TECHNISCHE UNIVERSITÄT MÜNCHEN
DEPARTEMENT CHEMIE
LEHRSTUHL FÜR BIOCHEMIE

Structural Characterization of the Argonaute protein

from *Methanocaldococcus jannaschii*

Christine A. Oellig

Table of Contents

1. Summary

RNA interference (RNAi) is routinely employed in biological research to effectively silence specific target genes and is being intensively investigated as a therapeutic weapon against various human diseases. In eukaryotes, Argonaute (Ago) proteins play an essential role in all RNAi pathways in which they are guided by small RNAs to complementary target RNAs, thus resulting in RNA degradation or translational repression. Ago proteins are also present in prokaryotic organisms (termed pAgos) where they are most likely crucial components of a novel class of defence system that utilize guide DNA or RNA molecules to destroy invading viruses or other foreign nucleic acid. Interestingly, scientists implied the idea that specific pAgos could be employed as a precise and powerful tool for genome editing in human cells.

The Ago protein of the archaeal organism *Methanocaldococcus jannaschii* (MjAgo) is a DNA-guided endonuclease, able to cleave complementary target DNAs or double stranded DNAs. This thesis presents the crystal structures of MjAgo in its free state and in complex with a bound 21mer guide DNA strand (binary complex), thus providing a snap-shot on guide DNA recognition.

The apo enzyme adopts the typical conserved fold of Ago proteins consisting of four distinct domains: the N-terminal (N), PIWI Argonaute and Zwille (PAZ), middle (Mid) and P-element-induced wimpy testis (PIWI) domains. Despite low sequence similarity, the overall architecture of MjAgo is comparable with the archaeal homologue of PfAgo.

In the binary complex, the 5′-phosphorylated end is anchored in the nucleotide binding pocket in the Mid domain, whereas the 3′-end of the guide is inserted into the PAZ domain. The structural architecture of the PAZ domain which is further defined by an α-helix implied the idea that MjAgo is less accessible for a guide strand carrying a purine at the 3′-end. This observation could be confirmed by activity assays, showing that MjAgo has a unique preference for a pyrimidine at the 3′-end of the guide DNA.

Due to the fact that it is possible to compare the apo enzyme with the complex bound form, new insights into the DNA-loading mechanism of pAgos could be obtained. Here, substrate interaction provokes an opening of a binding channel, which is accompanied by a movement of the PAZ domain. In addition, a conserved structural element, known as 'helix 7' undergoes

a structural transition which has not been observed in other Ago variants before. Furthermore, our structural data of the binary complex revealed that MjAgo features an additional DNA binding channel, which has been proven to be essential for the formation of an active competent ternary complex.

Taken together, the structural characterization of MjAgo in both its apo and complex bound forms provides new insights into the DNA-mediated binding mechanism of pAgos.

2. Zusammenfassung

Der Mechanismus der RNA-Interferenz (RNAi) wird in den Biowissenschaften regelmäßig angewendet um zielgerichtet bestimmte Gene auszuschalten und wird intensiv erforscht, um als therapeutische Methode gegen unterschiedlichste menschliche Krankheiten eingesetzt zu werden. In Eukaryoten spielen Argonauten (Ago) Proteine bei allen vorkommenden RNAi-Signalwegen eine essentielle Rolle. Hierbei werden sie durch kleine RNAs zu komplementären *target* RNAs gesteuert, was schließlich die Spaltung der RNA oder die Inhibition der Translation zur Folge hat. Ago Proteine sind auch in prokaryotischen Organismen vertreten (pAgos), in denen sie sehr wahrscheinlich ein essentieller Bestandteil eines neuen Abwehrmechanismus sind, welcher DNA oder RNA Moleküle einsetzt um aggressive Viren oder Fremd-DNA abzuwehren. Interessanterweise vermuten einige Wissenschaftler, dass bestimmte pAgos als präzise und effektive Werkzeuge zur Veränderung menschlicher DNA eingesetzt werden können.

Das Ago Protein des Archaebakterium *Methanocaldococcus jannaschii* (MjAgo) ist eine DNA-spezifische Endonuklease, welche fähig ist, komplementäre *target* DNAs oder doppelsträngige DNAs zu spalten. Die vorliegende Arbeit beschreibt die Kristallstrukturen des MjAgo Proteins sowie im Komplex mit einer 21 Nukleotide langen *guide* DNA (Binary-Komplex), und zeigt somit eine Momentaufnahme der Erkennung der *guide* DNA.

Das Apoenzym nimmt die typische konservierte Proteinstruktur der Ago-Proteine an, welche aus vier verschiedenen Domänen besteht: der N-terminalen (N), der PIWI-Argonaute-Zwille (PAZ), der Mittleren (Mid) sowie der P-element-induced wimpy testis (PIWI)–Domäne. Trotz geringer Sequenzhomologie ist die gesamte Anordnung von MjAgo mit der des homologen archaealen PfAgo vergleichbar.

Im Binary-Komplex ist das 5′-phosphorylierte Ende in der Nukleotid-Bindetasche der Mid-Domäne verankert während das 3′-Ende des *guide* Stranges in der PAZ-Domäne gebunden ist. Der strukturelle Aufbau der PAZ-Domäne, welche zusätzlich durch eine Alpha-Helix verkleinert wird, lies vermuten dass MjAgo weniger zugänglich für *guide* Stränge ist, die eine Purin-Base am 3′-Ende aufweisen. Dies konnte durch Aktivitätstest bestätigt werden, welche zeigten, dass MjAgo eine einzigartige Präferenz für Pyrimidine am 3′-Ende der *guide* DNA

3

hat.

Durch die Möglichkeit die Struktur des Apoenzyms mit der des im Komplex gebundenen Proteins zu vergleichen konnten neue Erkenntnisse in den DNA-Bindemechanismus von pAgo Proteinen gewonnen werden. Hierbei führt die Wechselwirkung mit dem Substrat zur Öffnung eines Bindekanals was gleichzeitig mit einer Rotationsbewegung der PAZ-Domäne verbunden ist. Zusätzlich durchläuft ein konserviertes Struktur-Element – bekannt als Helix 7 – eine Strukturänderung, die bei anderen Ago-Proteinen bisher nicht beobachtet werden konnte. Des Weiteren zeigte die Analyse der Strukturdaten des Binary-Komplexes, dass MjAgo über einen zusätzlichen DNA-Bindekanal verfügt, welcher nachweislich für die Bildung eines aktiven Ternary-Komplexes äußerst wichtig ist.

Zusammenfassend liefert die strukturelle Analyse des Apoenzyms sowie der Komplexgebundenen Form von MjAgo neue Erkenntnisse in den DNA-Bindemechanismus von pAgo Proteinen.

3. Introduction

3.1. RNA Interference

RNA interference is a eukaryotic, sequence-specific gene-silencing pathway, in which small non-coding RNAs guide destruction and destabilization of complementary target RNAs.

The inhibitory function of RNAs was first discovered in the early 1990s when Napoli and Jorgensen applied additional gene-copies, encoding a key enzyme responsible for deep, violet pigmentation into petunia plants [1]. Instead of the enhanced expression of the target gene, which would have led to a darker pigment synthesis, they observed white flowers resulting of endogenous as well as introduced gene suppression. Continuous investigation of this phenomenon by plant scientist revealed, that the down-regulation and suppression of gene expression was a consequence of increased levels of mRNA degradation in the cytoplasm and therefore referred to as „post transcriptional gene silencing" [2, 3].

In 1998, Fire and Mello finally achieved the major breakthrough by discovering that actually long double-stranded RNA (dsRNA) was the trigger of this type of sequence-specific gene silencing in *C. elegans* [4]. Consequently, they termed it RNA interference (RNAi) and won the Nobel Prize in 2006 for Physiology or Medicine for their work. Further biochemical and biophysical research in plant and invertebrates revealed, that an endoribonuclease enzyme called Dicer is actually trimming the long dsRNAs in short oligonucleotides, typical 21-25 nucleotides (nt) in length [5-10]. These small fragments, referred to as small interfering RNAs (siRNA) are later incorporated into a protein from the Argonaute (Ago) protein family, building up the molecular apparatus of RNAi, the so called RNA-induced silencing complex (RISC) [10-13]. Ago proteins are considered to be the key catalytic component within the RISC complex and are guided directly by small RNAs to the corresponding target mRNA, thus resulting in mRNA degradation and/or translational repression.

RNAi is widely distributed in eukaryotes and plays a major role in various processes such as antiviral defence, gene surveillance and gene regulation [14, 15]. In the last decade, RNAi has developed into an essential research tool to repress genes specifically and it is applied in many fields, ranging from experimental research to medical therapy [16, 17].

3.1.1.   Cellular mechanism of gene silencing in eukaryotes

Small RNAs (sRNAs) are small non-coding regulatory RNAs which guide the RNA induced silencing complex to complementary target mRNAs, thus provoking its translational repression or destruction [18, 19]. Based on their origins and structures, they are subdivided in three main classes of small interfering RNAs (siRNAs), microRNAs (miRNAs) and P-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs) [20]. Albeit their various biological roles and different biogenesis pathways, they all have to be incorporated within a member of the Ago protein family in order to fulfil their function.

Although several types of RNAs have been shown to knock down gene expression, two types of small RNA molecules - siRNAs and miRNAs - are considered to be critical to RNA interference. They both have in common that they are processed by the ribonuclease (RNase) III enzyme Dicer out of long dsRNA precursors, resulting in 21- to 25 nt mature dsRNAs, competent for RISC loading. In contrast to siRNAs and miRNAs, piRNAs are processed in a Dicer-independent fashion from single stranded RNA precursors. These piRNAs are primarily found in animals where their expression is restricted to the germline [21-23]. They associate with a member of the PIWI subfamily of Ago proteins and maintain genome integrity by controlling mobile genetic elements [24, 25].

3.1.1.1.   Biogenesis of miRNAs and siRNAs

Small interfering RNAs are cleavage products of long, fully complementary dsRNAs. These dsRNAs can be either transcribed endogenously (endo-siRNAs) as e.g. natural antisense transcripts or introduced into cells by viral infection (exo-siRNAs) or transfection [26-28]. After entering the cytoplasm, the long dsRNA is cleaved by Dicer resulting in RNA duplexes, with characteristic 2 nt overhangs at the $3'$-end and a phosphate group at the $5'$-end [29-31]. Although Ago proteins can load single stranded RNAs [32], binding of RNA duplexes requires the association of Dicer and a dsRNA binding cofactor (TRBP in humans), building up the so called RISC loading complex [33].

After unwinding of the duplex, the strand that is less tightly paired at its $5'$-end (termed the guide strand) remains incorporated within the Ago protein [34], whereas the other strand,

termed passenger strand gets degraded.

Mature miRNAs quite resemble siRNAs and comprise various similarities in their processing pathways. They are highly conserved, endogenously encoded RNAs and can be can be found in many organisms such as plants, animals and some viruses [20].

The production of miRNAs is starting in the nucleus as large hairpin loop transcripts (termed primary transcripts), which are further processed by the RNase III endonuclease Drosha into ~70-nt long precursor duplexes (pre-miRNA) [35]. These duplexes are exported from the nucleus by the shuttler Exportin-5 [36-38] and afterwards cleaved by Dicer to form a siRNA-like imperfect duplex termed mature miRNA duplex [5]. These miRNAs duplexes are about 22 nt in length [39], containing the mature miRNA and the corresponding duplex counterpart which is called "miRNA star" (miRNA*) [5].

As with siRNAs, the miRNA duplex is short lived. The mature miRNA strand, which is less thermodynamically stable remains associated within an Ago protein and mediates target recognition within the so called microRNA-induced silencing complex (miRISC). In contrast to the siRNA unwinding mechanism, the miRNA* gets in most cases released and not degraded. This might be a consequence of the mismatches and bulgs, which are present in miRNA duplexes, blocking the catalytic activity of Ago proteins [40].

3.1.1.2.   The RISC effector complex

RISCs are crucial for RNA interference and comprise a heterogeneous family of distinct gene-silencing complexes. Common to all is a member of the Ago protein family, which binds the small regulatory RNA. Once incorporated, the small regulatory RNA serves as a template, mediating RISC through Watson-Crick base pairing to it′s corresponding target RNA transcript. Though the complete mechanism is at present poorly understood, binding and unwinding of an RNA duplex requires additional RISC-loading cofactors [41] such as Dicer and a dsRNA-binding protein (known as TRBP in humans) [33].

However, the minimal RISC complex, able to interact and cleave the complementary target RNA was demonstrated to be simply an Ago protein bound to a small single-stranded RNA [32]. Interestingly, kinetic analysis have demonstrated that once incorporation within a RISC complex, the same guide strand can find and anneal 10 times faster with its target RNA than

in would anneal in free solution [42].

SiRNAs characteristically base pair perfectly with their corresponding mRNA, thus promoting Ago-catalysed cleavage of the target strand, termed as "slicing" [43, 44]. This cleavage activity is very precise, resulting in a single nucleolytic cut of the mRNA [43], 10 to 11 nt downstream of the first base pair between the formed siRNA-mRNA duplex [45]. After cleavage, the target strand gets ejected from the complex, enabling RISC to find additional target molecules. Surprisingly, mismatches introduced at or close to the centre of the formed duplex abolish slicer endonucleolytic activity of Ago [46].

