21 nt siRNA (Figure 1.8). Consistent with this, p19 does not interact with single-stranded nucleic acids or dsDNA, and has reduced binding affinity for siRNAs with longer or shorter double-stranded regions (Vargason et al., 2003; Ye et al., 2003). In one of the structures, the 2 nt single-stranded 3’-overhang makes few interactions with p19 (Vargason et al., 2003), whereas in the other structure no electron density is observed for the two nucleotides at the 3’-end (Ye et al., 2003). This suggests that recognition of the 2 nt overhang is not essential, a proposal that is consistent with the comparable binding affinities of p19 for a 19 bp RNA duplex with and without a 2 nt single-stranded 3’-overhang (Vargason et al., 2003; Ye et al., 2003). By contrast, a hydrogen bond between the indole amide of the conserved W42 and the 5’-phosphate seems to be important for binding, since the affinity of p19 for an siRNA lacking a 5’-phosphate is reduced 23-fold (Vargason et al., 2003). The major determinant of siRNA specificity is thus size selective recognition of the 19 bp duplex - a binding mode that is unprecedented and provides a fascinating example of structure-specific recognition of the key mediators of RNAi.
General nucleic acid binding of the Beet yellow virus p21 protein

Recently, the crystal structure of a second suppressor of RNA silencing was solved by Ye et al. (Ye & Patel, 2005), namely of the p21 protein of Beet Yellow virus (BYV). This virus is a positive-sense, single-stranded RNA virus that comprises a large 15.5 kb genome and is a member of the genus *Closterovirus*. The p21 protein, one of nine proteins encoded in the BYV genome, was the only one that suppressed the silencing of green fluorescence protein (GFP) induced by GFP dsRNA in the transitive *agrobacterium* infiltration experiment (Reed et al., 2003). Additional biochemical studies showed that p21 binds siRNA duplexes both *in vitro* and *in vivo*, but not single-stranded miRNAs (Chapman et al., 2004). It was also shown that p21 has no strict binding specificity for siRNA duplexes.

Ye et al. have determined the crystal structure of p21 in the free state. The structure shows that p21 forms an octameric ring of previously unknown topology, in which individual p21 monomers adopt an all α-helical structure (Figure 1.9).

![Figure 1.9](image.png)

Interestingly, the eight p21 monomer subunits associate into the closed ring structure through two types of inter-monomer interactions, namely a head-to-head and a tail-to-tail arrangement. Electrostatic potential analysis revealed that the ring surface displays a highly polarized distribution of charges: negative charges dominate the outer surface of the ring, while positive charges dominate the inner surface of the ring. Therefore, the inner surface was proposed to bind RNA via electrostatic interactions, but mutations of conserved residues in this putative binding site gave no conclusive results. Electromobility shift analysis revealed that p21 is a general nucleic acid binding protein. Thus, p21 is very different from p19 in terms of structure and RNA binding. The mechanism of RNAi suppression by p21 is therefore not revealed by the structure, because it can interact with all kinds of nucleic acids. However, it has been shown that transgenically expressed p21 can co-immunoprecipitate siRNAs *in vivo* (Chapman et al., 2004).
1.4 NMR spectroscopy: a tool to understand the function of biomolecules

Over the last 20 years, nuclear magnetic resonance (NMR) spectroscopy has developed into a very important tool in the repertoire of physical methods to study biomolecules (Wüthrich, 1986). It is employed mainly to determine the three-dimensional structure of both proteins and nucleic acids at atomic resolution, to study their dynamics in solution and to characterize molecular interactions (Krishna & Berliner, 2003). Unlike in x-ray crystallography, the investigated sample is usually in liquid phase and not in solid phase, which makes it possible to study the structural and dynamical parameters in a close to physiological environment. The obtained results are thus essentially complementary to crystallographic investigations. The main disadvantage of NMR spectroscopy is the size limit concerning the biomolecule of interest. This limitation is slowly pushed towards larger molecules but further improvement still remains one of the major tasks in method development in modern NMR spectroscopy.

In the course of this Ph.D. thesis, three main application of biomolecular NMR were employed: solution structure determination, characterization of binding events and molecular recognition, and analysis of dynamics. All three topics will be introduced in the following chapters, in addition to some basic concepts.

1.4.1 The basic principle of nuclear magnetic resonance

Nuclear magnetic resonance in condensed phase was first detected by Bloch and Purcell in 1946 (Bloch et al., 1946; Purcell et al., 1946). The physical basis for observation of nuclear magnetism by NMR spectroscopy is the nuclear spin angular momentum \( \mathbf{I} \) (representing a vector, indicated in bold), a purely quantum mechanic property that has no analog in classical mechanics. The magnitude of \( \mathbf{I} \) is given by equation

\[
|\mathbf{I}|^2 = \mathbf{I} \cdot \mathbf{I} = \hbar^2 [I(I + 1)]
\]

in which \( I \) is the angular momentum quantum number. The value of \( I \) is specified by the physical nature of the nucleus. Table 1.2 (adapted from Cavanagh et al., 1996) gives an overview of angular momentum quantum numbers \( I \), the gyromagnetic ratios \( \gamma \) (see below) and on the natural isotopic abundance of the respective nuclei that are commonly used in NMR spectroscopy of biomolecules. Most commonly, only nuclei with \( I=1/2 \) are employed.
by biomolecular NMR, which includes the nuclei $^1\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$ and $^{31}\text{P}$. Because of the low natural abundance of $^{13}\text{C}$ and $^{15}\text{N}$ (see Table 1.2), proteins are usually uniformly labelled to be enriched in these isotopes.

The value of the z component of the nuclear spin angular momentum is

$$I_z = \hbar m$$

in which $m$ is the magnetic quantum number with $m = (-I, -I+1, \ldots, I-1, I)$. Thus, $I_z$ has $2I+1$ possible values. The orientation of the spin angular momentum vector in space is quantized, because the magnitude of the vector is constant and the z component has a set of discrete values.

Nuclei that have a non-zero spin angular momentum also possess a nuclear magnetic moment $\mu$, which is collinear with the vector representing the nuclear spin angular momentum vector and is proportionally linked to it:

$$\mu = \gamma I$$

with a z component of

$$\mu_z = \gamma I_z = \gamma \hbar m$$

with $\gamma$ being the gyromagnetic ratio, a constant characteristic for a given nucleus (see Table 1.2). Because the angular momentum is a quantized property, so is the nuclear magnetic moment.

---

### Table 1.2 Properties of selected nuclei.

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>$I$</th>
<th>$\gamma (\text{T} \cdot \text{s})^{-1}$</th>
<th>Natural abundance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1\text{H}$</td>
<td>1/2</td>
<td>$2.6752 \cdot 10^8$</td>
<td>99.98</td>
</tr>
<tr>
<td>$^2\text{H}$</td>
<td>1</td>
<td>$4.107 \cdot 10^7$</td>
<td>0.02</td>
</tr>
<tr>
<td>$^{13}\text{C}$</td>
<td>1/2</td>
<td>$6.728 \cdot 10^7$</td>
<td>1.11</td>
</tr>
<tr>
<td>$^{14}\text{N}$</td>
<td>1</td>
<td>$1.934 \cdot 10^7$</td>
<td>99.64</td>
</tr>
<tr>
<td>$^{15}\text{N}$</td>
<td>1/2</td>
<td>$-2.712 \cdot 10^7$</td>
<td>0.36</td>
</tr>
<tr>
<td>$^{17}\text{O}$</td>
<td>5/2</td>
<td>$-3.628 \cdot 10^7$</td>
<td>0.04</td>
</tr>
<tr>
<td>$^{19}\text{F}$</td>
<td>1/2</td>
<td>$2.5181 \cdot 10^8$</td>
<td>100.00</td>
</tr>
<tr>
<td>$^{23}\text{Na}$</td>
<td>3/2</td>
<td>$7.080 \cdot 10^7$</td>
<td>100.00</td>
</tr>
<tr>
<td>$^{31}\text{P}$</td>
<td>1/2</td>
<td>$1.0841 \cdot 10^8$</td>
<td>100.00</td>
</tr>
</tbody>
</table>
In an external magnetic field $B$, the spin states of the nucleus have energies given by

$$E = -\mu B = -\gamma IB$$

in which $B$ is the magnetic field vector. The magnetic dipole moment $\mu$ aligns with the main axis of the field and becomes quantized with energies proportional to their projection onto $B$. In a high field NMR spectrometer, the static external magnetic field $B_0$ is directed along the $z$ axis of the laboratory coordinate system resulting in an energy term:

$$E_m = -\gamma I_z B_0 = -m\hbar\gamma B_0$$

The projection of the angular momentum of the nuclei onto the $z$ axis of the laboratory frame results in $2I+1$ equally spaced energy levels, which are known as the Zeeman levels (Figure 1.10). For the $I=1/2$ nuclei usually observed in biomolecular NMR, this results in two distinct energy levels which are commonly termed $\alpha$ and $\beta$, corresponding to spin states with $m=+1/2$ and $m=-1/2$, respectively.

At equilibrium, the different energy states are unequally populated because lower energy orientations of the magnetic dipole vector are more probable. The relative population of the states is given by the Boltzmann distribution

$$\frac{N_\alpha}{N_\beta} = e^{\frac{-\Delta E}{k_BT}}$$

The bulk magnetic moment $M$ of a macroscopic sample is given by the vector sum of the corresponding quantities of $\mu$ for individual nuclei. The small population excess in the lower energy levels gives rise to a bulk magnetization of the sample parallel to the static magnetic field. This macroscopic magnetization is referred to as the longitudinal magnetization.
Transitions between Zeeman levels can be stimulated by applying electromagnetic radiation, with the selection rule for magnetic dipole transitions being $\Delta m = \pm 1$. The energy required to excite a transition between the $m$ and $(m+1)$ state ($\alpha$ and $\beta$ state) is proportional to the static magnetic field $B_0$ (see Figure 1.11).

![Figure 1.11](image)

**Figure 1.11** A magnetic field $B_0$ acting on a nucleus with $I=1/2$ results in two different spin states which are referred to as Zeeman levels, or $\alpha$ and $\beta$ states. The energy difference $\Delta E$ between the states is dependent on the gyromagnetic ratio of the particular nucleus and on the field strength of $B_0$.

The corresponding frequency is given by

$$\omega_0 = \gamma B_0 \quad \text{or} \quad \nu_0 = \omega_0 / 2\pi = \gamma B_0 / 2\pi$$

This frequency $\omega_0$ is also called the Larmor frequency and can be imagined, in analogy to classical mechanics, as the frequency of precession of the magnetic dipole moment around the main axis of the static magnetic field. The sensitivity of NMR spectroscopy depends on the population differences between Zeeman states. Because the difference is very marginal, in the order of $1$ in $10^5$ for $^1H$ spins in an 11.7 Tesla magnetic field, NMR is an insensitive spectroscopic technique compared to other methods such as visible or ultraviolet spectroscopy. As the difference is directly correlated with the magnitude of the static magnetic field, it is an ongoing effort to build more powerful magnets, especially for biomolecular NMR.