In contrast, most animal miRNA interact through mismatches and bulges and match only perfect complementary with their target along nt 2-7/8 (from the 5′-end) referred to as the "seed sequence" [45], which is essential for their function [47, 48]. As a consequence, miRISCs remain associate with their target strands, resulting in silencing by translational repression and/or target deadenylation [49].
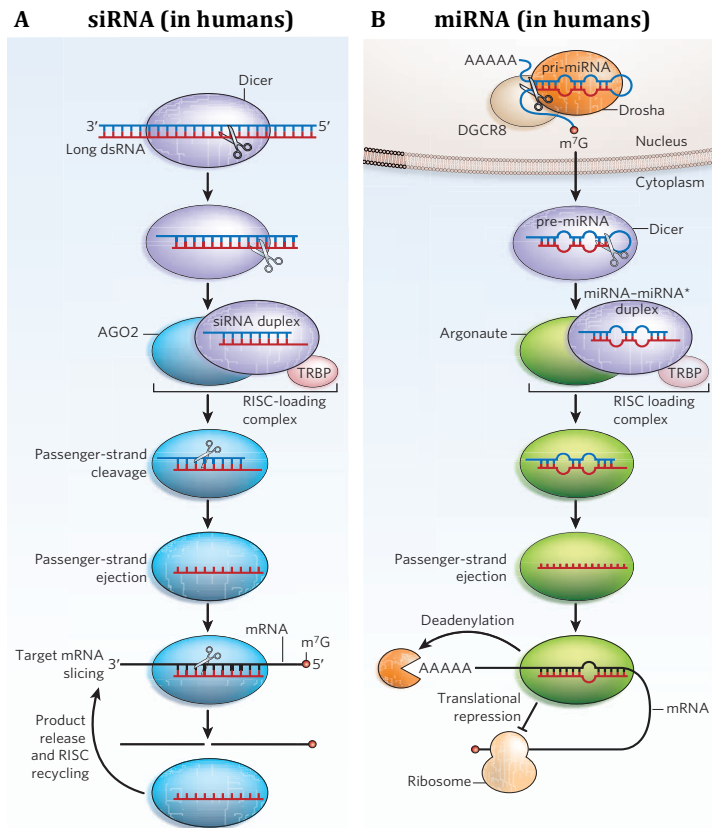
Figure 1: Synthesis and mechanism of function of the two main classes of small regulatory RNA in humans. A) Biogenesis pathway from siRNAs. Long dsRNAs are cleaved by Dicer and loaded onto an Ago protein within the RISC-loading complex. After unwinding of the siRNA duplex, the guide strand is selected and mediates the Ago protein to a complementary target mRNA. After binding, the Ago protein catalyses the cleavage of the target strands. **B)** Biogenesis pathway from miRNAs. MiRNAs are transcribed endogenously in the nucleus as primary-transcripts. After processing by Drosha, the pre-miRNAs are exported into the cytoplasm where they are further cleaved by Dicer, resulting in imperfectly paired miRNA-miRNA* duplexes. These duplexes are loaded onto an Ago protein within the RISC-loading complex. As with siRNAs, the duplex is unwound and the guide interacts through mismatches and bulges with the corresponding target mRNA. As a result, the Ago protein remains associated with the target strand and induces deadeylation or translational repression. (Figure reprinted with permission from M. Jinek, J. A. Doudna, *Nature* **2009**, 457, 405-412)

3.1.2.   DNA interference in prokaryotes

RNA interference is a common mechanism in eukaryotes, which is responsible for several of functions, including antiviral defence and gene regulation [14, 15]. All RNA interference pathways are mediated by small single-stranded RNA molecules, which lead the Ago protein to its corresponding target RNA. Although Ago homologs are also present in 12% of all prokaryotic organisms (termed pAgo) [31], their role has still remained a mystery. However,

the deep structural conservation of Ago proteins throughout nature implies the idea that these proteins operate in each organism basically the same way. Surprisingly, no other recognizable RNAi-related proteins (such as Drosha or Dicer) could be identified in prokaryotic organisms [30, 51, 52]. Based on the discovery that pAgo encoding genes frequently cluster with genes involved in host defence [53], several scientists implicated their involvement as key components of a novel class of defence system. Bioinformatic and expression analysis showed that the Ago from *Marinitoga piezophila* (MpAgo) is encoded in close proximity of a clustered regularly interspaced short palindromic repeats (CRISPR) gene loci and coexpressed with their gene products [53]. CRISPR-associated proteins use short segments of DNA (which are retrieved from a previous phage or plasmid attack and integrated in the genome) as a guide to destroy invasive mobile genetic elements, thus in a manner similar to RNAi. Furthermore, studies on recombinant Ago from the archaeal organism *Methanocaldococcus jannaschii* (MjAgo) as well as on other pAgos variants revealed, that these proteins bind single stranded DNAs (ssDNA) with high affinity and are able to cleave complementary target DNAs [54, 55]. Additionally, the Ago proteins of *Thermus thermophilus* (TtAgo), *Natronobacterium gregoryi* (NgAgo) and *Pyrococcus furiosus* (PfAgo) are capable to linearize target plasmids, thus able to protect the host against invasive genomes [54, 56, 57]. Consequently, the fundamental difference between pAgos and eukaryotic Agos (eAgos) lies in the capacity of pAgos to use either DNA or RNA guides to silence complementary target DNA. This preference for guide and target binding might be explained by the surface charge distribution of the guide-target duplex binding channel and the binding pocket of the Mid domain. Here, eAgos are strongly positive charged whereas in TtAgo and PfAgo the duplex binding channels are much less positive charged and their binding pockets are hydrophobic [54]. Taken together, these findings demonstrate that the role of pAgos lies most likely in a DNA-guided DNA-interfering defence system, able to protect the host against invasive plasmids, transposons or other mobile genetic elements.

3.2. The Argonaute protein family

Ago proteins are highly specialized RNA/DNA-binding proteins and are regarded as the key enzyme in all RNA-mediated silencing pathways [58]. They are represent throughout nature

and received their name from a knockout phenotype in *Arabidopsis thaliana* where the leaves resembled and curled up like squid tentacles of the octopus *Argonaute argo* [59]. Many species encode multiple family members, ranging from eight in humans, five in mice, ten in *Arabidopsis thaliana,* five in *Drosophila* to twenty-seven in *C. elegans* [60] all with different expression profiles and substrate specificity [61, 62]. On the basis of sequence homology the Ago protein family can be subdivided in two subfamilies [58]; the AGO subfamily and the PIWI subfamily [63, 64]. The expression of the PIWI subfamily is mainly restricted to the germ line in which they interact with piRNAs in order to control mobile genetic elements [58]. In contrast, the AGO subfamily is ubiquitously expressed in many organisms where they interact with either DNA or miRNAs/siRNAs.

Humans express eight Ago proteins whereas four belong to the AGO clade, namely Ago1, Ago2, Ago3 and Ago4. The functional difference between the four proteins is still unknown, however only Ago2 is able to cleave complementary target RNA [65, 66].

3.2.1.   Overall domain architecture

Ago proteins have a typical molecular weight of ~100 kDa and are composed of four predominant domains: N-terminal (N), PIWI Argonaute and Zwille (PAZ), middle (Mid) and PIWI.

The first reported full-length structure of the archaeal organism *P. furiosus* (PfAgo) [67] revealed a bi-lobal architecture where the PAZ lobe, comprising the N, linker 1 (L1) and PAZ domains is connected by an external linker 2 (L2) to the PIWI lobe, build up by the Mid and PIWI domains [29]. Crystals of TtAgo in complex with a bound DNA (binary complex) [68] revealed that the guide strand is located within a basic channel between the two lobes, thereby contacting all domains and linkers. The 5′-phosphorylated end is anchored within a specific binding pocket in the Mid domain, whereas the 3′-end of the guide is positioned in the PAZ domain [68, 69] **(Figure 2)**.

Over the past years, structural insights of the eukaryotic counterparts could be obtained by crystallization of human Ago2 (hAgo2) as well as the Ago of the budding yeast *Kluyveromyces polysporus* (KpAgo), both in complex with small guide and target RNAs (ternary complex) [70-72]. Despite low sequence similarity (12%) the global structure of

prokaryotic and eukaryotic Agos is exceptionally well conserved, comprising a bi-lobal architecture with a central cleft for binding the guide and target molecules [58]. However, notable structural differences could be observed in the relative orientation of the core domains as well as in extended loops and additional insertion elements, likely to be involved in RNA binding, target recognition or potential recruitment of other proteins [58, 70, 71].

3.2.1.1.   Structure and function of individual domains

The N-domain shows the highest structural variation between all characterized Ago variants. However, domain swapping from the active human Ago2 to Ago3 [73] turned the inactive protein into an active slicer enzyme, indicating a critical, yet unknown role of the N domain in catalytic activity as well as in the release of the cleaved target strands [74].

The PAZ domain, which is also typical for Dicer enzymes, resembles an SH3-like barrel or modified oligosaccharide/oligonucleotide-binding (OB) fold [58, 67]. A hydrophobic pocket in between, lined up with conserved aromatic and basic residues recognize the typical 2-nt overhangs of the guide strand, thus anchoring the 3′-end in the PAZ domain [68, 70, 71]. These interactions, which protect the guide from degradation [75] occur predominately with the final phosphate and sugar of the backbone, thus in a non-specific manner [72, 76]. Additionally, structural kinetic and single-molecule studies indicate this domain as highly flexile, even in the absence of a bound guide strand [68].

The PIWI domain adopts an RNase H-like fold, thereby providing endonuclease activity to the protein [67, 77, 78]. The catalytic core comprises a DDX triad (where X is either histidine or aspartic acid; in fewer cases its asparagine [79]) with two coordinated divalent metal ions, essential for slicing activity [80]. A glutamate residue, sitting on top on a flexible PIWI loop (the glutamate finger) can be plugged in, completing the catalytic pocket, forming an active DEDX tetrad [71, 81]. However, studies on human Ago3 revealed, that the presence of the catalytic tetrad is essential but insufficient for providing slicer activity to the protein [43, 82].

The Mid domain resembles a typical Rossmann fold, composed of a four-stranded beta-sheet core flanked by four alpha-helices and two additional short alpha-helices [83]. Highly conserved residues within a basic nucleotide-binding pocket interact with the final phosphate group at the 5′-end, thus anchoring the guide strand at the junction of the Mid and PIWI

domains [83-85]. Interistingly, this domain can, using a structural feature termed the 'nucleotide specificity loop' recognize specific 5′-end bases, thus conferring guide specificity to hAgo2 [83, 86]. A magnesium ion, found in the binding pocket of pAgos is essential for binding the two phosphates of the guide (nucleotides 1 and 3) [68, 87]. In contrast, the ammonium group of a conserved lysine achieves charge neutralization in eAgos. [74, 88].

In addition, the characteristic binding pocket of the Mid domain in TtAgo appears to be more hydrophobic compared to that of the eukaryotic counterparts [74, 54]. This discovery might be one explanation for the preference of guide DNAs over RNAs in pAgos.
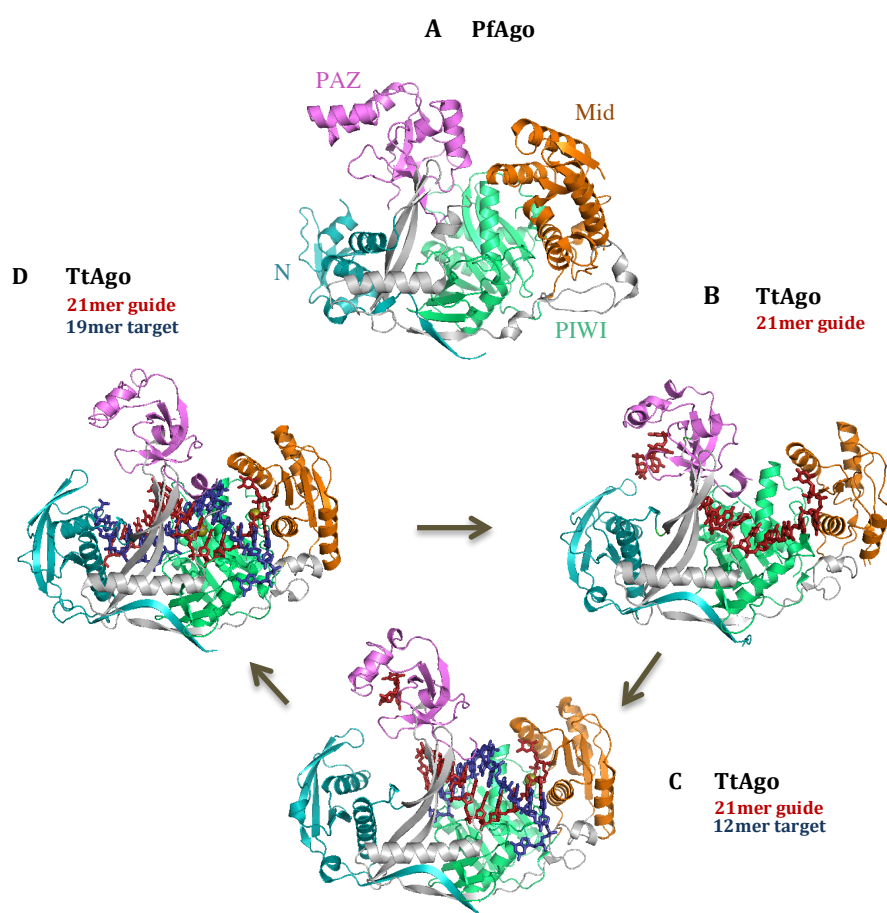
Figure 2: Structural features of the cleavage mechanism of Ago proteins. A) Crystal structure of apo-PfAgo (PDB code 1Z25) shown in ribbon representation with the N domain in teal, the PAZ domain in violet, the Mid domain in orange, the PIWI domain in green and the linkers in gray. **B)** Binary complex of TtAgo bound to a 21-base DNA guide strand (PDB code 3DLH). The DNA (shown in stick representation) is coloured in red and anchored within the binding pockets of the PAZ and Mid domains. **C)** Ternary complex of TtAgo (PDB code 4N47). The 21-mer guide DNA strand (shown in red) is still anchored in the the respective binding pockets and pairs along the seed sequence with the 12-mer target DNA (shown in blue). **D)** Cleavage-compatible architecture of TtAgo (PDB code 4NCB) in complex with a 21 mer guide and 19 mer traget DNA. Due to intensive duplex pairing, the 3′-end of the guide strand gets released from the PAZ domain, that is accompanied by the insertion of the glutamete finger within the catalytic pocket to complete the catalytic tetrad. The Mg^{2+} ions located in the active site are shown in olive.

3.2.2. Argonautes proteins in complex with guide and target molecules

The crystal structure of TtAgo in complex with a 5′-phosphorylated 21-base DNA guide strand [69] revealed that the guide strand is incorporated between a specific binding channel, mediated by all distinct domains of the protein **(Figure 2)**. Interestingly, the majority of the contact points are predominately between basic side chains and the sugar-phosphate backbone of the guide strand, explaining why Agos can be programmed to target nearly any gene for silencing.

As described earlier, the guide strand is anchored at both its ends within the protein, with the 5′-phosphate inserted in the Mid domain and nt 20 and 21 at the 3′-end in the PAZ domain [29] **(Figure 2)**. Insertion of the 5′-end within the Mid pocket strongly bends the guide strand, resulting in a helical arrangement of nt 2-10 with exposed Watson-Crick edges of the seed segment (nt 2-7/8) [29, 84, 87], thus ready for base pairing with the target molecule. These structural insights explain the enhancement in target recognition by the guide strand when incorporated within an Ago protein [41] and why the seed region is so crucial for target pairing. Additionally, the binary structure shows an unanticipated orthogonal alignment within the helix at bases 10 and 11, caused by an insertion of a conserved residue into the guide strand [28].

Further structural investigation of the ternary complex of TtAgo with a 19-mer target RNA [70] shows that the guide DNA-target RNA duplex pairs in an A helical conformation (ranging from nt 2-15), placing bases 10 and 11 stacked and centred on the cleavage site of the PIWI domain **(Figure 2)** [70, 88]. In order to facilitate insertion, alignment and pairing of the formed duplex, pronounced conformational changes, mainly of the N-terminal and PAZ domain within the protein, are essential [55]. Additionally, due to extensive duplex pairing, the 3′-end of the guide gets finally released out of the binding pocket in the PAZ domain [70]. Higher resolution structures of the ternary complex of TtAgo revealed [81], that extensive duplex pairing together with the PAZ release is the actual trigger that leads to the insertion of the glutamic finger [71] into the catalytic pocket, completing and activating the catalytic tetrad. Recent single-molecule fluorescence resonance energy transfer studies on MjAgo [55] confirmed this two-state model of Ago function. However, this unplugged to plugged

transition is unique to pAgos. All to date characterized structures of active eAgos comprise the complete catalytic tetrad, irrespective of the PAZ release [70, 71].

3.2.3. The Argonaute protein from *M. jannaschii*

M. jannaschii is a hyperthermophillic methanogen organism that belongs to the kingdom of archaea. It was first isolated from the East Pacific Rise close to white smokers [89] and is specialized to live in circumstances where water reaches boiling temperature (48-94 °C) or pressure is extremely high (up to 200 atm) [90]. It was the first archaeon which complete genome was sequenced in 1995, therebye identifying a total of 1738 predicted protein-coding genes [90].

A recently characterized member belongs to the Ago protein familiy, namely MjAgo which shares 28% sequence identity with PfAgo and 23% sequence identity with human Ago2 [55] **(Figure 3)**. Bioinformatic and biochemical analysis indicated that MjAgo is a catalytic active Ago variant, splicing exclusively target DNAs out of DNA/DNA hybrids [55].

The protein accepts 5'-phosphorylated ssDNA guide strands ranging from 18-40nt in length, resulting in multiple cleavge products of the corresponding target DNA. Additionally, MjAgo is not capable to cleave RNA target strands [55].

Further experiments revealed that MjAgo requires temperatures above 75 °C for efficient cleavage activity, demands bivalent cations and is inhibited by sequence-mismatches within the substrate [55].

4. Objective

RNA interference is today widely applied in biological research to effectively silence specific target genes [791-93] and is being intensively investigated as a therapeutic tool against pathogens and various human diseases [94, 95].

Ago proteins are highly specialized RNA/DNA-binding proteins and are regarded in eukaryotes as the key enzyme in all RNA interference pathways. In prokaryotes, Ago proteins are considered to be involved in a DNA-guided DNA-interfering defence system which protects the host against invasive genomes.

To date, many protein crystal structures from eukaryotic and prokaryotic organisms have provided precious insights into the structure-function relationship of these proteins and greatly contributed to a better understanding of the underlying mechanism of RNAi.

A new, recently characterized member of the Ago family represents the protein from the archaeal organism *M. jannaschii* (MjAgo). Biochemical research revealed that the most striking feature of this enzyme lies in its capacity of DNA cleavage out of double-stranded DNA (dsDNA), resulting in multiple splicing products [55]. However, no structural data of MjAgo or any other DNA-associated archaeal Ago that employs a DNA-guided DNA cleavage mechanism are currently available.

Therefore, the aim of this thesis was the structural characterization of the Ago protein from *M. jannashi* by X-ray crystallography, both in its apo and the substrate bound forms (binary and ternary complex). The received structural data should explain the above-mentioned key feature of the DNA-guided DNA cleavage mechanism, thereby providing a better understanding of pAgo action.

5. Results and Discussion

5.1. Structural similarity of MjAgo to other Ago variants

Crystallization experiments revealed that MjAgo adopts the typical conserved biolobal Ago fold, consisting of four functional domains and two linker domains. The N-terminal lobe comprises the N (residues 18-129) and PAZ domains (residues 163-258) connected by the linker L1 while the C-lobe is build up by the Mid (residues 346-488) and PIWI domains (residues 489-713) **(Figure 3)**. The lobes are joined by the linker L2, that forms the base of the central cleft **(Figure 7)**. Despite low sequence similarity (28 %), MjAgo is most similar to PfAgo **(Figure 3)**. However, MjAgo differs in the secondary structure composition and orientation of the PAZ domain **(Figure 8)**. Although no divalent metal ion could be found within the slicer site, highly conserved amino acids, located at structural equivalent position as active-site residues found in other Ago variants could be identified within the PIWI domain **(Figure 3)**.

```
                     β1      β2    η1      β3              η2    α1      β4      β5
MjAgo        →→→→→→→    →→→   ɷɷɷ  →→→→→→       ɷɷɷ ɷɷɷɷɷ.ɷɷɷ   →→→→    →→→→
                   1        10         20           30            40        50        60

MjAgo  ........MVLNKVTYKINAYKIKEEFIPKEVHFYRIKS.......FVNEAFNFYRF.VNFYGGMIINKKDKSFVLPY
PfAgo  ..........SMKAIVVINLVKINKKIIPDKIYVYRLFNDPE..EELQKEGYSIYRL.AYENVGIVIDPENLIIATTK
AaAgo  ..........MGKEALLNLYRIEYRPKDTTFTVFKPTH......EIQKEKLNKVRWRVFLQTGLPTFRREDEFWCAG
TtAgo  .......MNHLGKTEVFLNRFALRP.LNPEELRPWRLEVVLD..PPPGREEVYPLLAQVARRAGGVTVRMGDGLASWS
hAgo2  MYSGAGPALAPPAPPPPIQGYAFKPPPRPDFGTSGRTIKLQANFFEMDIPKIDIYHYELDIKPEKCPRRVNREIVEHM

                           β6    η3                α2           α3                  ▬
MjAgo            →→→→→  ɷɷɷ   ɷɷɷɷɷɷɷɷɷɷ....ɷɷɷɷ  ɷɷɷɷɷɷɷɷɷɷ
                  70        80        90       100       110       120       130

MjAgo  KVDNKVLKYKDGNNEIPIDIEYIKSLKLEYVKPEIAEKLVRG.......YLKSVHKIEPELSRIIKNIRKHKVVENIK
PfAgo  ELEYEGEFIPEGE............ISFSELRNDYQSKLVLR.......LLKENGIGEYELSKLLRKFRKPKTFGDYK
AaAgo  KVEKDTLYLTLSNG............EIVELK.RVGEEEFRG.......FQNERECQELFRDFLTKTKVKDKFISDFY
TtAgo  PPEVLVLEGTLARMG..........QTYAYRLYPKGRRPLDP.......KDPGERSVLSALARRLLQERLRRLEGVWV
hAgo2  VQHFKTQIFGDRKPVFDGRKNLYTAMPLPIGRDKVELEVTLPGEGKDRIFKVSIKWVSCVSLQALHDALSGRLPSVPF

                  β7       TT    β8        α4         α5          β9          β10
MjAgo     →→→→→→→→  TT   →→→→→→   ɷɷɷɷ   ɷ....ɷɷɷɷ      →→→→→       →→→→→
                 140 ★       150       160       170       180       190       200

MjAgo  VESYCEYEVKKHDGDYYLTLNFRHTASITKHLWDFVNRDK....ALLEEYVG..KKIIFKPNP...KVRYTISLVDA
PfAgo  VIPSVEMSVIKHDEDFYLVIHIIHQTQSMKTLWELVNKDP....KELEEFLMTHKENLMLKDIASPLKTVYKPCFEEY
AaAgo  KKFRDKITVQGKNRKIALIPEVNEKVLKSEEGYFLLHLDL....KFRIQPFETLQTLLERNDFNPKRIRVKPIGIDFV
TtAgo  EGLAVYRREHARGPGWRVLGGAVLDLWVSDSGAFLLEVDP....AYRILCEMSLEAWLAQGHPLPKRVRNAYDRRTWE
hAgo2  ETIQALDVVMRHLPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSVRPSLWKMMLNIDVSATAFYKAQPVIEFV

                  α6                α7                    β11             η4   β12
MjAgo     ɷɷɷɷɷɷɷɷɷɷɷɷɷ    ɷɷɷɷɷ.........ɷ        →→→→→          ɷɷɷ →→→→
                 210       220              230 ★        240              250

MjAgo  PNPQKIEEIMSHIIKYYKWSEDMVKST.........FGETDYNQPTM..YCEEILEP.........FAPQFCNLVFY
PfAgo  TKKPKLDHNQEIVKYWNYHIERYWNTPEAKLEFYRKFGQVDLKQPAI..LAKFASKIKKNKNYKIYLLPQLVVPTYN
AaAgo  GRVQDVFKAKEKGEEFFRLCXERSTHK.........SSKKAWEELL..KNRELREK..........AFLVVLEK
TtAgo  LLRLGEEDPKELPLPGGLSLLDYHASK.........GRLQGREGGR..VAWVADPK.................DPRK
hAgo2  CEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMRKKYRVCNVTRRPASHQTFPLQQESGQTVECTVAQYF
```
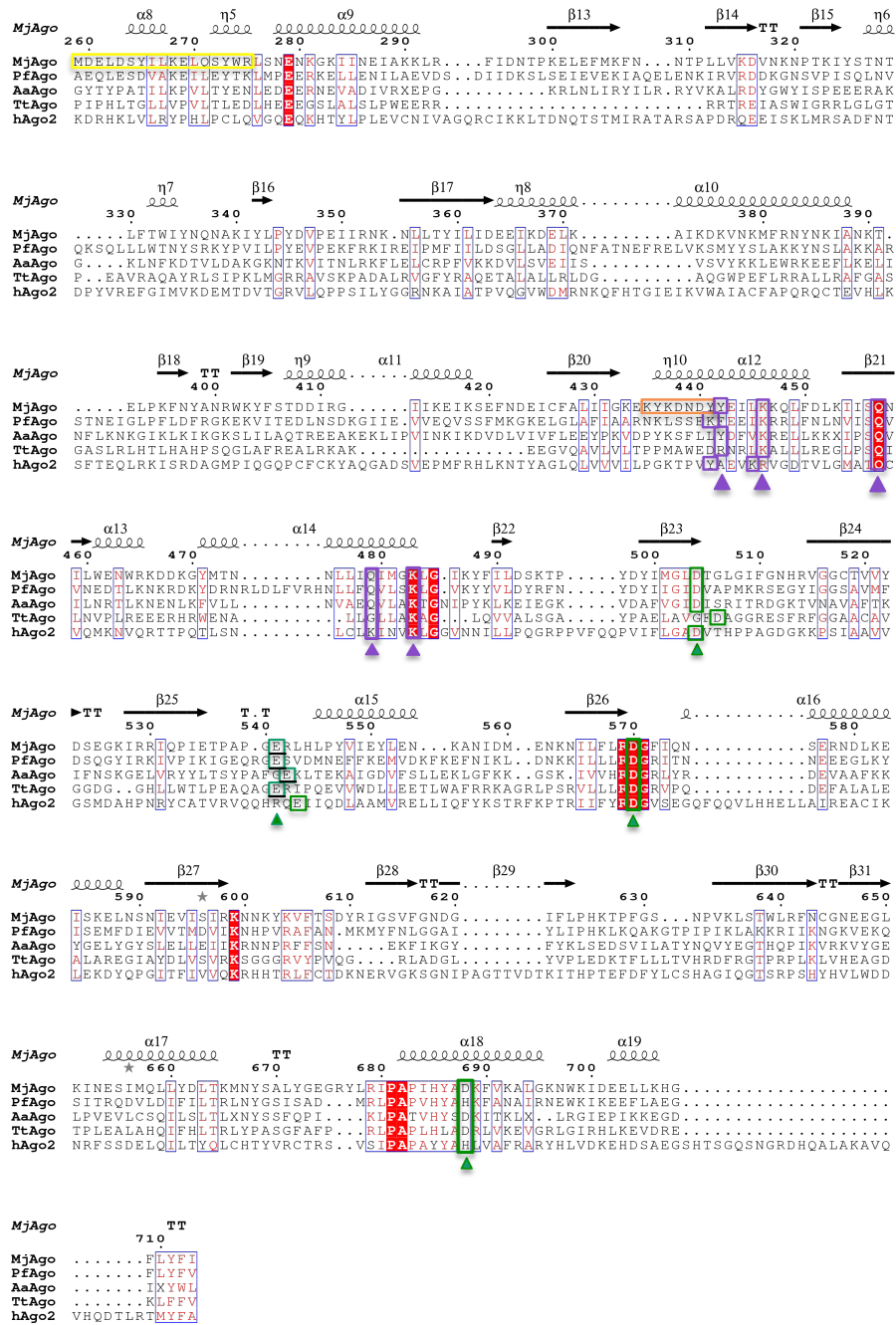
Figure 3: Structure-based sequence alignments of Ago proteins. Secondary structure assignment and numbering shown on top correlate with MjAgo. The aligned sequences are in the order of *M. jannaschii*, *P. furiosus* (PDB entry 1Z25), *Aquifex aeolicus* (PDB entry 1YVU), *T. thermophiles* (PDB entry 3DLH) and hAgo2 (PDB entry 4OLA). Residues building up the catalytic slicer site DED-D/H/N within the PIWI domain are highlighted in a green box with a green arrow. The amino acids located on a flexible PIWI loop that are required for the induced-fit mechanism in the catalytic tetrad in prokaryotic Argonautes are underscored by a black line. Critical residues that coordinate the 5′-phosphate and Mg^{2+} ion (K/R/Y-KQK) are highlighted in a purple box with a purple arrow. The nucleotide specificity loop (residues 435-441) within the Mid domain is highlighted in an orange box while the the the switch helix (residues 259-275) is highlighted in a yellow box. Conserved residues are shown in white and highlighted in a red frame. Similar residues are in red and columns are framed in blue if more than 70 % of its residues are similar. The alignment was performed using ESPRIPT.