### 1.4.2 The chemical shift

In addition to the static magnetic field $B_0$, a given nucleus experiences a local magnetic field $B_{\text{local}}$ that depends on the chemical environment. Thus, the resonance frequency of a spin is governed by the effective field $B_{\text{eff}}$

$$B_{\text{eff}} = B_0 + B_{\text{local}}$$

The local magnetic contribution is mainly constituted by the electronic environment of a given nucleus, e.g. high electron densities result in large shielding of the $B_0$ field. Thus, the bond structure and polarity of the surrounding atoms of a spin give rise to slightly different
frequencies of precession, which therefore directly correspond to different local chemical environments. This phenomenon is called chemical shift $\delta$. In practice, chemical shifts are measured in parts per million (ppm) relative to a reference resonance signal of a standard molecule (which is water in biomolecular NMR):

$$\delta = \frac{\Omega - \Omega_{\text{ref}}}{\omega_0} \times 10^6$$

in which $\Omega$ and $\Omega_{\text{ref}}$ are the offset frequencies of the signal of interest and the reference signal, respectively, and $\omega_0$ is the Larmor frequency. Differences in chemical shifts of individual nuclei are responsible for the dispersion that is observed in NMR spectra.

1.4.3 The basic NMR experiment

In a simple NMR experiment, a radio frequency with the associated temporary external magnetic field $B_1$ is applied along an axis perpendicular to the main axis of the static magnetic field $B_0$. This causes the equilibrium longitudinal magnetization $M_{z,\text{eq}}$ to precess around the main axis of $B_1$. If the radio frequency irradiation is set as a pulse that makes the magnetization to rotate by 90 degree, the magnetization is left in the xy plane, giving rise to a so-called transverse magnetization (see Figure 1.12).

![Figure 1.12](image)

**Figure 1.12** In the static external magnetic field $B_0$, the macroscopic magnetization is oriented along the z axis and has an equilibrium value of $M_{z,\text{eq}}$ (green arrow). A 90 degree radio frequency pulse along the y axis causes the magnetization to rotate around this axis resulting in a transverse magnetization $M_{xy}$ in the xy plane. Precession in the xy plane is recorded by a detector. Relaxation phenomena cause the signal to decay exponentially resulting in the detection of a free induction decay (FID). This information is transferred into a spectrum by Fourier transformation (FT).
As a result of the 90 degree pulse, zero magnetization is left along the z axis. Under the influence of the static magnetic field, the magnetization is precessing in the xy plane with different radial frequencies for individual spins (because of their different chemical shift, see above). The rotation of the magnetization is recorded by a detector in the xy plane. Because of relaxation phenomena (see chapter 1.4.8), the transverse magnetization is lost by time which results in an exponential decay of the detected signal, which is therefore called a free induction decay (FID). The intensity of the FID is detected over a certain time range. This data represents the so called time domain that is transferred into a spectrum representing the signal as a function of frequency by Fourier transformation. In case the pulse is double in time, the magnetization is flipped by 180 degree, resulting in a magnetization with a value of \(-M_z_{eq}\). Thus, by tuning the time and power of a radio frequency pulse, the magnetization can be changed by a desired angle, which is used extensively in pulse programs implemented in NMR spectroscopy.

### 1.4.4 The interaction of spins through chemical bonds: scalar coupling

Spins that are close to each other can influence each other, and these interactions are employed by NMR spectroscopy to obtain information about the chemical structure of the investigated molecule. For example, spin-spin interactions or couplings mediated by electrons that form chemical bonds between nuclei are called scalar couplings. The strength of the interaction is measured by the scalar coupling constant, \( ^nJ_{IS} \), in which \( n \) designates the number of covalent bonds separating the two nuclei, I and S. The magnitude is usually expressed in Hertz (Hz) and is independent of the strength of the external field \( B_0 \). Values are large (~10-140 Hz) for nuclei separated by a single bond and become smaller (~0-10 Hz) for two- and three-bond couplings. Typical values observed in proteins are shown in Figure 1.13.

**Figure 1.13** Typical values for one- and two-bond scalar couplings (\(^1J\) and \(^2J\)) in polypeptides are shown above the bond that mediates the spin interaction. Numbers are in red and given in Hertz (Hz). C’ is the carbonyl (CO) carbon.
The scalar coupling modifies the energy levels of the involved spin systems and modifies thereby the NMR spectrum. A coupled two-spin system constituted of spins I and S displays four energy levels corresponding to the possible combination of spin states, these are parallel orientation of spins (up-up and down-down, $\beta\beta$, $\alpha\alpha$) and anti-parallel orientations (up-down and down-up, $\beta\alpha$, $\alpha\beta$). An illustration of the energy levels is shown in Figure 1.14. Transitions between the states occur according to the selection rule $\Delta m = \pm 1$ (where $m$ is the magnetic quantum number of each state) and result in two signals at resonance frequencies of the I and S states. Introducing a scalar coupling between I and S with the value of $J_{IS}$ results in the modulation of the energy levels and each of the previously observed signals is now split into two with a distance corresponding to the magnitude of the J-coupling (Figure 1.14).

![Energy level diagram of an IS spin system.](image)

**Figure 1.14** Energy level diagram of an IS spin system. The four possible energy states (horizontal lines with numbers 1-4) of an uncoupled IS spin system are indicated on the left, whereas the same J-coupled spin system is shown on the right. Allowed transitions between them obeying the selection rule $\Delta m = \pm 1$ are indicated by arrows and result in the corresponding spectra which are outlined below. The I and S spins and the correlated resonance lines are shown in black and red, respectively. Identical J-couplings and therefore line splittings are observed for both signals, as energy levels change by the same magnitude in opposite directions.

For J-coupling to be active, one component of the coupled spin systems needs to contain transverse magnetization. When transfer of magnetization from one nucleus to another is desired, scalar coupling is allowed to evolve, thus J-couplings serve as a fundamental
mechanism to transfer magnetization between spins in NMR spectroscopy. In case coupling of spins is not wanted, the spins of a coupled heteronuclear spin system can be inverted by applying a 180 degree pulse on one of the components (decoupling).

1.4.5 Resonance assignment in proteins

In order to be able to correlate the information in the NMR spectrum with a specific atomic position in the molecule, the resonance frequency or chemical shift of each atom has to be assigned. The sequence specific assignment of NMR resonances requires many different experiments taking days of spectrometer time. The resonance assignment of the backbone atoms of a protein provides already information about the secondary structure and is a necessity to do interaction studies (see chapter 1.4.10) and to analyze the dynamics of the protein (chapter 1.4.8). Nevertheless, backbone assignment does not reveal much of information about the three dimensional structure of a protein and the complete assignment is a necessary prerequisite to achieve this goal.

Assignment of proteins requires in general uniformly isotopically labelled protein samples. This is because of the large amounts of protons in proteins that give rise to very crowded spectra. For that reason, the signals are correlated with the resonances of heteronuclei, namely $^{15}$N and $^{13}$C, and thereby dispersed over additional frequency axes. As the natural abundance of these nuclei is very low (see Table 1.2), protein samples have to be enriched for these isotopes. This is achieved by recombinant expression of the protein in bacteria grown in isotope enriched media which results in the uniform incorporation of $^{15}$N or $^{15}$N/$^{13}$C. The following description is based on a $^{15}$N/$^{13}$C-labelled protein sample.

Resonance assignment is most commonly based on triple-resonance experiments that correlate different nuclei by the transfer of magnetization via the bonds connecting them, i.e. via J-couplings (Sattler et al., 1999) (see chapter 1.4.4). Besides the advantage of increased resolution, this type of experiments also provide better signal to noise ratio. The strategy is usually to assign the atoms of the protein backbone first, and then the atoms of the amino acid side chains.

**Backbone assignment**

A basic set of experiments is outlined below to assign the resonances of the backbone atoms and to obtain connectivities between individual amino acids. Due to the fact that the chemical shifts of the C$_\alpha$ atoms are often not dispersed enough for unambiguous assignment, the C$_\beta$
resonance is normally also employed to obtain reliable assignments. A sufficient set consists of four experiments: the HNCA-, HN(CO)CA-, CBCANH-, and CBCA(CO)NH-experiments (Figure 1.15).

**Figure 1.15** Summary of NMR experiments for the resonance assignments of protein backbone atoms. Correlated spin systems are boxed for the respective experiments, spin systems which are employed to transfer magnetization without chemical shift evolution are outlined in gray and indicated in brackets in the respective pulse sequence names.

All of them rely on the correlation of the $C_{\alpha}$ and/or $C_{\beta}$ atoms with the amide nitrogen and proton of either the same amino acid (i) and/or the previous amino acid in the polypeptide sequence (i-1). The brackets represent thereby a step in the experiment at which the chemical shift of the specified nucleus is used only for the transfer of magnetization but not evolved. Experiments involving a transfer over the backbone carbonyl carbon (CO) therefore correlate only the carbon atoms of the previous amino acid (i-1) with a given amide group. In contrast, experiments without this step reveal the chemical shifts of both the previous and the same residue. Therefore, a comparison of peaks can be used to get connections between residues and to assign the resonances of the $H_N$, $N$, $C_{\alpha}$ and $C_{\beta}$ atoms.

**Side chain assignment**

In a second step, triple-resonance experiments are employed to assign carbon and proton chemical shifts of individual side chains. In general, magnetization is initially created on aliphatic side chain protons and transferred to the directly bound carbon nuclei. There, it is allowed to propagate or mix along the whole of the aliphatic spin system via a transfer termed TOCSY (total correlation spectroscopy), redirected to the amide nitrogen spin system via the carbonyl carbon (CO) and eventually transferred from there to the $H_N$ protons for detection. These experiments are referred to as (H)CC(CO)NH and H(CC)(CO)NH for correlation of the aliphatic carbon resonances and the side chain protons of the previous amino acid (i-1) to the backbone amide of residue i, respectively (see Figure 1.16).
In addition, a HCCH-TOCSY correlates all aliphatic proton resonances with all carbon spins systems within the same amino acid residue.

1.4.6 The interaction of spins through space: dipolar coupling and residual dipolar coupling

A second type of interaction between spins is the direct dipole-dipole interaction which is referred to as dipolar coupling. The magnetic field caused by one nucleus affects the magnetic field at the site of the other nucleus. More specifically, the z component of the dipolar field of nucleus $P$ will change the resonance frequency of nucleus $Q$ by an amount that depends on the internuclear distance and on the orientation of the internuclear vector relative to the static field $B_0$. This results in a splitting of resonances with a separation in frequency by

$$D_{PQ} = D_{PQ}^{\text{max}} \left(3 \cos^2 \theta - 1 \right)/2$$

where $\theta$ is the angle between the internuclear vector and $B_0$ and $D_{PQ}^{\text{max}}$ is the doublet splitting that applies for the case where $\theta$ is zero. Thus, the dipolar splitting $D_{PQ}$ provides direct information of the internuclear vector. Dipolar coupling splittings are very large in solid phase NMR, but because of the fast reorientation and isotropic orientation in an isotropic solution, dipolar couplings are usually averaged to zero and therefore not visible in liquid state NMR spectroscopy. Nevertheless, information about the orientation of bonds can now be even obtained from samples in solution because of the availability of tunable, weak alignment, which results in residual dipolar coupling.
Residual dipolar coupling

Recently, NMR methods have been developed which allow the extraction of structural restraints characterizing long-range order. Traditionally, structure determination was based on distance restraints derived from quantification of NOESY spectra (see chapter 1.4.7) as well as torsion angle restraints from measurement of $^3J$-couplings (chapter 1.4.4 and 1.4.9). A key limitation inherent to this approach concerns the strictly local nature of these parameters, since they solely define distances and angles between atoms close in space within the structure.

In particular, residual dipolar couplings (RDCs), which contain information about the orientation of the internuclear, usually one-bond vectors relative to the molecular susceptibility tensor, have been introduced to derive angular restraints and are now employed in NMR based structure calculations (Tolman et al., 1995; Tjandra & Bax, 1997; Tjandra et al., 1997; reviewed in Bax, 2003). Besides constraining local geometry, dipolar couplings also have a global ordering character as they restrain all bond orientations relative to a common frame, independent of their individual localization in the structure.