Furthermore it was possible to obtain co-crystals of MjAgo in complex with a 21-base DNA guide strand. As observed within other Ago variants, the DNA threats its way through the

protein, thereby contacting all domains and linkers **(Figure 13)**. However, no sequence-specific contacts are made to the guide DNA by the protein. Critical conserved residues that coordinate the 5′-phosphate and Mg^{2+} ion and are thus responsible for the proper insertion and anchoring of the guide strand within the Mid domain could be identified **(Figure 3)**. This extensive network puts the first base out of the helical stack (as observed for the seed region) and explains why this base does not extremely contribute to target recognition. Based on the unique architecture of the PAZ domain, the guide strand in MjAgo is orientated in a different binding position as in observed in TtAgo and provides a unique preference for the 3′-end base of the guide strand to the protein [144].

X-ray crystal structures revealed that some Ago proteins can recognize the identity of the 5′-end bases of the guide strand through direct interactions with the specificity loop within their Mid domains. Activity assays demonstrated that MjAgo is able to use 5′ phosphorylated guide strands with all different nucleobases [144]. However, the nucleotide specificity loop (residues 435-441) within the Mid domain is essential for full catalytic activity **(Figure 3)**.

In order to fully accommodate the DNA guide strand, several pronounced changes in the architecture of MjAgo can be observed. This include for example the reorganization of the domains within the two lobes that results in an opening of the nucleotide binding channel as well as the rotation of the PAZ domain around a hinge region **(Figure 15)**. Furthermore an important structural element, termed the switch helix (residues 259-275) gets rearranged and structured up on substrate binding **(Figure 3)**. This helix is equivalent with a so called helix 7 that introduces in some Agos a distinct kink between nucleotides 6 and 7 and thus interferes with the A-form arrangement of the guide strand. Cleavage assays showed that in MjAgo a complete structured helix is required for the formation of a cleavage competent ternary complex (by using guide strands that carry a pyrimidine base at the 5′-end) [144].

.

5.2. Optimization of protein production and purification

For structural characterization, Adrian Zander - a member from our collaboration group of Dr. Dina Grohmann from the University of Regensburg - expressed and purified appropriate amounts of recombinant MjAgo.

Samples were tested for their purity by SDS-Page (**Figure 4B**), but analysis of the protein by dynamic light scattering (DLS) indicated the presence of large, nonspecific aggregates. This result was verified by applying the sample on gel filtration chromatography, showing multiple oligomerization states of the protein (**Figure 4A+B**). Efforts to remove aggregates via heating, spinning and filtering as well as the performance of a thermal shift assay (TSA) failed to enhance the quality and stability of the sample (data not shown). However, filtered protein was subjected to initial crystallization screens but resulted only in clear drops or aggregated protein, caused probably by polydispersity and low concentration of the protein.

In order to avoid damage caused by freezing or transportation, recombinant protein was expressed and purified in our laboratory according to an established protocol, initial set up for characterization of MjAgo [55]. This work showed that recombinant MjAgo is heat stable up to 85 °C and requires high temperatures for DNA binding and slicing activity (65 °C and 75 °C) [55]. Therefore, MjAgo was purified using a heat denaturation step (85 °C for 45 min), followed by immobilized metal affinity chromatography (IMAC). However, analysis of the aggregation status of the protein during the purification-steps revealed, that despite the high purity, protein aggregation still occurred at low concentrations (data not shown).

In order to increase the likelihood for crystallization, the expression and purification strategy were optimized to improve stability and homogenity of the target protein. Based on the high efficiency to selectively separate recombinant protein from crude sample as well as the possible instability of the recombinantly expressed protein towards high temperatures, the heat-denaturation step was discarded and IMAC was selected as initial purification step. Therefore, MjAgo was isolated by a two-step purification protocol, including IMAC, tag removal and size-exclusion chromatography. In combination with using high ionic strength buffers and shortening of the expression time, purified protein remained stable and eluted on gel filtration chromatography as a single peak (retention volume 77 mL), indicating monomeric protein (**Figure 4C**).

A

B

C

D
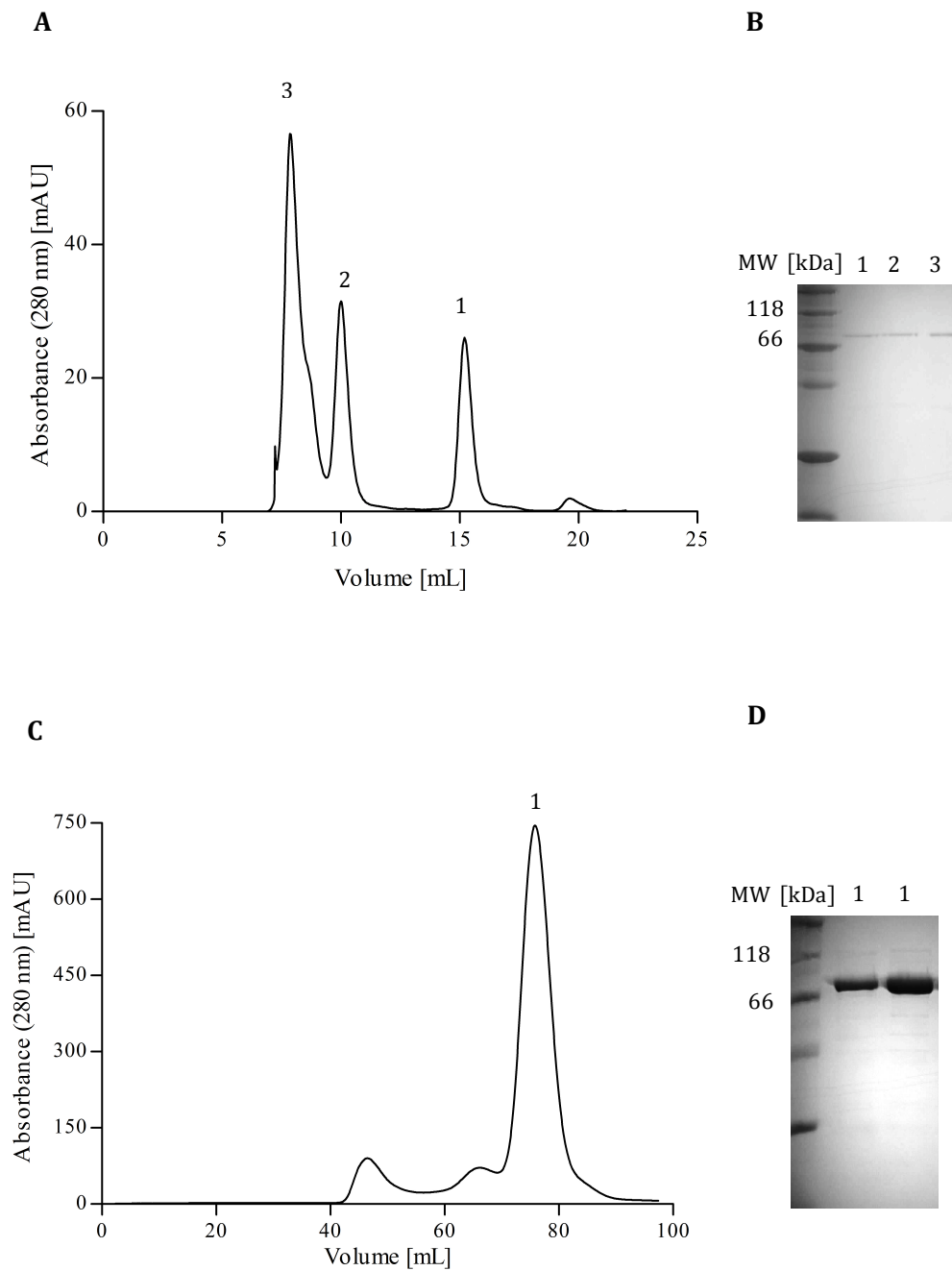
Figure 4: Purification of recombinant MjAgo. Gel filtration profiles of recombinant MjAgo **A)** before and **C)** after optimization of the purification strategy. Purification was performed on an ÄKTAxpress system using a **A)** HiLoad 10/30 Superdex 200 column and **C)** HiLoad 16/60 Superdex 200 column (GE Healthcare). **B)** and **D)** 10 % SDS-PAGES showing numbered fractions of corresponding peaks of gel filtration chromatography.

The purity of the protein in the major elution peak was verified by SDS-PAGE (**Figure 4D**) and recombinantly expressed MjAgo could be concentrated at low speed up to 7-8 mg/mL. In order to analyse homogeneity as well as to exclude the presence of aggregates, concentrated protein was further analysed by DLS.

5.2.1.   Characterization of recombinant MjAgo

5.2.1.1.   Dynamic Light Scattering

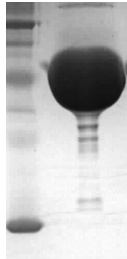DLS represents a powerful tool to measure the size distribution and polydispersity as well as the aggregation state of proteins in solution [96]. Therefore, concentrated MjAgo was analysed by DLS to verify homogeneity as well as to exclude the presence of aggregates prior to setting up crystallization trails. Additionally, recombinantly expressed MjAgo was tested for its stability and homogenity at high temperatures (55 °C and 85 °C).