As described before, the source of this structural information is the direct dipole-dipole interaction between two nuclei, which is normally averaged to zero in solution. In the case of weak, partial anisotropic alignment, however, a certain residual orientation of the dipole-dipole vector in the magnetic field remains. Therefore, the major breakthrough with respect to the potential use of dipolar coupling derived structural restraints was the demonstration that tunable degrees of molecular alignment can be achieved by placing the molecule under investigation into a dilute liquid crystalline media. The alignment relies on the steric interaction between the highly aligned medium and the biomacromolecules, which induces a weak alignment of the biomolecule itself. In this case, not all orientations of the molecule are equally likely to occur and a residual dipolar coupling can be observed.

The most commonly measured residual dipolar coupling in structure determination of proteins is the coupling between the $^{15}$N nucleus and a proton (Figure 1.17).

**Figure 1.17** Magnetic dipole-dipole coupling, illustrated for a $^{15}$N-$^1$H spin pair. $^{15}$N and $^1$H moments are aligned parallel (or anti-parallel) to the static magnetic field $B_0$. The total magnetic field in the $B_0$ direction at the $^{15}$N position can increase or decrease relative to $B_0$, depending on the orientation of the $^{15}$N-$^1$H vector and the spin state of the proton (parallel or anti-parallel to $B_0$). Adapted from Bax, 2003.
The measurement of $^{15}\text{N}-^{1}\text{H}$ RDCs provides information about the orientation of the bond vector of amide groups in the protein backbone (Figure 1.17). Similar RDC values therefore indicate similar orientation of the bond vectors, as observed for example for amino acids located in helices. Tjandra et al. developed a procedure to incorporate these experimental dipolar couplings into the structure calculation (Tjandra & Bax, 1997). This resulted in an substantial improvement of the percentage of backbone angles in the most favoured region of the Ramachandran plot and in considerably higher cross-validation statistics. Thus, RDC measurements are a valuable tool in structure validation and refinement and were employed in the structure determinations presented in this thesis (see results).

### 1.4.7 The nuclear Overhauser effect

In general, the nuclear Overhauser enhancement or nuclear Overhauser effect (NOE) is the fractional change in intensity of one NMR line when another resonance is irradiated in a double irradiation experiment. The NOE has its origin in population changes as a result of a particular form of relaxation (see next chapter), namely dipole-dipole cross-relaxation (Neuhaus & Williamson, 2000). This is caused by the dipolar interactions between spins, where the magnetic moment of a particular spin induces a magnetic field at the position of the interacting spin that is close in space (see previous chapter).

The coupling gives rise to modified energy levels shown in Figure 1.18 for a two-spin system IS.

![Energy diagram of a two spin system (IS) with dipolar coupling.](image)

**Figure 1.18** Energy diagram of a two spin system (IS) with dipolar coupling. The NOE is a result of $W_0$ and $W_2$ transitions which occur via spin-spin cross-relaxation.

Although the dipolar interaction is averaged to zero for molecules in solution, the interaction of spins results in relaxation, which is a second order perturbation. The transitions include $W_0$ and $W_2$ transitions, which involve the simultaneous change of both spins and thus give rise to an NOE. These transitions are forbidden in the conventional sense that they cannot be directly
excited by a radio frequency or cause directly detectable NMR signals, but they are not forbidden in the context of relaxation.

The NOE is so important for NMR spectroscopy of biomolecules because it is manifested by a through space interaction and the magnitude of it directly correlates with the distance between the interacting nuclei. The NOE is therefore used to derive proton-proton distance restraints for the structure determination by NMR. The intensity of an NOE is measured using NOE spectroscopy (NOESY) experiments. The cross-peak intensity is approximately proportional to the inverse sixth power if the inter-nuclear distance and can be observed for distances up to 6 Å.

### 1.4.8 Relaxation and dynamics

The macroscopic longitudinal and transverse magnetizations were introduced in chapter 1.4.1 and 1.4.3, respectively. The longitudinal magnetization ($M_z$) is aligned along the static field axis of $B_0$ with a certain equilibrium value. The transverse magnetization ($M_{xy}$) is perpendicular to $B_0$, around which it precesses at the Larmor frequency $\omega_0$ and it has an equilibrium value of zero. After any disturbance caused by the action of a radio frequency irradiation, two processes must occur for equilibrium to be restored: the longitudinal magnetization must return to its original value and the transverse magnetization must decay to zero. The first process is called longitudinal relaxation and the second transverse relaxation. Any nuclear spin relaxation is a consequence of coupling of the spin system to the surroundings and is induced by fluctuations in the magnetic field.

**Longitudinal relaxation and measurement of T1 relaxation times**

Longitudinal relaxation necessarily involves transitions between spin states as the longitudinal magnetization is a result of difference in populations of spin levels $\alpha$ and $\beta$ (chapter 1.4.1). The longitudinal relaxation time is also called spin-lattice relaxation time or simply T1. The term lattice means basically the surrounding that includes a set of energy levels that is generally provided by the various translations, rotations and internal motions of the molecules that constitute the sample. Because of the very large number of degrees of freedom that the lattice possesses, this set is effectively a continuum. The lattice modifies the local magnetic fields at the locations of nuclei and thereby couples the lattice and the spin system. For any given NMR transition, there will be always a possible change within the lattice that involves...
the same quantity of energy. Thus, exchange of energy between the spin systems and the
lattice brings the spin system in thermal equilibrium with the lattice and the distribution of
spins returns to the equilibrium population.

The T1 relaxation time describes the longitudinal relaxation and is measured by the inversion
recovery experiment (Figure 1.19).

Following an inversion by applying an 180 degree pulse, the value of $M_z$ equals $-M_{z,eq}$. Then,
the decay of $M_z$ to its equilibrium value will follow the simple expression

$$M_z(\tau) = M_{z,eq} \left[ 1 - 2e^{-\tau/T_1} \right]$$

After a delay $\tau$, a 90 degree pulse is applied to generate magnetization in the xy plane that is
recorded as an FID. The delay is varied incrementally to obtain a set of data to fit the T1 time.

Transversal relaxation and measurement of T2 relaxation times

In contrast to T1, the transverse relaxation is a result of loss and dephasing of individual
contributions to the macroscopic transverse magnetization. This is due to transitions, but also
due to any other process that perturbs the individual Larmor frequencies like loss of phase
coherence. The transverse relaxation time is also known as spin-spin relaxation time or T2.
As the line width of a resonance signal is related to \( T_2 \) by equation

\[
\Delta v_{1/2} = \frac{1}{\pi \cdot T_2}
\]

(\( \Delta v_{1/2} \) is the full line width at half height of the line), a shorter transverse relaxation time \( T_2 \) gives rise to a broader line width. Due to the fact that \( T_2 \) is inversely correlated to the molecular tumbling rate (see below and Figure 1.21), this is a major obstacle for NMR spectroscopy of large molecules.

As explained, \( T_2 \) is the relaxation rate of \( M_{xy} \) and \( T_2 \) times are generally measured by employing a Hahn spin-echo experiment, which is using the principle of symmetry reversal (Figure 1.20).

**Figure 1.20** The transversal relaxation time \( T_2 \) is determined by the Hahn spin-echo experiment. At equilibrium, the macroscopic magnetization (green arrow) is oriented along the z axis and has a value of \( M_{z,eq} \). A first 90 degree pulse along the y axis flips the magnetization into the xy plane. Spins are allowed to evolve for a delay time \( \tau \). The spins are inverted by a 180 degree pulse along the x axis, followed by a second delay time \( \tau \). The echo is then recorded as an FID. The \( T_2 \) time can be extracted from the magnitude of the echo, because all contributions of spin-lattice relaxation (\( T_1 \)) are cancelled out due to the symmetrical concept of the experiment.

The Hahn echo is constituted first by a 90 degree pulse that flips the magnetization in the xy plane. During the first \( \tau \) delay, the magnetization evolves according to its chemical shift and field inhomogeneity. Then a 180 degree pulse is applied, which inverts the magnetization. Following this inversion pulse, another \( \tau \) delay is applied. During this delay, the magnetization refocuses. This will give rise to an echo after the \( 2\tau \) delays which is recorded as an FID. The size of this echo will only be affected by the spin-spin relaxation processes. To obtain a whole data set, the delay \( \tau \) is incrementally changed. The Carr-Purcell-Meiboom-Gill
(CPMG) sequence, which is derived from the Hahn spin-echo sequence, can be also used to measure T2 relaxation times. This sequence is equipped with a "built-in" procedure to self-correct pulse accuracy error.

Relaxation and dynamics in biomolecules

Longitudinal and transverse relaxation rates are intimately linked to motion and the interactions between the spins and the surroundings. This is due to the fact that relaxation is a result of spin interactions that are modulated by molecular motion. These interactions involve mainly dipolar coupling (see chapter 1.4.6), and chemical shift anisotropy (CSA). Other terms are responsible for additional line broadening caused by T2. The most important is chemical exchange, which describes the interconversion of two or more states (conformations) into each other. Thus, relaxation rates are affected by intra-molecular mobility in flexible structures and correlated with the overall rotational tumbling of the molecule in solution.

Small and medium sized molecules have a fast tumbling rate which is reflected by a small correlation time $\tau_c$. The rotational correlation time $\tau_c$ represents the time that it takes for a molecule to reorient, which is much shorter for small molecules compared to large molecules, like biomolecules. For molecules having a small $\tau_c$, T2 is normally equal to T1 (Figure 1.21). For biomolecules that are usually larger in size resulting in a slower tumbling rate and a larger correlation time $\tau_c$, T2 is shorter than T1 (Figure 1.21).

![Figure 1.21](image)

Figure 1.21 Logarithmic plots of the T1 and T2 times versus the correlation time $\tau_c$ for dipolar relaxation. The curves were computed for proton resonance frequencies $\nu_o$ of 400, 600 and 800 MHz.
Because the slow tumbling limit applies for typical biomacromolecules with a molecular weight above \(\sim 6\) kDa, T2 is inversely proportional to the correlation time \(\tau_c\), which itself is proportional to the molecular weight. Thus, by measuring T2 relaxation times (see above) it is possible to estimate the molecular weight of a macromolecule in solution.

As the relaxation rates are directly linked to motion, they are an optimal probe to characterize the dynamical properties of a molecule. For proteins, the aforementioned measurements are usually set up as two dimensional experiments, correlating \(^{15}\text{N}\) nuclei with a directly bound proton. Thus, in proteins, relaxation rates are determined for the backbone amide nitrogens. Most commonly, the obtained data is represented in a plot of the ratio of the T1 and T2 times of the amide nitrogens versus the position of the corresponding residue in the amino acid sequence. If the observed value differs significantly from the average T1/T2 ratio, it indicates regions of internal motion at time scales above or below the correlation time \(\tau_c\) of the overall tumbling. Thus, information about the rotational diffusion tensor of a molecule can be extracted from relaxation data. Values of T1/T2 will differ significantly within one molecule if the shape of the molecule can not be approximated by a sphere, resulting in anisotropic diffusion. Measurements of dynamical properties were obtained of the investigated proteins in this thesis (chapter 2.1 and 2.4).