As seen in **Table 1**, after optimization of the purification protocol, concentrated protein was monodisperse, free of aggregates and displayed a size of 87 kDA, that is in good agreement with the deduced molecular weight of MjAgo (~ 85 kDa).

	% PD	MW-R [kDA]	Radius [nm]
1	7.1	87	4.1
2	20.3	94	4.2
3	Multimodal	-	294

MW [kDa]

118

66

A

Table 1: Mean values of recombinant MjAgo samples, analysed by DLS. **1)** Concentrated MjAgo at 7 mg/mL. **2)** Sample after incubation at 55 °C. **3)** Sample after incubation at 85 °C. Sum of squares (SOS); Polydispersity PD; Molecular weight (MW) hydrodynamic radius R_h. **A)** Concentrated MjAgo analyzed by DLS and used for crystallization**.** 10 % SDS Page showing 3 µL of purified MjAgo at approximately 5 mg/mL.

In addition, incubation at 55 °C for 10 min let to an increase in polydispersity (from 7.1 to 20.3 %). Furthermore, heat treatment up to 85 °C caused visible flocculation of the sample, which was immediately confirmed by DLS (**Table 1**). Efforts to remove aggregates via filtering and centrifugation failed to regain the quality of the sample.

In summary, after changing the production and purification strategy, recombinantly expressed MjAgo could be isolated using a two-step purification protocol, resulting in a pure, monodisperse protein solution. Although derived from a hyperthermophilic organism, recombinantly expressed MjAgo is less thermophilic and thermostable than observed previously [55].

These different results might be explained by the use of a heterologous expression host as well as by the high protein concentration required for crystallization. Thereby, the high concentrations might promote aggregation which does not occur at low protein cocnentrations that are required for the analysis of enzyme activity (< 10µM).

5.2.1.2.   Thermal shift assay

In order to optimize the buffer formulation for crystallization, the thermofluor screen of EMBL (Hamburg, GER) was utilized. Thereby, a mixture of 96 chemical reagents is used to systematically analyse the stability of the target protein over a broad pH and buffer range during thermally induced unfolding [97]. As the temperature increases, an environmentally sensitive dye interacts with exposed hydrophobic protein regions upon heat denaturation, thus regaining significantly its fluorescence. To detect a condition that stabilizes the protein, the melting temperature (T_m) (the temperature where the protein is 50 % unfolded) is evaluated for each buffer condition and compared with the reference T_m (which is the protein in water).
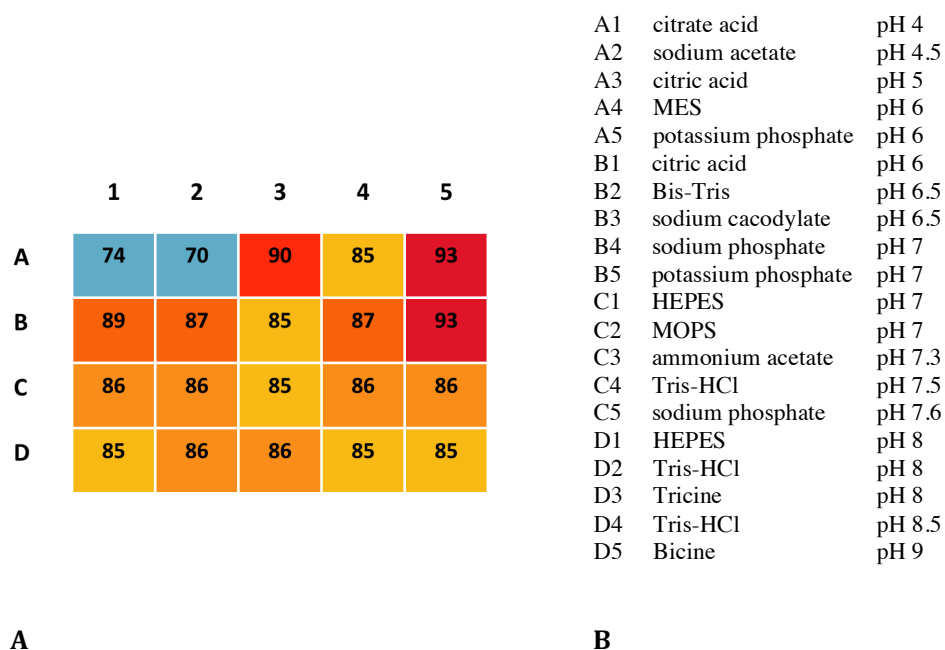
	1	2	3	4	5
A	74	70	90	85	93
B	89	87	85	87	93
C	86	86	85	86	86
D	85	86	86	85	85

A1	citrate acid	pH 4
A2	sodium acetate	pH 4.5
A3	citric acid	pH 5
A4	MES	pH 6
A5	potassium phosphate	pH 6
B1	citric acid	pH 6
B2	Bis-Tris	pH 6.5
B3	sodium cacodylate	pH 6.5
B4	sodium phosphate	pH 7
B5	potassium phosphate	pH 7
C1	HEPES	pH 7
C2	MOPS	pH 7
C3	ammonium acetate	pH 7.3
C4	Tris-HCl	pH 7.5
C5	sodium phosphate	pH 7.6
D1	HEPES	pH 8
D2	Tris-HCl	pH 8
D3	Tricine	pH 8
D4	Tris-HCl	pH 8.5
D5	Bicine	pH 9

A   **B**

Figure 5: Heat map of the T_m. values for MjAgo during thermally induced unfolding. A) Heat map of different T_m values for recombinant MjAgo in the presence of various buffer compositions as measured by TSA. The red colours correspond to the highest T_m values, orange to the average T_m and blue the lowest T_m. **B)** Buffer formulations with 250 mM NaCl used for TSA.

As expected, MjAgo displayed an overall high resistance towards elevated temperatures, resulting in high T_m values for all buffers ($T_m > 70$ °C) (**Figure 5**). Interestingly recombinant protein remained constantly stable above the pH range from pH 5.0 to pH 9, with a marginal preference for potassium phosphate at pH 6 and pH 7 (T_m of 93 °C).

A destabilizing effect could only be observed at very low pH values using citric acid at pH 4 (T_m of 74 °C) and acetate buffer at pH 4.5 (T_m of 70 °C). This response is probably caused by the acidic pH and not by the properties of the buffer system since a shift in the pH of citric acid resulted in a shift of ΔT_m of 16 °C respectively. Additionally, citric acid at pH 5 and 6 were together with potassium phosphate the most effective buffers in enhancing conformational stability (**Figure 5**). This is in agreement with a previous study, where the authors discovered that potassium phosphate and sodium citrate were among the most stabilizing buffers for 11 of total 17 *E. coli* proteins tested and analysed by TSA [98].

However, testing citric acid at pH 6 for purification resulted in precipitation of the protein. Based on incompatible properties of phosphate buffer for purification (phosphates sequester divalent cations such as Ca^{2+} and Mg^{2+}) and crystallization, the original buffer formulation (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 % glycerol, 5 mM $MgCl_2$) was used for all further experiments.

5.3. Crystallization and structure determination

For initial crystallization screens, sparse matrix screens from Quiagen were utilized using the sitting drop vapour diffusion method. Therefore, freshly prepared or thawed protein at approximately 7 mg/mL were mixed with equal amounts of mother liquor and crystals appeared within 3 to 5 days in 0.1 M tri-sodium citrate, 20 % isopropanol, 20 % PEG 4000 or 0.1 M tri-sodium citrate, 20 % isopropanol and 0.1 M HEPES-HCl, pH 7.5, at 20 °C (**Figure 6**). In order to optimize crystal quality, crystallization was further refined using the hanging drop vapour diffusion method with a 1:1 (v/v) ratio of protein to mother liquor. The quality and size of crystals was significantly improved using microseeding**.** Therefore, 0.5 µL microseeds were added to a crystallization drop containing 1 µL mother liquor with 1 µL protein solution and crystals appeared after 2 to 4 days in 0.1 M tri-sodium citrate, 10 % isopropanol, 8 % PEG 4000 or 0.1 M HEPES-HCl, pH 7.5, 12 % isopropanol and 0.05 M tri-sodium citrate (**Figure 6**). Prior to flash-freezing in liquid nitrogen, crystals were cryoprotected by transferring them into the crystallization solution supplemented with either 30 % ethylene glycol or 30 % 2-Methyl-2,4-pentanediol (MPD).

5.3.1. Dehydration

Reducing the solvent content within a protein crystal can result in more ordered and closely packed crystals, thus extending its diffraction pattern [99-102]. One of the most widely applied methods for dehydrating crystals is the transfer of the crystal into a new drop of dehydration solution (which is, in most cases, the original crystallization condition with a higher concentration of precipitant or supplemented with other cryoprotective agents) and equilibrated for a specific time period against a reservoir containing the same dehydration formula [103].

To test if dehydration could have an impact on the diffraction quality, various MjAgo crystals were reproduced via microseeding **(chapter 5.3)** and soaked for 12-16 h hours in dehydration solution containing the crystallization condition (0.1 M tri-sodium citrate, 10 % isopropanol and 8 % PEG 4000), enriched with either glycerol, MPD, ethylene glycol or an increased PEG

concentration. Analysis of the dehydration process revealed that incubation at 20 °C resulted in dissolving of the crystals by all tested substances, whereas incubation at 4 °C led to more robust crystals which did not show any signs of cracking during dehydration. MPD did not show any effect on the crystal quality, whereas ethylene glycol improved the diffraction resolution from 2.9 to 2.75 Å. Additionally, a significant improvement in the crystal diffraction quality was observed by using glycerol and a higher PEG concentration, which extended the resolution to 2.45 and 2.3 Å, respectively (**Figure 6**).

Taken together, reducing the solvent content led to an extended diffraction patter and caused no significant changes in the asymmetric unit of MjAgo crystals.
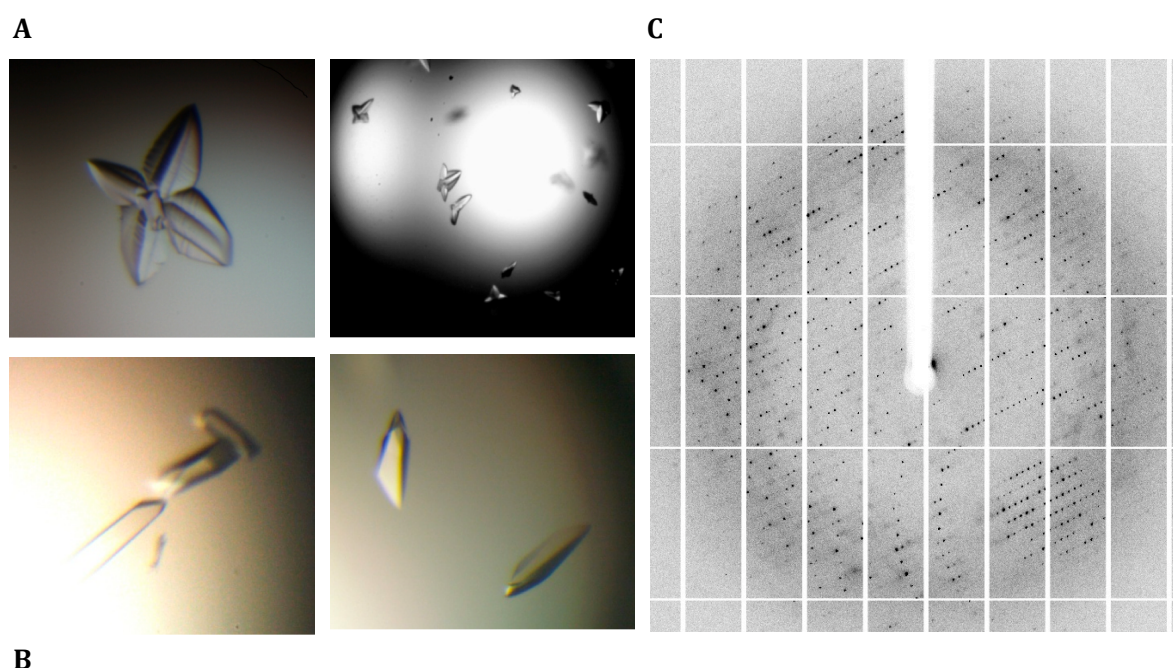
A   **C**

B

Figure 6: Protein crystals and diffraction pattern of MjAgo. Crystals of recombinant MjAgo using the hanging drop vapour diffusion method **A)** before and **B)** after optimization of the crystallization process. **C)** Diffraction pattern of a dehydrated MjAgo crystal.

Crystals of MjAgo belonged to the orthorhombic space group $P2_12_12_1$ with one MjAgo in the asymmetric unit. Diffraction data were collected and solved according to **chapter 7.2.14** with following unit cell axes: a = 63.5 Å, b = 104.9 Å, c = 115.1 Å. Data processing statistics are summarized in **Table 2**. Initial phases required for structure determination of MjAgo were obtained experimentally through Mercury-derivatized crystals (**chapter 5.3.3**).

5.3.2.   Solving the phase problem

Determination of crystallographic phases is crucial in order to define novel protein crystal structures. In principle, three existing techniques are at hand which can be used to receive initial phases for large protein molecules: molecular replacement, isomorphous replacement and anomalous diffraction [104, 105]. Molecular replacement requires the pre-existence of a fairly complete homologous structure-model, which shares at least 35 % sequence identity with the unknown structure [104]. If no structure homology model can be utilized, phases must be determined experimentally either by the attachment of heavy atoms to the native crystals (isomorphous replacement) or by introducing of anomalous scatters into the crystal, which is usually achieved by incorporation within the polypeptide itself (anomalous scattering) [105, 106]. The standard method for obtaining phases in anomalous scattering is to introduce selenomethionine (SeMet) in place of methionine residues (achieved by blocking methionine biosynthesis in the host during protein expression). The selenium atoms (which replace the sulfur atoms) have a strong anomalous signal at wavelengths near the absorption edge and are in most cases nearly complete and uniformly incorporated within the protein itself.

Although many structures of the Ago protein family have already been characterized, none existing homologous structure model could be used for molecular replacement based on low sequence identity to MjAgo (28 % sequence identity to PfAgo) [67]. Due to a high methionine content within the sequence (14 in 720 amino acids), the expression and incorporation of SeMet substituted MjAgo was explored first in order to receive initial phases for structural characterization. Therefore, protein production was tested in M9 media using several *E. coli* strains in combination with different expression temperatures. However, none of them yielded soluble protein.

Consequently, different heavy atom types were selected based on their reactivity towards specific amino-acid ligands represented within the primary sequence of MjAgo. In order to increase the chance of success, MjAgo was successfully crystallized using a broader pH and buffer range (MES pH 6.5, HEPES-HCl pH 6.5 to pH 7.5), since solubility and reactivity of the heavy-atom reagents strongly depend on the mother liquor [105]. Additionally, various concentrations, ranging from 1 mM to 10 mM in combination with different soaking periods were also tried out since both long and quick soaks have been shown to be highly effective during derivatization [105, 107-109]. However, using high concentrations for 10 min resulted

in cracking and a much lower diffraction pattern by all tested crystals, whereas lower concentrations over a longer time period at 4 °C were more suitable. Soaked crystals in platinum retained good resolution but analysis of the received data revealed only a small number of derivatized sites, causing changes in unit cell parameters. In contrast, using grains of Hg(OAc)2 in the crystallization drop for 4 h resulted in a high diffracting quality crystal which was suitable for data collection (Data collection and refinement statistics are summarized in **Table 2**). After cryoprotection with 30 % ethylene glycol, the data set was collected and the crystal structure of MjAgo was solved at a maximum resolution of 2.9 Å by using the single-wavelength anomalous diffraction method (SAD) for phase calculation.

Table 2: Crystallographic data and refinement statistics

	free MjAgo*	MjAgo binary complex*
Wavelength		
Resolution range	49.1-2.3 (2.4-2.3)	44.4-2.85 (3.0-2.85)
Space group	P $2_1 2_1 2_1$	P $4_3 2_1 2$
Unit cell	63.5 104.9 115.1	118.0 118.0 134.6
Total reflections	250,495 (22,010)	122,590 (17,050)
Unique reflections	34,953 (3,103)	22,659 (2,222)
Multiplicity	7.2 (7.1)	5.4 (5.2)
Completeness (%)	0.99 (0. 99)	0.99 (0.99)
Mean I/sigma(I)	18.8 (1.19)	7.8 (1.1)
Wilson B-factor	68.0	87.1
R-merge	0.05 (1.43)	0.12 (1.2)
R-meas	0.05 (1.54)	0.13 (1.38)
R-pim	0.02(0.57)	0.05 (0.59)
CC $_{1/2}$	0.999 (0.568)	0.997 (0.520)
Reflections used in refinement	34,944 (3,103)	22,652 (2,222)
Reflections used for R-free	1,795 (153)	1,146 (113)
R-work	0.203	0.228
R-free	0.239	0.285
CC (main chain / side chain)	0.95 / 0.81	0.94 / 0.82
Number of non-hydrogen atoms	5,568	5410
macromolecules	5,557	5374
ligands	9	36
solvent	2	
Protein residues	692	671
RMS(bonds)	0.015	0.015
RMS(angles)	1.9	1.3
Ramachandran favored (%)	95	92
Ramachandran allowed (%)	4.4	7.8
Ramachandran outliers (%)	0.15	0.46
Rotamer outliers (%)	6.6	13
Clashscore	6.8	6.9
Average B-factor	90.3	96.1
macromolecules	90.3	95.9
ligands	107.3	127.5
solvent	65.4	

Statistics for the highest-resolution shell are represent in parentheses.
(* Friedel mates were averaged when calculating reflection statistics.)

5.3.3.   Structure analysis

5.3.3.1.   Overall structure

As expected, MjAgo adopts the characteristic structure and organization of Ago proteins, consisting of four predominant domains, the N, PAZ, Mid and PIWI. The overall bilobal architecture composed of the PAZ and PIWI lobes is connected through a structured linker L2, that forms the base of the central cleft **(Figure 7)**. The N terminal half, build up by the N and PAZ domain is separated by the internal linker L1, that holds the PAZ domain above the Mid an PIWI domains. Due to this arrangement, the PAZ, Mid and PIWI domains are aligned in the free state of MjAgo in a triangular fashion, with each domain contacting the other two domains of the protein **(Figure 7)**.
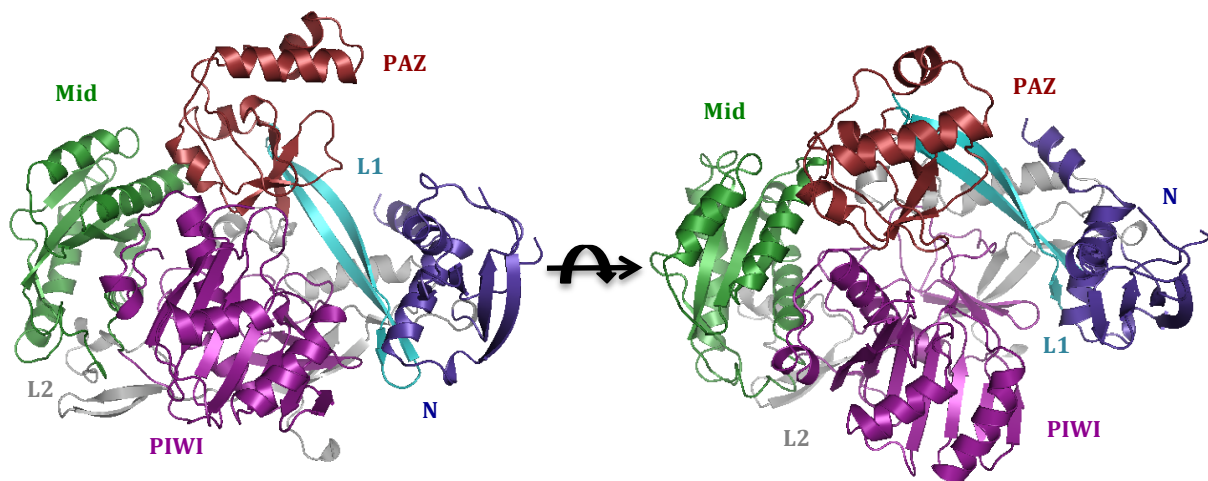

**Figure 7: Structural features of the Argonaute protein from *M. jannaschii.* Crystal structure of MjAgo showing the N-terminal domain (blue), the linker L1 (cyan), the PAZ domain (red), the interdomain connector L2 (gray). the Mid domain (green) and the PIWI domain (purple).

Despite low sequence similarity, the overall structure of MjAgo is remarkably similar to the Ago protein from *P. furiosus* (PDB code 1Z25), consisting of 30 beta-sheets and 18 alpha-helices **(Figure 8)**. However, marginal differences between the two proteins in both the N and PAZ domains is compatible with previous structural studies, demonstrating that these domains comprise the highest structural variety of all Ago proteins. In particular, the N domain of MjAgo is slightly shorter compared to that from PfAgo and the PAZ domain differs in both its β barrel fold and its position relative to the Mid/PIWI lobe **(Figure 8)**.
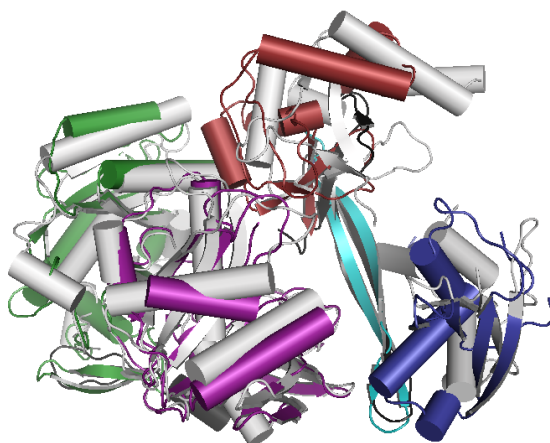
Figure 8: Superposition of MjAgo in its free state with PfAgo. The MjAgo structure
(coloured according to figure 7) is aligned with PfAgo (shown in grey; PDB entry 1Z25).

The PIWI domain located at the C terminus of the protein comprises the highest structural
conservation among Ago proteins and harbours the catalytic tetrad DEDX. Although included
in the crystallization and purification buffer, no Mg^{2+} could be found at the active site.
However, three highly conserved amino acids, located at structural equivalent position as
active-site residues found in other Ago variants [32] could be identified within MjAgo (D504,
D570, D688) **(Figure 9)**. Additionally, conserved Glu541 can be found at a structural
equivalent location as in TtAgo (termed the glutamate finger) [81] supporting the idea of the
plugged in mechanism, responsible for slicer activity. Therefore, we suggest the catalytic
tetrad DDED (D504, E541, D570, D688) building up the catalytic core within the PIWI
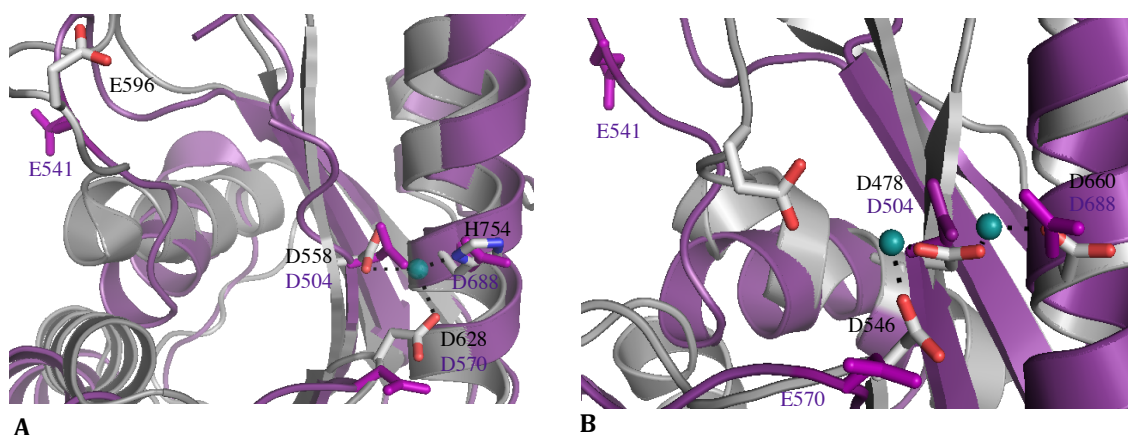domain of MjAgo.

Figure 9: Positioning of catalytic residues within the PIWI domain of MjAgo. Superpositioning of the PIWI domains in cartoon representation of MjAgo (shown in purple) with **A)** PfAgo (shown in grey; PDB entry 1Z25) and **B)** TtAgo (shown in grey; PDB entry 4NCB). **A)** Active site residues in the catalytic invariant version of MjAgo and PfAgo are shown in stick representation with the Glu 596/541 outside and directed away from the catalytic pocket. The Mg^{2+} ion in the active site of PfAgo is shown in cyan. **B)** Active site residues in the catalytic invariant version of MjAgo and cleavage-compatible version of TtAgo are shown in stick representation with the Glu512 of TtAgo inserted into the catalytic pocket. The Mg^{2+} ions in the active site of TtAgo are represent in cyan.

However, supposed amino acids involved in catalytic activity of MjAgo require further characterization. To assess whether either of these conserved residues is involved in slicer activity, each position should be mutated to alanine and examined on endonucleolytic activity. Furthermore, to aid in location of residues involved in metal coordination, native MjAgo crystals could be soaked with $MnCl_2$. The metal has been shown to increase cleavage activity in pAgos (compared to $MgCl_2$) [53, 81] and is heavier and more comfortable to locate crystallographically.