1.4.9 Structure determination with NMR derived constraints

As described in the previous chapters, a variety of structural restraints can be extracted from NMR experiments. Considering the input data for structure calculations, the most important ones are short proton-proton distances derived from NOESY experiments (Table 1.3). With increasing size of the biomolecule, overlap of cross-peaks becomes a problem for the unambiguous assignment of distances. This overlap can be reduced by the introduction of additional frequency dimensions, and therefore the NOESY spectrum is in practice usually acquired as three dimensional \(^{13}\text{C}-\) and \(^{15}\text{N}\)-edited NOESY experiments. The distances \(r_{ij}\) can be calculated from the NOE intensities \(I_{ij}\) by equation

\[
r_{ij} = r_{\text{ref}} \left( \frac{I_{\text{ref}}}{I_{ij}} \right)^{1/6}
\]

using an appropriate reference for distance calibration. In practice, the experimental distance is imprecise because of peak integration errors, peak overlap, spin diffusion and internal
dynamics. For this reason, the distance is defined to be between a lower and an upper bound. Usually, the lower bound separation of a pair of protons is set to the sum of the van der Waals radii.

In addition to NOEs, structural information can be obtained by analyzing $^3$J-coupling constants. As was first described by Karplus (Karplus, 1959), the magnitude of a $^3$J-scalar coupling constant is a function of the dihedral angle $\theta$, formed by the three covalent bonds

$$^3J = A \cos^2 \theta + B \cos \theta + C$$

The constants $A$, $B$, and $C$ depend on the particular nuclei involved in the covalent bonds. Most commonly, the coupling constant between the amide proton $H_N$ and the $C_\alpha$ bound proton $H_\alpha$ is analyzed (Table 1.3). The value of the coupling constant provides a direct correlation with the backbone dihedral angles $\Phi$ and $\Psi$.

<table>
<thead>
<tr>
<th>NMR observable</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOEs ($^{15}$N- and $^{13}$C-resolved NOESY)</td>
<td>Interatomic distances (&lt;6 Å)</td>
</tr>
<tr>
<td>$J$-couplings ($^3J{H_N, H_\alpha}$)</td>
<td>Dihedral angles (backbone angles $\Phi$ and $\Psi$)</td>
</tr>
<tr>
<td>RDCs ($H_N$-$N$)</td>
<td>Bond projection angles (backbone $H_N$-$N$)</td>
</tr>
<tr>
<td>Solvent exchange ($H_N$)</td>
<td>Hydrogen bonds</td>
</tr>
</tbody>
</table>

Table 1.3 Structure determination by NMR spectroscopy. NMR observable parameters (left) are translated into structural restraints (right), which are employed in calculations resulting in a 3D structure ensemble.

Another set of restraints is generally obtained from experimentally observed H/D exchange rates. To detect slowly exchanging amide protons which might be part of H-bonds, the protein is first lyophilized and then redissolved in 100% $D_2O$. Protected amide protons do not or only slowly exchange with the surrounding deuterium of the heavy water and can therefore be observed in a $^{15}$N-$^1$H-correlation experiment. The information that a specific proton is located in a regular secondary structure is then translated into distances between this proton and another atom, based on the previously determined secondary structure (Table 1.3).

As described in chapter 1.4.6, more recent NMR observables that are implemented in structure calculations are residual dipolar couplings. The most commonly used are the $H_N$-$N$
RDCs. They provide information about the directionality of the vector of the bond between the amide nitrogen and the amide proton.

The experimental information has to be extended by chemical knowledge about bond length, bond angles and van der Waals radii of the molecule.

The obtained experimental restraints together with the fixed chemical restraints are then applied in a simulated annealing protocol, e.g. by using the programs CNS (Brünger et al., 1998) and ARIA (Linge et al., 2001). The structure calculation is basically an iterative search for an energy minimum of a given data set. The result is a family of structures, also called the NMR ensemble, with the individual solutions fulfilling the input restraints (Table 1.3). The final ensemble of NMR structures is then refined in a shell of water molecules (Linge et al., 2003). The quality of the structure and of the input restraints can then be estimated by the similarity of these structures (precision), which is normally expressed as the root mean square deviation of the position of selected atoms in the individual structures from a mean structure.

Another set of quality checks concerns the accuracy of the structures. This includes the analysis of backbone dihedral angle values in the Ramachandran plot and of the so-called Q-factor for a set of RDC constraints.

For the structural analysis of complexes between a polypeptide and a ligand, a special type of NOESY experiment is employed. Usually, the protein under investigation is fully labelled and the ligand, may it be a peptide or a nucleic acid, is added to it as an unlabelled compound in saturating amounts. The edited-filtered NOESY experiment allows then to either select or suppress magnetization evolution of hetero-nucleus bound protons and can thereby provide NOE intensities between protons bound to $^{13}$C nuclei and protons bound to $^{12}$C (unlabelled) nuclei. Thus, short inter-molecular distances between protons of the protein and the ligand can be identified and used as constraints in structure calculations of the investigated complex (see chapter 2.2).

1.4.10 Interaction studies by NMR spectroscopy

NMR spectroscopy is widely used to study the interaction between molecules (reviewed by Zuiderweg, 2002). Compared to other techniques that report interactions between biomolecules like electromobility shift assays, Biacore experiments or calorimetry and fluorescence based assays, NMR based studies have the main advantage that not only binding is detected but also structural details of the binding process become available. This is because
the chemical shift can be used as a very sensitive probe that is reporting changes in its physico-chemical environment.

**Chemical shift perturbation mapping**

Chemical shift perturbations can be analyzed in principal for all atoms of the biomolecule that were previously assigned (see chapter 1.4.5). In practice, one is employing most frequently $^{15}\text{N},^{1}\text{H}$-correlation spectra ($^{15}\text{N},^{1}\text{H}$-HSQC experiments) to detect amides in the protein backbone that are located in or in close vicinity to the binding site. For this, the spectrum of the free and the bound form of the biomolecule under investigation are measured and compared by superimposition. Observation of chemical shift perturbations upon addition of a ligand reveals atoms in which close spatial proximity the other molecule binds. An example showing binding of ssRNA to the *Drosophila* Argonaute2 PAZ domain is shown in Figure 1.22 (see also chapter 2.1).

![Chemical shift perturbation mapping](image)

**Figure 1.22** A superimposition of $^{15}\text{N},^{1}\text{H}$-HSQC spectra for the Ago2 PAZ domain in the absence (blue) and presence (red) of a six fold molar excess of a 21 nt ssRNA. Green arrows indicate the direction of peak movements. The insert shows additional spectra at 0.5-, 1-, 1.5- and 3-fold molar excess of RNA. Significant chemical shifts changes were observed for labelled residues. This example is taken from the publication presented in chapter 2.1 (Lingel et al., 2003).

Chemical shift perturbation measurements just yield the locations of the interfaces on the individual binding partners. It is then still unknown how the partners interact on an atom-to-atom basis. This can be achieved by observing inter-molecular NOEs and employing them in calculations of complex structures (see previous chapter and chapter 2.2).
Titrations with NMR

In addition to the mapping of the interface, titrations can be carried out to estimate the affinity, stoichiometry and the kinetics of the binding. How the chemical shifts of the labelled protein change during the titration is determined by the kinetics of the interaction.

If the complex dissociation is very fast, there is only a single set of resonances observable with chemical shifts being the fractionally weighted average of the free and bound chemical shifts (Figure 1.22 and chapter 2.1). This is referred to as fast chemical exchange and is often observed for weaker interactions. By titration with increasing amounts of ligand and by following the peak position in the acquired spectra for each of the changing signals, a data set is obtained which correlates the amount of ligand with the fractional change of the NMR signal. This data can be fit to obtain the dissociation constant.

If the complex dissociation is very slow, one observes one set of resonances for the free protein and one set for the bound protein (see chapter 2.4). During the titration, the ‘free set’ will disappear and will be replaced by the ‘bound set’. This regime is referred to as slow chemical exchange and is observed usually for high affinity interactions. In slow exchange one does not automatically know to which new location the resonance has moved, and therefore a quantitative analysis of chemical shift changes is difficult unless the bound state is assigned. However, the binding constant can still be quantitated by measuring the intensities of the disappearing and/or appearing peaks as a function of the ligand concentration.

In the intermediate chemical exchange case, the frequencies of the changing resonances become poorly defined and extensive kinetic broadening sets in, which might result in the complete disappearance of resonance lines from the NMR spectrum.

The range of binding constants that can be determined in a quantitative way by NMR is set by the concentration of the interaction partner that is observed by the NMR experiment. As a rule, quantitative measures can be obtained for $K_D$ values within an order of magnitude of the concentration of the studied species.

A practical application of interaction studies by NMR is the identification of new drugs, as the design of drugs initially relies on the detection of binding activity of substances to target proteins. Binding of low molecular weight compounds to large receptor molecules is a therefore prominent example of the implementation of NMR in applied research. This concept of identification and characterization of binding compounds is commonly known as structure activity relationship (SAR) (Shuker et al., 1996).
2. Results

The major part of the results I obtained during my thesis was published in the following papers, listed in chronological order:


The following part is divided into four subchapters, one for each publication. Each of them contains a summary of the results and a statement about my contribution, followed by a reprint of the original article. The statements of contributions were confirmed by Dr. Elisa Izaurralde and Dr. Michael Sattler.
2.1 Structure and nucleic acid binding of the *Drosophila* Argonaute2 PAZ domain

**Summary of results**

The solution structure of the *Drosophila* Argonaute2 PAZ domain was solved. The construct comprised 143 residues and was chosen based on an alignment that initially defined the boundaries of this conserved domain (Cerutti et al., 2000). The protein was expressed recombinantly in *E.coli* and purified to homogeneity. NMR experiments showed that the purified protein is properly folded. A complete set of spectra was recorded and the resonances of the polypeptide backbone and side chains atoms were assigned. A refined NMR structure was determined from experimentally determined short distances within the protein (NOEs), HN-N residual dipolar couplings, backbone dihedral angles and hydrogen bonds. Measurement of relaxation times and heteronuclear {1H}-15N NOE experiments showed that 20 amino acids at the C-terminus are unstructured in solution. This information together with the determined secondary structure elements enabled it to make a structure based sequence alignment that redefines the domain borders.

An analysis of the structure revealed that the domain is composed of two structural modules, a 5-stranded β-barrel, that is covered on one side by two N-terminal helices, and on the other side by a conserved module comprising a β-hairpin and a short α-helix. Inbetween these two modules, a cleft aligned by basic and hydrophobic residues is formed. The central β-barrel is a topological variation of the oligonucleotide/oligosaccharide (OB) fold, which is often implicated in single-stranded nucleic acid binding. As a consequence, the possibility of nucleic acid binding was explored. Using NMR spectroscopy, binding of the PAZ domain to a single-stranded 21 nucleotide RNA was detected. The binding site was mapped by chemical shift perturbation analysis and located to the cleft mentioned above. Mutation of conserved residues in the binding site of the PAZ domain confirmed the binding site and identified critical residues for binding, as their mutation strongly reduced the affinity for RNA. Based on a titration experiment, the dissociation constant for the interaction of the Ago2 PAZ domain with a 21 nt RNA was determined to be around 20 µM. In addition to NMR titrations, the binding of nucleic acids was also shown by UV-crosslinking assays, in which the purified protein could be efficiently crosslinked to ssRNA. Detection of binding by electromobility shift assays (EMSA) was not successful. To determine the relative affinities of different
nucleic acid ligands, competition assays were performed with detection of binding by the crosslinking experiment. It was found that not only ssRNA, but also ssDNA binds to the Ago2 PAZ domain, the DNA even with slightly higher affinity. Furthermore, a double-stranded RNA with a 19 bp duplex region and a 2 nt 3’-overhang on each side of the duplex (an siRNA duplex) was shown to bind, whereas the 19 bp duplex without the overhangs showed a reduced affinity. This suggested that the 3’-end of nucleic acids could be critical for binding to the PAZ domain. All binding data were confirmed always by both methods, NMR spectroscopy and the UV-crosslinking assay.

**Contribution**

My contribution to the published data was to clone, express and purify the protein necessary for all the NMR measurements. In addition, I did part of the NMR data acquisition and processing. Analysis of the spectra to obtain data for the structure calculation and the structure calculation itself was done in majority by me. I also did the NMR titration experiments with RNA. I expressed the PAZ domain mutants as glutathione S-transferase (GST) fusion proteins and purified them. I also contributed to the writing of the manuscript and the preparation of figures. Overall, I estimate my contribution to the manuscript to be around 50%.
2.2 Nucleic acid 3’-end recognition by the Argonaute2 PAZ domain

Summary of results

After solving the structure of the free *Drosophila* Argonaute2 PAZ domain, one important goal was to get further insight into the interaction between the PAZ domain and nucleic acids. Therefore, the solution structures of the domain bound to single-stranded RNA and DNA oligomer ligands were determined. Based on the definition of the exact domain borders of the PAZ domain in the previous study (chapter 2.1), a construct that is 20 residues shorter than the construct investigated in the structure determination of the free protein was used. Using the previously established competition crosslinking assay, a 5 nucleotide long oligomer could be defined as a ligand and was used in NMR structure determination. Two complete sets of spectra were measured to determine the complex structures with two ligands, a DNA oligomer with the sequence CTCAC and the corresponding CUCAC RNA oligomer. As before, the structures were calculated based on experimentally determined short distances within the protein (NOEs), H\textsubscript{N}-N residual dipolar couplings, backbone dihedral angles and on hydrogen bonds. Inter-molecular short distances between the ligand and the protein revealed that only the two 3’-terminal nucleotides are bound by the PAZ domain. The interaction surface of the protein mapped essentially to the region that was identified before as the binding site for nucleic acids. In both complexes, the interaction is based mainly on hydrophobic contacts of aromatic and nonpolar residues of the protein with the bases and sugar backbone of the two 3’-terminal nucleotides. The 3’-OH group of the last nucleotide is buried deeply inside of the protein, making it impossible to extend the polynucleotide chain into the 3’-direction. These results enabled the definition of the PAZ domain as a 3’-end recognition domain.

Contribution

I prepared all samples of the *Drosophila* Argonaute2 PAZ domain needed for the determination of the two complex structures. Data acquisition and processing was done in part by me. I did all assignments of the DNA and RNA oligomers and of the protein in its bound state and analyzed the spectra to get short distances within the protein and between the PAZ domain and the nucleic acid ligands. I was involved in writing the manuscript and prepared all figures. My contribution to the publication is around 70%.
Results
2.3 NMR assignment of the *Drosophila* Argonaute2 PAZ domain

**Summary of results**

In the process of the structure determination of the *Drosophila* Argonaute2 PAZ domain, chemical shifts of the protein backbone and side chain atoms were assigned. Triple-resonance NMR experiments were performed to get essentially complete sequence-specific assignments. To allow public access to the obtained results, the chemical shifts were deposited in the BioMagResBank (BMRB) and an assignment note was published.

**Contribution**

My contribution to this publication was to do the major part of chemical shift assignment of the *Drosophila* Argonaute2 PAZ domain. Furthermore, I wrote the manuscript and prepared the figure. The overall contribution is estimated to be 80%.
2.4 The structure of the FHV B2 protein, a viral suppressor of RNAi, reveals a novel mode of dsRNA recognition

Summary of results

The structure of the Flock House virus B2 protein was solved by NMR spectroscopy. The full length protein comprises 106 amino acids and was expressed recombinantly in E.coli. NMR experiments indicated that a portion of the full length protein is not well defined. Based on secondary structure predictions and NMR spectra, C-terminally deleted versions of the protein were made. The shortest folded construct comprised 72 amino acids, lacked flexible residues and was therefore chosen for further studies. Gel filtration of the purified protein and the line width of the NMR signals suggested that both the full length and the B2 (1-72) construct are dimers in solution.

Based on triple-resonance NMR experiments, chemical shift assignment was performed for the polypeptide backbone and side chains atoms to almost completeness. The structure of the B2 (1-72) dimer was determined by structure calculations based on experimentally determined short distances in the protein (NOEs), residual dipolar couplings (RDCs), dihedral angles and hydrogen bonds.

The NMR structure shows that B2 (1-72) is a symmetric, anti-parallel dimer in solution. The dimeric conformation is confirmed by NMR relaxation experiments. One monomer is comprised of three α-helices, which are arranged in a triangular manner and thereby the N- and C-termini are placed in close proximity. Helix α1 and α2 of one monomer are anti-parallel and the interaction between them is stabilized by hydrophobic residues. The B2 dimer is formed by a head-to-tail interaction of the two monomers, with the interface being stabilized mainly by hydrophobic side chains of residues in helix α1 and α2. In addition, the dimer is held together by three electrostatic interactions.

To characterize the interaction of B2 with RNA and to map the binding surface, a double-stranded siRNA was added to the B2 dimer and 15N,1H-correlation spectra were recorded. Slow exchange in the NMR timescale between the free and the RNA-bound form of the protein was observed, which indicates that B2 binds the siRNA with a high affinity in the nanomolar range. Chemical shift perturbations were located mainly at helix α2, with some additional changes also in the small helix α3. Helix α2 contains a number of positively charged residues with one of them being arginine 54, a residue previously identified to be
critical for dsRNA binding. The spacing of positively charged side chains indicates that B2 contacts the dsRNA primarily by polar interactions with the sugar phosphate backbone. This is in-line with the reported sequence-independent binding specificity of B2. Based on a titration experiment, the stoichiometry of the dsRNA-B2 dimer complex could be determined to be 1:1. Together with the observed length of the binding site, this suggests that the siRNA binds with its helix axis oriented in parallel to the $\alpha_2$ helix axis.

**Contribution**

I cloned, expressed and purified all constructs of the B2 protein and prepared all NMR samples. I also did part of the data acquisition and processing. I did the complete chemical shift assignment, the analysis of the spectra to get constraints for the structure calculations and the structure calculations themselves. Furthermore, I performed the characterization of the dsRNA binding to the B2 protein by NMR. I was involved in writing the manuscript and prepared all the figures. I estimate my overall contribution to this publication to be 80%.
Results
3. Discussion

3.1 The PAZ domain represents a novel nucleic acid binding fold

The PAZ domain is one of the two signature domains of Argonaute proteins and was initially identified by a bioinformatics approach which searched for conserved sequence blocks in proteins involved in RNA interference and cell differentiation (Cerutti et al., 2000). It was found that the PAZ domain is exclusively present in only two protein families: the Argonaute and the Dicer protein family, both of them being known to constitute core components of the initiation and effector step of RNA silencing, respectively (Figure 1.1). In the course of RNAi or miRNA-mediated gene silencing, the small RNAs generated by Dicer are transferred into RISC to perform their function in the effector complex. A direct transfer seems likely, as Dicer was shown to interact with Argonaute proteins in several studies (Hammond et al., 2001; Caudy et al., 2002). This was the reason for the proposal that the PAZ domain might act as a protein-protein interaction domain, mediating physical contact between Dicer and Argonaute proteins.

Structure of the PAZ domain

The structure of *Drosophila* Argonaute2 PAZ domain, presented in chapter 2.1 of this thesis, provided the first molecular insight into the function of this conserved protein domain. The structure revealed that the domain has a bimodular composition, with a cleft formed inbetween the two modules. The core of the PAZ domain is formed by a central five-stranded \( \beta \)-barrel that resembles structures of known \( \beta \)-barrels, such as the Sm fold and the oligosaccharide/oligonucleotide binding (OB) fold (Figure 3.1, dark blue), which were both shown to have a function in nucleic acid binding. The Sm fold is shared by members of the LSM protein family and is known to form hexameric or heptameric rings that bind to single-stranded RNA (reviewed by Khusial et al., 2005). It constitutes a closed barrel comprising 5 anti-parallel \( \beta \)-strands with an \( \alpha \)-helix on one side (Figure 3.1B). RNA binding is found not for the monomer alone, but only when the protein is part of a multimeric ring. The binding site is located in the lumen of the ring with one nucleotide binding to each subunit (Thore et al., 2003) and is formed of the loops between strands \( \beta 2 \) and \( \beta 3 \) and between \( \beta 4 \) and \( \beta 5 \) (see Figure 3.1B, black arrows). Members of the closely related OB fold superfamily consist of two 3-stranded anti-parallel \( \beta \)-sheets, where strand \( \beta 1 \) is shared by both sheets (Figure 3.1C,
The observed variability in length among OB fold domains is due to differences in the length of the loops between the well-conserved elements of secondary structure. The nucleic acid binding subfamily of OB fold proteins is the largest within the OB superfamily, and members of this subfamily are involved in ssRNA or ssDNA binding. OB folds tend to use a common ligand binding interface centered on β-strands 2 and 3 (Figure 3.1C, arrow).

Figure 3.1 Comparison of β-barrel protein folds. The strands of the central β-barrel are colored in dark blue in all panels. (A) The DmAgo2 PAZ domain, with the conserved insertion shown in light blue. (B) Sm fold of the human Sm D2 protein (PDB code 1B34, chain B) (Kambach et al., 1999). The black arrows indicate the common Sm fold binding site. (C) OB fold of the E. coli Rho transcription termination factor (PDB code 1A8V) (Bogden et al., 1999). The black arrow indicates the common OB fold nucleic acid binding site on β2 and β3.

The OB-like barrel of the DmAgo2 PAZ domain is covered on one side by two N-terminal helices and on the other side by an inserted β-hairpin/α-helix module (Figure 3.1A, light blue), which is around 35 amino acids long and contains several highly conserved residues. One of them is a phenylalanine (F72) in strand β4, which is invariant in PAZ domains. In between the central β-barrel and the conserved module, a hydrophobic cleft can be identified. This domain architecture of a β-barrel covered by a β-hairpin/α-helix module was not observed in previously solved structures and thus represents a novel domain fold. The structure of the DmAgo2 PAZ domain reported in chapter 2.1 was confirmed by a crystal structure of an MBP-DmAgo2 PAZ fusion protein that was solved by Song et al. shortly afterwards (Song et al., 2003). The two structures determined by different methods are
essentially identical. The same is true for the PAZ domain of *DmAgo1*, that was also solved as a solution structure and published simultaneously (Yan et al., 2003). Both structures show the same secondary structure and domain architecture.

**Function of the PAZ domain in nucleic acid 3’-end recognition**

Unlike previously suggested, the PAZ domain was found to bind nucleic acids in vitro. Binding of ssRNA, ssDNA, and dsRNA with 3’-end overhangs to the *DmAgo2* PAZ domain could be detected by chemical shift perturbation analysis and crosslink experiments (chapter 2.1). Using NMR, the binding site for a single-stranded siRNA could be mapped to be in the hydrophobic cleft between the β-barrel and the β-hairpin/α-helix module. Furthermore, two critical residues for nucleic acid binding, namely F50 and F72, which are located on opposite faces of the clamp in the PAZ domain, could be identified by mutational analysis. Notably, the two mutant proteins were still folded as confirmed by NMR spectroscopy, but showed no chemical shift changes upon addition of ssRNA and were not able to form a photo-induced product in the UV-crosslinking assay. Competition experiments revealed that ssRNA and ssDNA are bound with similar affinities by the *DmAgo2* PAZ domain. Different siRNAs competed equally well in the assay, indicating that the RNA is bound with little sequence preference. The observation that blunt-ended dsRNA molecules were bound with a reduced binding affinity to the PAZ domain pointed towards a specificity for ssRNA.