5.4. MjAgo in complex with guide and target DNA

Although derived from a hyperthermophilic organism, recombinant MjAgo showed a decreased resistance towards elevated temperatures (> 55 °C) when expressed in *E. coli*, thus resulting in aggregation and oligomerization of the protein **(chapter 5.2.1)**. However, further studies demonstrated that the enzyme requires temperatures above 65 °C for DNA binding and a minimum of 75 °C for catalytic activity [55]. In order to obtain a pure, homogenous protein-DNA complex solution suitable for crystallization, binding of a 5′-phosphorylated 21mer guide strand was carried out and analyzed at increasing temperatures. The sequence of the used guide strand (**Figure 11**) is based on the human miRNA hsalet- 7a-5p and binding was demonstrated in a concentration-dependent manner [55].

Therefore, DNA was added to diluted protein (≤ 1.5 mg/mL) using a ratio of 1:2 (protein to DNA) and incubated for 10 min at various temperatures. After binding, possible aggregates were removed and the thermostability of the formed complex was tested by TSA. As seen in **Figure 10**, adding the substrate to the protein gives a positive shift in the ΔT_m by all tested temperatures, thus indicating a reduced conformational flexibility and binding of the DNA to the protein. Interestingly, an increase in stability can already be observed at low temperatures. Additionally, increased temperatures above 40 °C caused a simultaneously rise of precipitation of the complex as analyzed by SDS-PAGE **(Figure 10)**.

Based on these results, a low temperature scale was selected (20 °C and 37 °C) for all further experiments, in order to ensure the formation of the binary complex as well as to avoid aggregation of the protein.
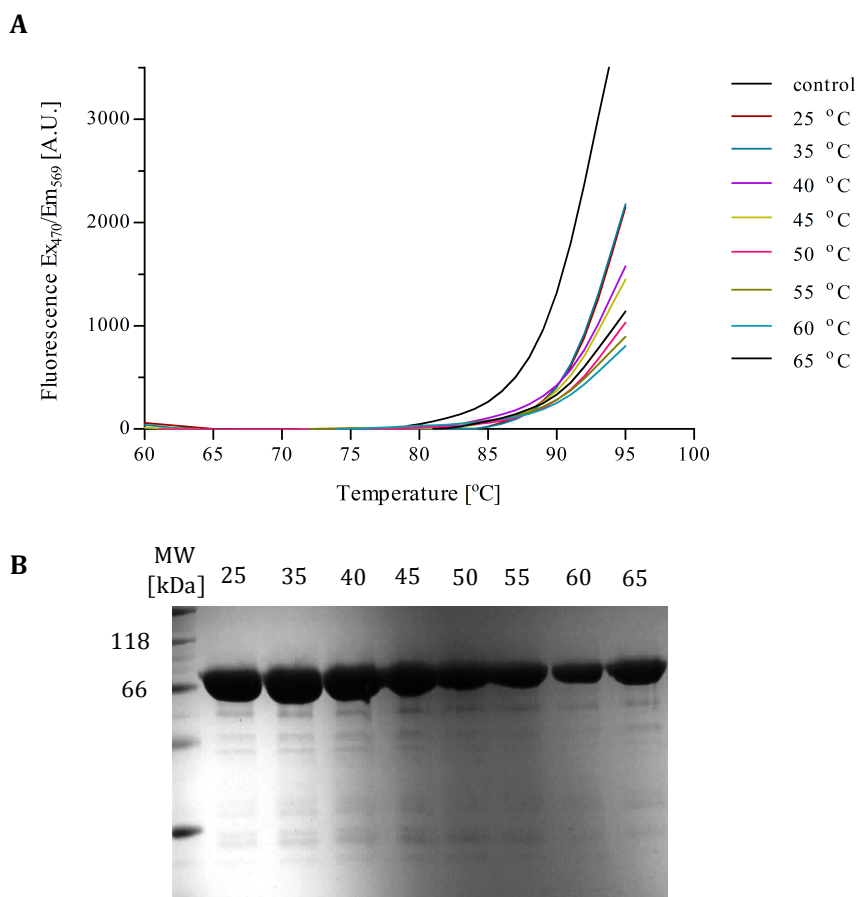
A

B

Figure 10: Analysis of DNA binding by MjAgo at elevated temperatures. A) Effect of temperature on binary complex thermal stability. Tested pre-incubation temperatures for complex formation are shown on the right hand site of the graph, whereas the control represents pure protein. **B)** Analysis of soluble protein after complex formation on 10 % SDS-PAGE. The corresponding pre-incubation temperatures are shown on top of each lane.

5.4.1. Crystallization and structure determination of the binary complex

Although the crystallization of protein DNA complexes often represents a unique challenge, several strategies can be applied to obtain co-crystals of macromolecular complexes. These include for example the variation of the sequence, length and termini (blunt-end or overhang) of the oligodeoxynucleotide itself, the stoichiometric ratio of protein to DNA as well as the pH and ionic strength of the buffer system [110-113].

Therefore, recombinant MjAgo was purified by using different Tris-HCl buffer formulations, (with varying pH and ionic strength) and screened for crystallization by testing different guide strands in combination with various molar ratios (1:1.2; 1:1.5; 1:2) of protein to DNA. Since the binary structure of TtAgo revealed that the guide strand is anchored at both its ends within

33

the complex [67], testing oligonucleotides differing in termini was not carried out. Additionally, due to weak affinity of the substrate, further purification of the formed complex by gel filtration prior to crystallization has not been applied.

A   guide strand 5′-phos-TGA GGT AGT AGG TTG TAT AGT 3′

           target strand 5′-phos-TAT ACA ACC TAC TAC CTC GT 3′

B   guide strand 5′-phos-TAG AGG TAC GTG CTG AGG CTT 3′

Figure 11: DNA alignment of guide and target strands used for crystallization. Nucleotide sequences of the guide and target strand used for crystallization of the binary and ternary complex of MjAgo. The sequence of the first guide strand (**A**) is based on the human let-7 miRNa, whereas the second guide strand **B)** is sequence-optimized for binding to MjAgo.

In one attempt, the first guide strand was added to concentrated MjAgo at approximately 7.5 mg/mL and complex formation was performed *via* incubation at 37 °C for 3-4 hours. To assess whether the protein aggregates or oligomerize upon binding, the sample was centrifuged and analyzed by DLS. As seen in **Table 3 (chapter 5.4.2)**, multiple peaks with an increase in polydispersity indicated unbound DNA as well as aggregation of the protein.

However, filtered binary complex was subjected to a wide variety of initial crystallization screens using the sitting drop vapour diffusion method. Therefore 0.15 µL of the complex (in 50 mM Tris-HCl pH. 7.4, 300 mM NaCl, 10 % glycerol) were mixed with 0.15 µL reservoir solution and crystallization plates were stored at 4, 20 and 37 °C, respectively.

After two months at 20 °C, triangle shaped crystals appeared using a molar ratio of 1:2 of protein to DNA in 0.1 M citric acid, pH 4 and 1.6 M Ammonium sulfate (**Figure 12**). Crystals were briefly soaked in reservoir solution enriched with 2 M Li_2SO_4 (lithium sulphate) as cryoprotectant and flash frozen in liquid nitrogen.
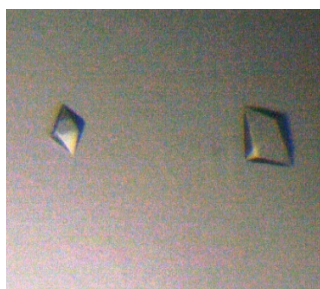

Figure 12: Protein crystals of the binary complex. Crystals of MjAgo in complex with a 21-base guide DNA strand using the sitting drop vapour diffusion method.

After cryoprotection and freezing, diffraction data were collected to maximum resolution of 2.85 Å and the structure was solved by molecular replacement using apo-MjAgo as model structure (data processing was performed as described in chapter 7.2.14). The crystals of the binary complex belonged to the tetragonal space group P $4_3 2_1 2$ with one MjAgo in the asymmetric unit and cell dimension a,b=118 Å and c=134.6 Å.

Since the first guide strand is mainly disordered along the seed sequence **(Figure 13)**, an additional guide strand (varying in sequence composition) **(Figure 11)** was tested for crystallization of the binary complex. Using a sequence optimized oligonucleotide for binding should result in a more rigid conformation, thus enhancing the likelihood for crystal complex formation. However, applying the same approach as mentioned above, no crystals were obtained for this binary complex.

Therefore, since soaking and seeding were also unsuccessful, a different attempt was carried out for crystallizing the binary complex. Further studies implicated the formation of precipitate upon mixing protein and oligonucleotides at high concentrations suitable for crystallization [112]. Therefore, in order to avoid high local concentrations of DNA, the oligonucleotides were pre-mixed with buffer and added slowly to diluted protein (≤ 1 mg/mL). After binding over night at RT, the formed complex was filtered, concentrated by low speed to 7-9 mg/mL and inspected by DLS. Previous DLS measurements revealed that apo-MjAgo is monodisperse and has a hydrodynamic radius (R_H) of 4.1 nm, respectively (**chapter 5.2.1**). After mixing MjAgo with the guide strand, two peaks corresponding to a R_H of 0.85 nm and 4.7 nm are apparent, thus indicating structural rearrangement and incorporation of the DNA within the protein (**Table 3**). Additionally, adding the diluted guide strand to low concentrated protein led to a dramatic reduction in aggregation, resulting in a monodisperse, monomodal protein-complex solution.

However, testing the sample in few initial crystallizations screens, only spherulites and crystalline aggregates could be observed. Due to time limits, no further attempts could be tried out to obtain co-crystals of this binary complex. These would include for example the optimization of the existing hit conditions by varying the pH or temperature, concentration and type of precipitant, or by adding additives for an additional screen with a more restricted range of conditions [114]. Indeed, using monovalent or divalent metal ions such as $MnCl_2$ or $MgCl_2$ have been shown to be highly effective in stabilizing of the formed protein RNA/DNA complex, making it more favourable for crystallization [68]. Also varying the length and

screening with different DNA strands have successfully been applied in order to grow cocrystals of protein-DNA complexes [110, 68].

5.4.1.1.   Recognition of the guide strand by MjAgo

In its overall structure, MjAgo in complex with 21mer guide DNA is similar to the TtAgo and hAgo2 binary complex structures. Both ends of the guide strand are anchored in their respective binding pockets **(Figure 13)**. The bound DNA is traceable for nucleotides 1-6 and 19-21, whereas the 5′-phosphorylated end is only partially ordered. As within other binary structures, the guide DNA threads its way through a specific binding channel whereas the majority of the contact points are predominantly between the sugar-phosphate backbone and residues of the PIWI domain, explaining why Agos can target nearly any gene for silencing.
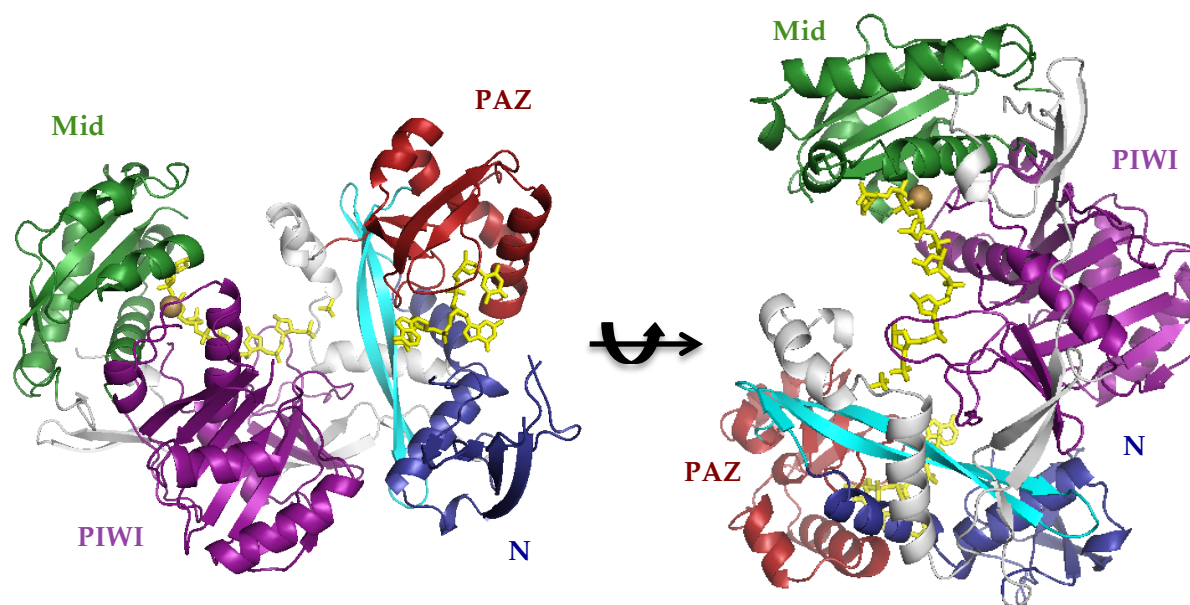
Figure 13: Crystal structure of MjAgo bound to a 5'-phosphorylated 21-base DNA guide strand. The protein is shown in cartoon representation with the N domain in blue, the PAZ domain in red, the Mid domain in green and the PIWI domain in purple. The DNA (colored in yellow) is shown in stick representation and can be traced for nucleotides 1 to 6 and 19 to 21. The metal ion within the Mid domain is shown in deepolive.