These results were in-line with UV-crosslinking experiments that were reported together with the crystal structure of the *DmAgo2* PAZ domain (Song et al., 2003). The authors showed binding of a single-stranded siRNA and of double-stranded siRNAs with 3’-overhangs, but observed low affinity for blunt-ended dsRNA and double-stranded RNA with 5’-overhangs. This suggested a role of a single-stranded 3’-end in nucleic acid binding.

As the identified region of interaction includes highly conserved residues, it was suggested that also other PAZ domains function in nucleic acid binding (chapter 2.1). Indeed, in the simultaneously published study of the *Drosophila* Argonaute1 PAZ domain, Yan et al. showed binding of a small, single-stranded RNA oligonucleotide (Yan et al., 2003). The authors also reported a reduced affinity for DNA, showing that the *DmAgo1* PAZ domain discriminates between RNA and DNA, whereas the Ago2 PAZ domain does not (chapter 2.1 and 2.2). Furthermore, RNA binding was shown by NMR for the PAZ domains of human Ago1, Ago2 and Piwi proteins (Yan et al., 2003).

In agreement with that, binding of nucleic acids was detected by the crosslinking assay for the human Ago1 PAZ domain and for the PAZ domains of all four Argonaute proteins of
Drosophila (A. Lingel and E. Izaurralde, unpublished). In these experiments, PAZ domains showed a clear preference for RNA over DNA, except for the DmAgo2 PAZ domain (see above). If this difference in specificity is employed in vivo in different RISCs is presently not known.

The molecular basis of nucleic acid recognition has been revealed by NMR solution structures of the DmAgo2 PAZ domain bound to single-stranded RNA and DNA oligonucleotides and by the characterization of the binding of an siRNA mimic to the PAZ domain, as described in chapter 2.2 of this Ph.D. thesis. In both complex structures, only the two 3’-terminal nucleotides are bound at similar binding sites, which are located in the previously identified hydrophobic cleft of the PAZ domain (Figure 3.2A). This established the molecular function of the PAZ domain as 3’-end recognition. Simultaneously, the crystal structure of the HsAgo1 PAZ domain in complex with an siRNA mimic was published (Ma et al., 2004) and confirmed these results. Like in the structures of the DmAgo2 PAZ complexes, the 2 nt 3’-overhang of the siRNA-like ligand is bound in the same cleft (Figure 3.2B). Recognition of the 3’-terminal nucleotide involves a stacking interaction with a conserved aromatic residue (F292 and F72 in HsAgo1 and DmAgo2, respectively) of the β-hairpin/α-helix module. The phosphate groups of the two 3’-terminal nucleotides interact with tyrosine and histidine residues in the HsAgo1-RNA (Y309, Y314, H269 and Y277) and DmAgo2-RNA (Y44) complexes. It is conceivable that some of these residues (such as Y314 and H269, which are not conserved in DmAgo2) contribute to the selectivity of most PAZ domains for RNA over DNA. This may also be linked to the different conformations of the sugar phosphate backbone in the complexes of DmAgo2 with RNA and DNA (chapter 2.2).

Very few contacts involve the sugar of the 3’-terminal nucleotide, indicating that the 3’-end of the nucleic acid is specified mainly by steric exclusion. In the complexes of the DmAgo2 PAZ domain with single-stranded nucleic acids, the penultimate nucleotide interacts with conserved residues Y57 and K107. Notably, this binding pocket for the penultimate nucleotide corresponds to the location of a single uridine nucleotide bound to the DmAgo1 PAZ domain (Yan et al., 2003), suggesting that it is a high-affinity site. By contrast, in the complex of HsAgo1 with an siRNA mimic, the base of the penultimate nucleotide is in a different more exposed location (Figure 3.2B, Ma et al., 2004). It acts as a spacer connecting the 3’-terminal nucleotide with the dsRNA stem and makes only a few contacts with the PAZ domain. This observation indicates that there is conformational plasticity with respect to the location of the penultimate nucleotide. The mode of recognition of this nucleotide may
depend on whether the upstream nucleic acid adopts a single-stranded or double-stranded conformation. In both studies, the double-stranded stem of the different siRNA mimics extends along the β3 strand of the central β barrel of the PAZ domain (chapter 2.2 and Ma et al., 2004).

The phosphodiester backbone of the RNA strand which is bound via its 3′-end mediates numerous contacts to positively charged arginine and lysine side chains (Ma et al., 2004). These additional contacts depend on the presence of the dsRNA stem and contribute substantially to the binding affinity, as the extension of the 3′-end single-stranded overhang from two to ten uridines reduces the interaction 50-fold (corresponding to a change in the dissociation constant, $K_D$, from $\sim$2 nM to 100 nM). Notably, the complementary RNA strand, including its 5′-end, does not seem to be recognized by the PAZ domain (Figure 3.2B).
Taken together, the structural studies on the *Drosophila* Argonaute2 PAZ domain show that the PAZ domain binds the two 3’-terminal nucleotides of single-stranded nucleic acids by burying them in a hydrophobic cleft between the characteristic β-hairpin/α-helix module and a central β-barrel.

The mode of recognition of the two terminal nucleotides is unique and distinct from that used by other single-stranded nucleic acid binding domains, like the Sm or the OB fold (Theobald et al., 2003; Messias & Sattler, 2004). In strong contrast to OB and Sm fold proteins, where the solvent-exposed surface of the β-barrel is used for nucleic acid recognition (Figure 3.1), the major part of the interaction surface of the PAZ domain involves the inner part of the central β-barrel near β7. An even more significant difference is the presence of the highly conserved β-hairpin/α-helix module in the PAZ domain, which is absent in the OB fold (Figure 3.1). The burial of the 3’-end by the monomeric PAZ domain also differs from 3’-end recognition by OB fold domains, such as the recognition of telomeric DNA by the OB fold of *Schizosaccharomyces pombe* Pot1 (Figure 3.2C, Lei et al., 2003). SpPot1 engages in several base- and DNA-specific contacts, including intra-DNA interactions, consistent with sequence-specific recognition and its preference for DNA over RNA. The 3’-end of the single-stranded telomeric DNA is exposed and recognized by specific contacts between its 3’-terminal hydroxyl group and the side chains of arginine and threonine residues (Figure 3.2C).

Thus, the PAZ domain represents not only a novel nucleic acid binding fold, also the mode of selectively binding the 3’-end of nucleic acids was not preceded.

**Diversity of PAZ domains**

In eukaryotes, only Dicer and Argonaute proteins contain the PAZ domain. Unlike for the PAZ domain of Argonaute proteins, little is known about the Dicer family PAZ domains. Sequences of Dicer PAZ domains align well with the sequences of Argonaute proteins, except for a stretch of approximately 30 amino acids in Dicer PAZ domains that are inserted in the loop between the β6 and β7 strand of Argonaute PAZ domains (for sequence alignments, see Supplementary Figure 1 in chapter 2.1 and Supplementary Figure 2 in chapter 2.2). This insertion is in close proximity to the nucleic acid binding site and is rich in arginine and lysine residues. This suggests that it might be involved in nucleic acid binding, but it could modulate the function of the PAZ domain. Conservation of hydrophobic residues in secondary structure elements suggests that the overall fold of Dicer PAZ domains is similar to that of Argonaute
proteins. A proposed model for the function of Dicer PAZ domains will be discussed in the next chapter.

Initially, the PAZ domain was believed to occur only in eukaryotes, as there were no conserved sequences identified in prokaryotes (Cerutti et al., 2000). Later, crystal structures of archaeal proteins with identified Piwi domains revealed that the PAZ domain is not restricted to eukaryotes. First, Song et al. solved the structure of an Argonaute protein of *Pyrococcus furiosus* (Song et al., 2004) and found a PAZ domain which can be superimposed with eukaryotic PAZ domains, despite very low sequence identity. However, the structural similarity is not complete, and there are structural elements unique to the PfAgo PAZ domain. The main difference is that in the PfAgo protein, the β-hairpin/α-helix module is replaced by two α-helices (Figure 3.3). In addition, some loop conformations are different. Despite these variations, amino acids reported to be important for binding to the 3'-end of nucleic acids (see chapter 2.2), including those in the β-hairpin/α-helix module, are conserved. Their side chains are located in similar spatial positions in the *Pyrococcus* protein, although some of these residues are at different positions in the protein sequence and protrude from different secondary structural elements (Figure 3.3).

![Figure 3.3](image)

**Figure 3.3** Comparison of the *Drosophila* Argonaute2 PAZ domain (*DmAgo2, A*) and the PAZ domain of the Argonaute protein from *Pyrococcus furiosus* (*PfAgo, B*). Structures are shown as ribbon diagrams. The β-hairpin/α-helix module is shown in light blue. This module contains two α-helices in the PAZ domain of *PfAgo*. Side chains of residues shown to contact nucleic acids in the *DmAgo2* PAZ domain are shown and labeled. Residues in equivalent positions in the *PfAgo* protein are indicated.

As described above, key features of nucleic acid recognition by the PAZ domain involve base interactions with conserved aromatic residues. Residues Y190 and Y216 of *PfAgo* are
equivalent to Y57 and Y84 of the \textit{Dm}Ago2 PAZ domain (Figure 3.3). The position of residue F72 of the \textit{Dm}Ago2 PAZ domain, which was identified to be critical for nucleic acid recognition, is occupied by another aromatic residue, W213, in \textit{Pf}Ago. On the basis of these similarities, Song et al. suggested that the PAZ domain of \textit{Pf}Ago could also bind to single-stranded 3’-ends of RNAs.

A PAZ domain is also found in the crystal structure of the Argonaute protein of another archaea, \textit{Aquifex aeolicus} (Yuan et al., 2005). Similarly to the PAZ domain of the Argonaute protein of \textit{Pyrococcus furiosus}, the β-hairpin/α-helix module found in eukaryotic PAZ domains is replaced by two α-helices. Also like in the \textit{Pf}Ago PAZ domain, conserved residues occupy similar spatial positions with their counterparts in other PAZ domains (Figure 3.3), suggesting a conserved function in 3’-end binding. However, nucleic acid binding to archaeal PAZ domains has not been demonstrated yet and thus the function of this class of PAZ domains remains to be established.

3.2 The PAZ domain as a specificity providing module in the RNAi pathway

The structural studies on eukaryotic Argonaute PAZ domains revealed that the PAZ domain provides a unique platform for the recognition of the two 3’-terminal nucleotides in single-stranded nucleic acids. A key feature of the binding is the burial of the terminal 3’-OH group. This binding mode provides new insights into the biological role of this domain within Dicer and Argonaute proteins, which constitute the cores of RNA silencing complexes.

In the initiation step of RNA-mediated gene silencing, the RNase III enzyme Dicer generates small double-stranded RNAs. Based on the established function of the Argonaute PAZ domain, a model of PAZ function in Dicer was proposed by Filipowicz and co-workers in a recent study addressing the function of human Dicer (Zhang et al., 2004). The authors suggest that the characteristic size (~20 bp) of the siRNAs resulting from Dicer cleavage may be linked to 3’-end recognition of the dsRNA substrate by the Dicer PAZ domain. In this model, the PAZ domain recognizes the 2 nt 3’-overhang generated by Drosha in the miRNA pathway, or the last cut made by Dicer in the RNAi pathway. PAZ and RNase III domains of Dicer are spaced such that when the 3’-overhang is bound, they act as a ruler to position the cleavage site exactly 22 nt away from the 3’-end. This model is consistent with the known preference of Dicer for free ends of dsRNAs (Zhang et al., 2002). This results in Dicer
cleavage at the ends of dsRNA and pre-miRNAs, shortening the substrate consecutively by siRNA-sized segments, and not cutting internally, which would lead to the generation of multiple larger fragments (Zhang et al., 2002; Zhang et al., 2004; Vermeulen et al., 2005). In-line with this model, both the PAZ and dsRBD domains are required for full Dicer activity. Intact Dicer can slowly process blunt-ended dsRNA, however deletion of the dsRBD makes Dicer more dependent on substrates with 3’-overhangs (Zhang et al., 2004). This suggests that the dsRBD contributes to non-specific binding of incoming substrates and potentially allows Dicer to cleave long dsRNA substrates processively. As no experimental structural data is available for the PAZ domain of Dicer proteins, the exact details of the catalytic mechanism and the role of the PAZ domain must await the structure of a Dicer-dsRNA complex.

In *Drosophila*, Dicer-2 is stably associated with the dsRBD containing protein R2D2 (Liu et al., 2003), and they are together required for cleavage of dsRNA and for RISC assembly (Lee et al., 2004b; Pham et al., 2004). After cleavage of long dsRNA, the Dicer-2-R2D2-siRNA ternary complex is an assembly intermediate in the formation of RISC (Pham et al., 2004; Tomari et al., 2004b) that is proposed to recruit Ago2 directly to the siRNA. This interaction might be mediated directly by Ago2 and Dicer-2 (Tahbaz et al., 2004). The siRNA duplex is thereby transferred from the initiator complex into the RISC effector complex, as it was demonstrated recently (Matranga et al., 2005). The results on the PAZ domain of Argonaute proteins presented in this thesis show that it specifically recognizes the 2 nt 3’-overhang of siRNAs. This suggests that the PAZ domain is involved in receiving the siRNA from Dicer and to facilitate its incorporation into RISC.

The role of the PAZ domain in the subsequent catalytic cycle of RISC can be envisioned as follows: The PAZ domain of Ago2 binds the 3’-end of the strand that is considered the guide strand of the siRNA duplex. At the same time, the 5’-end of the guide strand is docked in the phosphate binding pocket of the Ago2 Piwi domain (see chapter 1.2.4). The decision which strand of the siRNA duplex functions as guide strand is determined by the orientation of the siRNA in the RISC loading complex. The passenger strand would then be cleaved by Ago2 and the resulting fragments released (Matranga et al., 2005; Rand et al., 2005). This release might be facilitated by an ATP-dependent cofactor, like the release of the products of target mRNA cleavage which was shown to be stimulated by ATP (Haley & Zamore, 2004). The resulting RISC contains Ago2 with the single-stranded guide RNA bound. This ‘mature’ RISC provides the sequence-specific binding site for the complementary mRNA. The mRNA can then base pair with the guide strand and is cleaved by the endonucleolytic activity.
constituted by the Piwi domain of the Argonaute protein (see chapter 1.2.3). The reaction cycle is completed with the release of the cleavage products.

It is not established yet that the PAZ domain interacts with the 3’-end of double-stranded or single-stranded siRNAs \textit{in vivo}. Experiments testing this are possible and will be done certainly in the future. A strong indication that the PAZ domain is important for RNAi function \textit{in vivo} is provided by results of a genetic screen in \textit{C. elegans} (Tabara et al., 1999). Worms deficient in RNAi were isolated in which only an absolutely conserved glutamate in the PAZ domain of Rde-1 is mutated to lysine. The structure of the \textit{Drosophila} Ago2 PAZ domain suggests that this mutation destabilizes the fold, which most probably also disrupts the structural integrity of the PAZ domain \textit{in vivo}.

Despite the diversity of functions attributed to the different Argonaute-associated complexes (see chapter 1.2.1), a common feature of these complexes is the use of non-coding RNAs to select their targets in a sequence-specific manner. These non-coding RNAs (siRNAs or microRNAs) have the specific features of RNAs processed by the RNase III enzyme Dicer in common, i.e. a 19-22 nucleotide double-stranded RNA body with two nucleotide 3’-overhangs and 5’-monophosphate groups. This raises the question of how siRNAs or miRNAs are discriminated from other cellular RNAs and specifically incorporated into RISCs. With the binding specificity of the PAZ domain of Argonaute proteins, it is likely that it provides a contribution to the specific recognition of the 2 nt 3’-overhang feature of siRNAs. It would then act as a selectivity filter in RISCs, allowing only small RNAs processed by Dicer to enter the effector complex and excluding all other RNAs that are present in the cell. Without such a filter, all double-stranded RNA that is present in a cell could enter the RNAi pathway. This selectivity is very important, as the incorporation of unwanted dsRNA would lead to the down-regulation of the expression of complementary genes.
3.3 The FHV B2 protein binds dsRNA in a novel mode

To counteract the cellular response to the presence of dsRNA, viruses have evolved proteins that can act as suppressors of RNA interference. The detailed mechanism of suppression is unknown for most of the identified proteins, with the exception of the p19 protein of tombusviruses (see chapter 1.3). p19 was shown to bind specifically to siRNAs and thereby blocks their incorporation into RISC (Vargason et al., 2003; Ye et al., 2003).

As part of this thesis, the solution structure of the B2 protein of Flock House virus was determined and is described in chapter 2.4. B2 was previously shown to act as a potent suppressor of RNAi in insect cells, *C. elegans* and plants (Li et al., 2002; Lu et al., 2005). The NMR structure shows that B2 is an anti-parallel, symmetric dimer in solution with each monomer composed of three \( \alpha \)-helices. Searches in databases of known structures did not result in significant hits, indicating that this dimer fold is novel. Binding of dsRNA was detected by chemical shift perturbation analysis and the binding site could be mapped to an elongated surface formed by the two \( \alpha \)2 helices of the dimer (\( \alpha \)2/\( \alpha \)2’). Notably, the helices are highly positively charged, as many lysine and arginine residues are located on this surface. Arginine 54, which has been shown to be crucial for dsRNA binding (Lu et al., 2005), is located at the centre of this binding interface, which confirmed the importance of residues in the identified binding surface. The exposed arginine and lysine residues suggest that recognition of dsRNA is mediated by non-sequence-specific electrostatic contacts with the sugar phosphate backbone of the dsRNA stem. Interestingly, the spacing of positively charged side chains corresponds to known distances in regular A-form dsRNA. R54 and K47 are separated with a spacing of approximately 10 Å (Figure 3.4A), which is approximately the distance of phosphates across the major groove. Furthermore, R36 and K62 are positioned to bridge the distance over the minor groove of A-form dsRNA. These observations, together with the determined 1:1 binding stoichiometry and the fact that the complete \( \alpha \)2/\( \alpha \)2’ surface provides the binding site for dsRNA suggests that the dsRNA binds parallel to the \( \alpha \)2 helix axis. The length of the RNA binding surface is about 45 Å, which corresponds to an approximately 17 base pair dsRNA. This proposal was confirmed by a crystal structure of a B2 dimer bound to a palindromic 18 bp RNA duplex that was published simultaneously (Chao et al., 2005). In this protein-RNA complex, the \( \alpha \)2/\( \alpha \)2’ helices of the B2 dimer are indeed oriented almost parallel to the helical axis of the A-form RNA (the angle between the
B2 dimer axis and the axis of the dsRNA is ~10°) and B2-RNA interaction is made up almost entirely of contacts with the phosphate backbone and ribose sugars (Figure 3.4B).

As suggested by the solution structure of B2, K47 and R54 of both α2 and α2’ helices contact the phosphate backbone in the major groove, whereas R36 and K62 contact the phosphates over the minor groove. Both the monomer and the dimer folds of the B2 dimer in complex with dsRNA are very similar to the solution structure of the free B2 protein (backbone coordinate r.m.s.d. 1.2/1.3 Å for monomer/dimer). A slightly different orientation of the α3 helices is potentially linked to electrostatic contacts of residues in the preceding linker with the RNA. In addition, the helix dipole moment of the small helix α3 is oriented towards the RNA and the partial positive charge could be also used to interact with the phosphate backbone.
The structure of the B2 dimer reveals a novel mode of sequence-independent dsRNA recognition, a feature that is shared with the double-stranded RNA binding domain (dsRBD, also called double-stranded RNA binding motif, dsRBM). The dsRBD comprises about 70-75 amino acids and adopts an \( \alpha \beta \beta \alpha \) fold in which the two \( \alpha \)-helices pack against one face of the 3-stranded \( \beta \)-sheet. It uses residues in both the \( \alpha \)-helical and \( \beta \)-strand loop regions for RNA binding (Figure 3.5, Ryter & Schultz, 1998; Stefl et al., 2005). The dsRBD it found in many RNA binding proteins with various functions, including pre-mRNA editing, RNA transport, and RNA-mediating silencing. In the dsRBD, the \( \alpha \)-helix 1, the N-terminal part of \( \alpha \)-helix 2, and the loop between \( \beta \)-strands \( \beta_1 \) and \( \beta_2 \) recognize the shape of the dsRNA stem. Like the B2 dimer, the dsRBD discriminates between dsRNA and dsDNA but has no sequence-specificity. Both the dsRBD and the B2 dimer recognize two minor grooves and the intervening major groove along one face of the dsRNA, but the recognition is performed by completely different protein architectures and backbone contacts (Figure 3.4 and Figure 3.5). Thus, B2 and the canonical dsRBD represent distinct solutions to the problem of sequence-independent dsRNA recognition.

Notably, the topology of the four-helical bundle formed by \( \alpha_1, \alpha_2, \alpha_1' \) and \( \alpha_2' \) in the B2 dimer resembles the small dimeric Repressor of primer (Rop) protein (Figure 3.6B, Banner et al., 1987), which was shown to be important for the regulation of ColE1 plasmid copy number in *Escherichia coli*. The Rop protein forms a four-helical bundle and binds an RNA-kissing loop structure. Several models of the Rop-RNA complex have been built on the basis of
mutagenesis and NMR chemical shift mapping (Predki et al., 1995). The predicted binding interface with kissing RNA hairpins comprises an anti-parallel pair of the first helix (α1) of each monomer, somewhat resembling the B2 dimeric RNA binding interface. But in contrast to B2, in which all four helices are almost parallel to each other, Rop forms a canonical left-handed four-helical bundle, as reflected in the rather large r.m.s.d. of the atomic backbone coordinates (3.7 Å) between the B2 and Rop dimers (Figure 3.6A,B).

Similarly, the N-terminal RNA binding domain of the Influenza A virus NS1 protein is a symmetric homodimer formed by two three-helix monomers (Figure 3.6C, Chien et al., 1997; Liu et al., 1997). It corresponds to the first 73 amino acids of the NS1 protein and possesses all of the dsRNA binding activity of the full length protein (Qian et al., 1995). Site-directed mutagenesis experiments indicated that the side chains of positively charged amino acids in the second α-helix (α2) are the only amino acid side chains that are required for the dsRNA binding activity of the intact dimeric protein (Figure 3.6C).

Despite these similarities, the structure of B2 reveals clearly a novel fold and an unprecedented way of interaction with dsRNA.
3.4 Viral suppressors of RNAi act at different levels in the RNAi pathway

The B2 protein of FHV was the first identified suppressor of RNA silencing of an animal virus (Li et al., 2002), after the presence of suppressors of RNAi was already established for plant viruses. The fact that FHV suppresses RNA silencing during replication in plant, insect and worm hosts suggested that B2 has evolved a host-independent evasion mechanism (Li et al., 2002; Lu et al., 2005). The solution structure of the FHV B2 protein dimer and the novel dsRNA binding mode of contacting the sugar phosphate backbone to enable sequence-independent recognition of dsRNA provide an explanation of the suppressor activity on a molecular level. These results are consistent with a suppression of RNAi by B2 at two different stages: first, it can bind the viral dsRNA that is generated during viral replication, thereby protecting it from being cleaved by Dicer. Secondly, in the case viral siRNAs are generated, they can be bound by B2 blocking their incorporation into RISC (Figure 3.7).

![Figure 3.7 Suppression of RNAi by the FHV B2 protein. B2 can act at two stages of the RNAi pathway: by binding viral dsRNA it can block the initiation step (Dicer cleavage), preventing the generation of siRNAs. Like p19, it can also bind to siRNAs and thereby taking them out of the pathway.](image)

This model is in-line with genetic analysis in *C. elegans* showing that B2 is active in both wild-type worms and worms that are deficient of the Argonaute protein Rde-1, which indicated that B2 might act upstream of the RISC-mediated step of RNAi (Lu et al., 2005). B2 has been demonstrated to bind siRNAs *in vitro*, but unlike p19 of tombusviruses, B2 can also bind to long dsRNA, even with much higher affinity. Interestingly, inhibition of Dicer cleavage of long dsRNA after addition of B2 was indeed demonstrated very recently *in vitro* (Chao et al., 2005; Lu et al., 2005). This is supported by recent studies on the B2 protein of
Nodamura virus (NoV, Sullivan & Ganem, 2005). The two B2 proteins share only little sequence similarity, but probably adopt similar structures (Johnson et al., 2001). Interestingly, almost all of the residues that contact the RNA are conserved between these proteins. Sullivan et al. showed that NoV B2 is able to inhibit processing of long dsRNA into siRNAs in virus infected mammalian cells (Sullivan & Ganem, 2005). Furthermore, cellular NoV B2 was found to be associated with small ~20 nt RNAs which derived from a longer precursor. In addition to the inhibition of Dicer \textit{in vivo}, the authors showed the binding of B2 to dsRNA and blocking of Dicer \textit{in vitro}.

In conclusion, the B2 protein represents the first example of a dual mode of suppression of RNAi: it is not only targeting siRNAs as it was demonstrated for p19, but also the long precursor RNA that becomes thereby protected from Dicer cleavage. The protection of the viral dsRNA might be not complete, and a certain fraction of the replication intermediates might be processed anyway. Then, the binding of the generated siRNAs by B2 provides a second mechanism to protect the viral RNA from being cleared from the cell. The structure of the B2 dimer and the characteristics of dsRNA binding described in chapter 2.4 of this thesis provide an explanation for that on a molecular level.

The list of identified suppressors is long (Table 1.1) and future studies on the molecular function of these proteins might reveal novel modes of RNAi inhibition that are unknown to date. Suppressors of RNA silencing may provide very useful tools to study RNA silencing mechanisms, e.g. to identify endogenous targets of siRNA or miRNAs.
4. Additional Publications

In addition to the publications containing primary data, I was involved in writing one perspective and one review during the time of my Ph.D. thesis:

Perspective


Review


For each of the manuscripts, a summary of the discussed topic is given, followed by the reprint of the publication.
4.1 RNAi: Finding the elusive endonuclease

Summary

As described in the introduction, the process of RNAi can be divided into two steps. Briefly, in the first step, dsRNA is cleaved by the RNase III-type enzyme Dicer containing complex into smaller RNAs, called siRNAs. These are transferred and incorporated into a second complex, which contains an Argonaute protein as key player and name giving component. This complex is also called the effector complex, as it targets an mRNA that is complementary to the siRNA for degradation. The endonucleolytic activity of this Argonaute-associated complex was termed Slicer, according to the activity that cleaves RNA in the first step, Dicer. The molecular identity of Slicer remained unknown despite intensive studies until the middle of the year 2004.

Then, Song et al. solved the crystal structure of an Argonaute protein of the archaeabakterium *Pyrococcus furiosus* (Song et al., 2004). The key result of their study is that the Piwi domain adopts an RNase H fold. This class of enzymes is known to cleave the RNA strand of an RNA:DNA hybrid. Based on this structure, Liu et al. did mutations in the putative active site of human Argonaute proteins and showed that the Argonaute2-associated complex is not longer capable of cleaving a target mRNA (Liu et al., 2004). Additonal evidence for the hypothesis that the human Argonaute2 protein itself is Slicer came from results of Tuschl and coworkers (Meister et al., 2004). They showed that of different Argonaute-associated complexes purified from human cells, only the Ago2 complex was cleavage competent.

This perspective summarizes the results of these three papers which led to the identification of Ago2 as the molecular identity of Slicer.
4.2 Novel modes of protein-RNA recognition in the RNAi pathway

Summary

Over the last couple of years, an increasing number of structures of proteins or protein domains involved in RNAi were determined. These structures include PAZ domains, either in their free form or in complex with nucleic acid ligands, the structure of the Argonaute protein of the archaea *Pyrococcus furiosus*, structures of RNase III-type enzymes and crystal structures of the p19 protein, a viral suppressor of RNA silencing, in complex with an siRNA. In this review, the novel structures are described and the modes of interaction of the proteins with RNA are discussed.
5. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aa</td>
<td><em>Aquifex aeolicus</em></td>
</tr>
<tr>
<td>Af</td>
<td><em>Archaeoglobus fulgidus</em></td>
</tr>
<tr>
<td>At</td>
<td><em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosin triphosphate</td>
</tr>
<tr>
<td>BMRB</td>
<td>BioMagResBank</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BYV</td>
<td>Beet Yellow virus</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CMV</td>
<td>Cucumber Mosaic virus</td>
</tr>
<tr>
<td>CSA</td>
<td>Chemical shift anisotropy</td>
</tr>
<tr>
<td>DGCR8</td>
<td>DiGeorge syndrome critical region gene 8</td>
</tr>
<tr>
<td>Dm</td>
<td><em>Drosophila melanogaster</em></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ds</td>
<td>Double-stranded</td>
</tr>
<tr>
<td>DUF</td>
<td>Domain of unknown function</td>
</tr>
<tr>
<td>Ec</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EMS</td>
<td>Ethyl methanesulfonate</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electro mobility shift assay</td>
</tr>
<tr>
<td>FHV</td>
<td>Flock House virus</td>
</tr>
<tr>
<td>FID</td>
<td>Free induction decay</td>
</tr>
<tr>
<td>FMRP</td>
<td>Fragile X mental retardation protein</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transformation</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosin triphosphate</td>
</tr>
<tr>
<td>Hs</td>
<td><em>Homo sapiens</em></td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilo Dalton</td>
</tr>
<tr>
<td>miRNA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear overhauser effect / enhancement</td>
</tr>
<tr>
<td>NOESY</td>
<td>NOE spectroscopy</td>
</tr>
<tr>
<td>NoV</td>
<td>Nodamura virus</td>
</tr>
<tr>
<td>NPC</td>
<td>Nuclear pore complex</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>OB</td>
<td>Oligosaccharide/oligonucleotide</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacryl-amid gel electrophoresis</td>
</tr>
<tr>
<td>PAZ</td>
<td>Piwi, Argonaute, Zwille</td>
</tr>
<tr>
<td>Pf</td>
<td><em>Pyrococcus furiosus</em></td>
</tr>
<tr>
<td>PKR</td>
<td>RNA-dependent protein kinase</td>
</tr>
<tr>
<td>RBD</td>
<td>RNA binding domain</td>
</tr>
<tr>
<td>RBM</td>
<td>RNA binding motif</td>
</tr>
<tr>
<td>RDC</td>
<td>Residual dipolar coupling</td>
</tr>
<tr>
<td>RdDM</td>
<td>RNA-directed DNA methylation</td>
</tr>
<tr>
<td>RdRP</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA-induced silencing complex</td>
</tr>
<tr>
<td>RITS</td>
<td>RNA-induced initiation of transcriptional silencing</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonulease</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>Rop</td>
<td>Repressor of primer</td>
</tr>
<tr>
<td>S2</td>
<td><em>Drosophila</em> Schneider cell line 2</td>
</tr>
<tr>
<td>SAR</td>
<td>Structure activity relationship</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>Sm</td>
<td>Stephanie Smith</td>
</tr>
<tr>
<td>SN</td>
<td><em>Staphylococcus</em> nuclease</td>
</tr>
<tr>
<td>Sp</td>
<td><em>Schizosaccharomyces pombe</em></td>
</tr>
<tr>
<td>ss</td>
<td>Single-stranded</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch virus</td>
</tr>
<tr>
<td>TGS</td>
<td>Transcriptional gene silencing</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total correlation spectroscopy</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-violet</td>
</tr>
<tr>
<td>VIG</td>
<td>Vasa intronic gene</td>
</tr>
<tr>
<td>VIGS</td>
<td>Virus-induced gene silencing</td>
</tr>
</tbody>
</table>
6. References


References


References


Schwar 2. Evidence that siRNAs function as guides, not complexes is a Mg2+-


References


7. Curriculum vitae

Personal details

Name               Andreas Lingel
Date of birth      February 8th, 1976
Place of birth     Marbach am Neckar, Germany
Marital status     Single
Nationality        German

Education

Since 05/2002  European Molecular Biology Laboratory (EMBL) Heidelberg, Germany and
               Swiss Federal Institute of Technology (ETH) Zurich, Switzerland
               International Ph.D. Programme at EMBL Heidelberg
               Ph.D. research project under the supervision of Dr. Elisa Izaurralde: “Structural
               and functional characterization of proteins involved in RNA-mediated gene
               silencing”
               Supervisor at ETH Zurich: Prof. Dr. Ulrike Kutay

11/2001-02/2002  Rockefeller University, New York City, New York, USA
               Laboratory of Cell Biology, Prof. Dr. Günter Blobel
               Research internship, topic: “The Role of POM121 and gp210 on Nuclear Pore
               Complex (NPC) Structure and Assembly”

10/1999-11/2001  ETH Zurich, Switzerland
               Diploma in biology (biochemistry), passed with distinction, grade: 5.93

10/2000-04/2001  University of Zurich, Switzerland
               Department of Biochemistry, laboratory of Prof. Dr. Andreas Plückthun
               Cyclization”, grade: 6, with honors

04/2000-07/2000  ETH Zurich, Switzerland
               Institute of Biochemistry, laboratory of Prof. Dr. Ari Helenius
               Semester work, topic: “Quality Control of Newly Synthesized Proteins in the
               Endoplasmatic Reticulum”

04/1997-09/1999  Eberhard-Karls-University Tübingen, Germany
               Vordiplom in biochemistry, passed with distinction, grade: “sehr gut”

01/1996-01/1997  German Red Cross, District Ludwigsburg, Germany
               Civil service as an ambulance man in the emergency service

06/1995  Friedrich-Schiller-Gymnasium in Marbach am Neckar, Germany
          Abitur, grade: 1.4, passed with distinction
Publications


Conferences

1. EMBL Heidelberg, Germany
   Poster presentation: “Nucleic acid 3’-end recognition by the Argonaute2 PAZ domain”

2. Keystone Resort, Colorado, USA
   Keystone Symposium on “siRNAs and miRNAs”, 14.-19. April 2004
   Poster presentation: “Structural Basis for 3’-end recognition of nucleic acids by the *Drosophila* Argonaute2 PAZ domain”

3. Max Planck Institut for Biophysical Chemistry, Göttingen, Germany
   Poster presentation: “The Argonaute 2 PAZ domain adopts a novel nucleic acid binding fold”

4. Cold Spring Harbor Laboratory, New York, USA
   Oral presentation: “Structure and nucleic acid binding of the *Drosophila* Argonaute 2 PAZ domain”