The nucleotides 2-6 of the guide strand (termed the seed region) are exposed to the solvent and organized in an A-form helical conformation in order to simplify recognition and base pairing with the target DNA **(Figure 13)**.

The 5′-end of the guide strand is anchored at the interface between the Mid and PIWI domains. Here, four side chains of highly conserved residues contact the 5′-end phosphate group of the first nucleotide (Tyr442, Lys446, Gln457, Lys483) **(Figur 14)**. The main chain of the beta 23-sheet within the Mid domain is also participating in contacting the 5′ phosphate. This extensive network of salt linkages and hydrogen bonds puts the 5′-base out of the helical stack (as observed for the seed region) and explain why the first base does not extremely contribute to target binding.

The Mg^{2+} located in the binding cleft within the Mid domain is coordinated by Gln479 and Lys483 as well as the carboxy-terminal carboxylate end (Ile713) of the PIWI domain **(Figure 14)**. The importance of these residues for anchoring the DNA within the protein could clearly be demonstrated by single mutations of Gln457, Asn458 or Lys483 to Alanine, which completely abolished slicing activity [144]. Furthermore, Gln479Ala and Asp438Pro mutations resulted in an altered cleavage pattern of the target strand [144].

Activity assays demonstrated that MjAgo is able to use guide strands with all different nucleobases [144]. Moreover, MjAgo is more efficient and shows a distinct cleavage pattern by utilzing guide strands that carry a purine base at the 5′-end. This indicates that the identity and the resulting interactions with the 5′-nucleotide by the Mid domain effects and influence the positioning of the guide-target duplex within the protein.
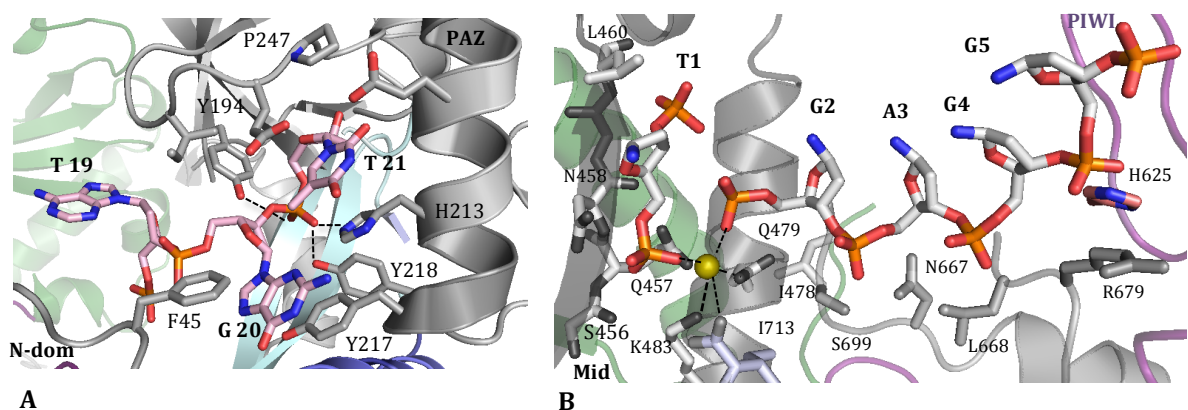
Figure 14: Guide recognition by the nucleotide-binding pockets of MjAgo. A) The PAZ and N domains are shown in ribbon representation with the residues that interact with the 3′-end of the guide DNA shown as sticks. **B)** The Mid and PIWI domains that anchor the 5′-end of the guide DNA are illustrated in ribbon representation with the residues that contact the guide strand present as sticks. The hydrogen bonds are shown in dotted lines with the Mg^{2+} as olive sphere.

These structural insights into the charged binding pocket of the Mid domain explain also the preference for guide-strand 5′ phosphate recognition in MjAgo. In contrary, MpAgo comprises a cleft that is lined with hydrophobic residues, responsible for anchoring the 5′ hydroxyl group of the RNA strand [53]. Here, the pocket is further confined by an ordered α-helix at the C terminus of the PIWI domain, which makes it sterically unlikely to accommodate a 5′ phosphate [53].

In contrast to TtAgo and hAgo2 binary complex structures, residues corresponding to the specificity loop within the Mid domain (residues 435-441) are disordered in the binary complex. In the apo form of MjAgo, these amino acids adopt a 3^{10}-helix at the N-terminal end of α12-helix.

The general thesis is that some Ago proteins such as hAgo2 can recognize different 5′-nucleotide identities through direct interactions with the specificity loop. Although a point mutation of a conserved Lys had no impact, a helix-break mutation in the 3^{10}-helix reduced slicing activity of MjAgo [144]. Furthermore, this mutant showed a significant change in the cleavage pattern using a guide strand that carries a pyrimidin base at the 5′-end. This implies the idea that in MjAgo this helix helps to position and stabilize the formed duplex within the binding pocket of the Mid domain [144].

Additionally, the loop linking the α17- and α18-helix positioned in the DNA-binding pocket of the PIWI domain becomes disordered upon protein-guide strand interaction.

In contrast to the 5′-end, residues 19 to 21 of the complex are well defined within the structure **(Figure 14)**. Here, the majority of the interactions occur between the final phosphate group and Tyr194, His213 and Tyr218, thus anchoring the 3′-end in the PAZ domain. As expected, there are no base-specific interactions between the DNA and the protein.

The last two residues of the DNA are almost aligned against the α6-helix in the PAZ domain with the bases thymine and guanine stabilized by stacking interactions between Glu246/ His213 and Phe45/ Tyr217, respectively. Furthermore, the final hydroxyl group of the guide strand forms also a hydrogen bond with Pro247 **(Figure 14)**. Here, cleavage activity assays revealed that point mutations of His213 and Tyr217 showed no impact, whereas mutations of Tyr194 and Glu246 completely abolished slicing activity [144].

Furthermore, the limited space observed in the structural architecture of the binding pocket implied the idea that a pyrimidine base might fit best into the PAZ domain **(Figure 14)**. This examination could be confirmed by *in vitro* cleavage assays, showing that MjAgo is less active with a guide DNA with a 3′-dG compared to a 3′-dT [144].

5.4.1.2. Conformational transitions in MjAgo on formation of the binary complex

By comparising the apo and the complex bound structures, several pronounced conformational rearrangements within the architecture of MjAgo can be observed **(Figure 15)**.

In order to fully accommodate the helical structure of the guide DNA, the insertion of the 5′-end at the junction of the Mid and PIWI domains leads to an opening of the nucleotide binding channel by 34 Å (distance is calculated between Cα-Leu 175 and Cα-Tyr 442). Furthermore, a significant movement is performed by the PAZ domain which rotates by 74° torwards the N domain around a 'hinge': thereby, the C-terminal 3^{10}-extension of α8-helix regains secondary structure and folds to an extendet α-helix with the disordered β8-strand adopts, together with β9 and β10 an anti-parallel beta sheet **(Figure 15)**.

This switch helix is comparable with a so called helix 7 in hAgo2 that has been shown to be responsible for introducing a helical kink at position 7 within the guide strand [70]. Here, introducing a helix break mutation at position 270 in combination with using a 5′-dT guide DNA completely abolished slicing activity of MjAgo [144]. This indicates that, in case a guide strand carries a pyrimidin base at the 5′-end, a structured α8-helix is essential to position and align the guide-target duplex within the protein.

Additional, residues building up the so called nucleotide specificity loop (residues 435-441) get disordered upon guide binding.
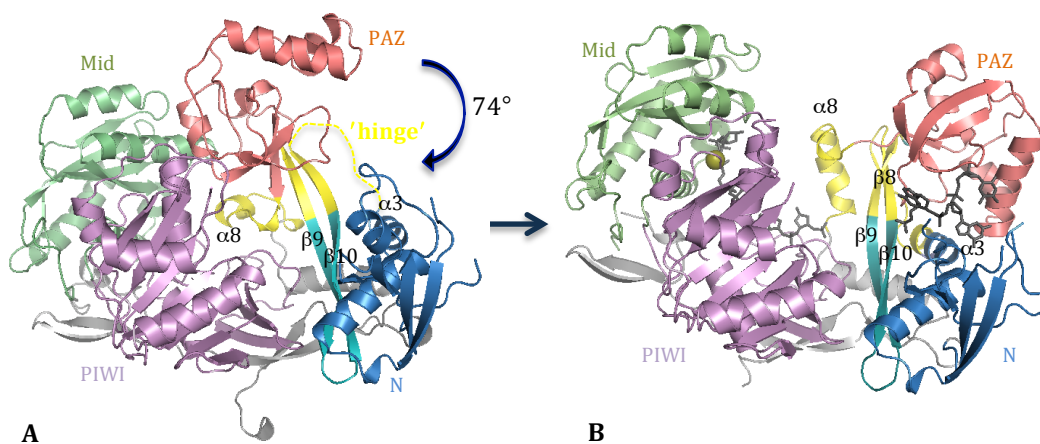

Figure 15. Conformational transition in MjAgo on formation of the binary complex. Crystal structures of **A)** the free state of MjAgo and **B)** in complex with a 5′-phosphorylated 21-mer guide DNA (shown as black stick representation). The conformational change of 74 ° performed by the PAZ domain around the hinge region (shown in yellow) is indicated by a blue arrow.

These mentioned conformational rearrangements induced by substrate binding are restricted to the hinge region and the disordering of the two loops in the Mid and PIWI-domain.

Compared to the overall TtAgo and hAgo2 binary structures, the major differences in the architecture of MjAgo can be found in the relative orientation of the well-conserved core domains as well as in the secondary structures building up the PAZ domain (**Figure 16**).
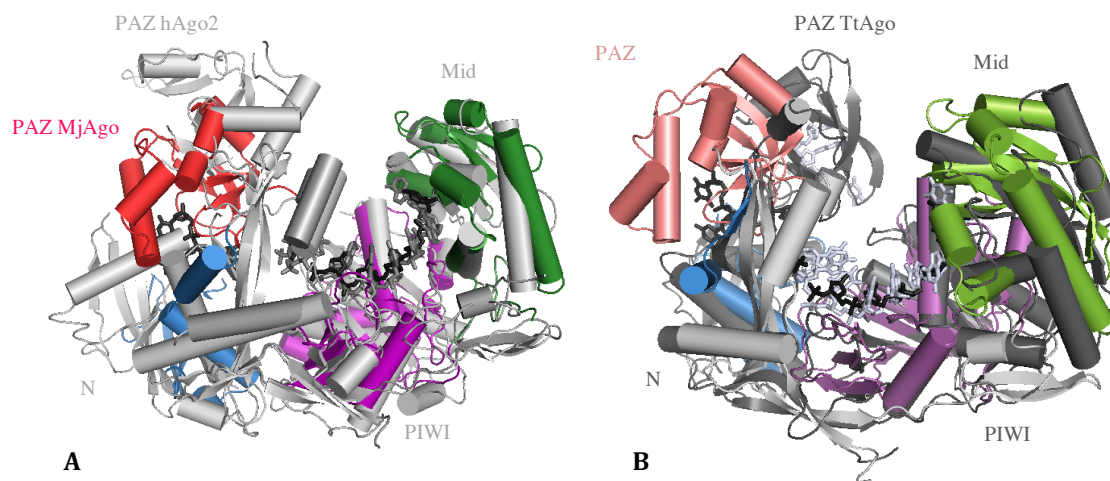
Figure 16: Comparison of guide recognition in the binary complex of MjAgo, hAgo2 and TtAgo with highlighting the varying position of the PAZ domains. Stereoview showing superimpositions of MjAgo (coloured according to Figure 7) in complex with a 21-base guide DNA (shown as sticks in black) and **A)** hAgo2 (PDB entry: 4OLA) with a heterogeneous mixture of guide RNAs (shown as sticks in grey) or **B)** TtAgo (PDB entry: 3DLH) bound to a 21-mer DNA guide (shown as sticks in bluewhite).

In MjAgo, the central beta-barrel in the PAZ domain is build up by four β-strands whereas in both TtAgo and hAgo2 this domain possess an additional beta strand. Additionally, hAgo2 contacts the guide strand through an extra ß-hairpin that can be also found in a rudimentary version in TtAgo. In contrast to MjAgo, these structural variations result in TtAgo and hAgo2 to a perpendicular binding position of the final bases of the guide strand.

Albeit these differences, the TtAgo, hAgo2 and MjAgo-PAZ domains alone can be superimposed quite well with an r.m.s.d of 2.1 - 5.5 Å, respectively **(Figure 16)**.

5.4.1.3. Evidence for a possible secondary nucleic acid binding channel

Analysis of the binary structure in more detail revealed next to the central cleft an additional binding channel between the N and PIWI domains. This cleft is 10 Å wide open and lined predominantly with positively charged amino acids **(Figure 17)**. Furthermore, we observed at this location electron density that could account for nucleotides, indicating that target strands might be traffic through this channel **(Figure 17)**.

A **B**
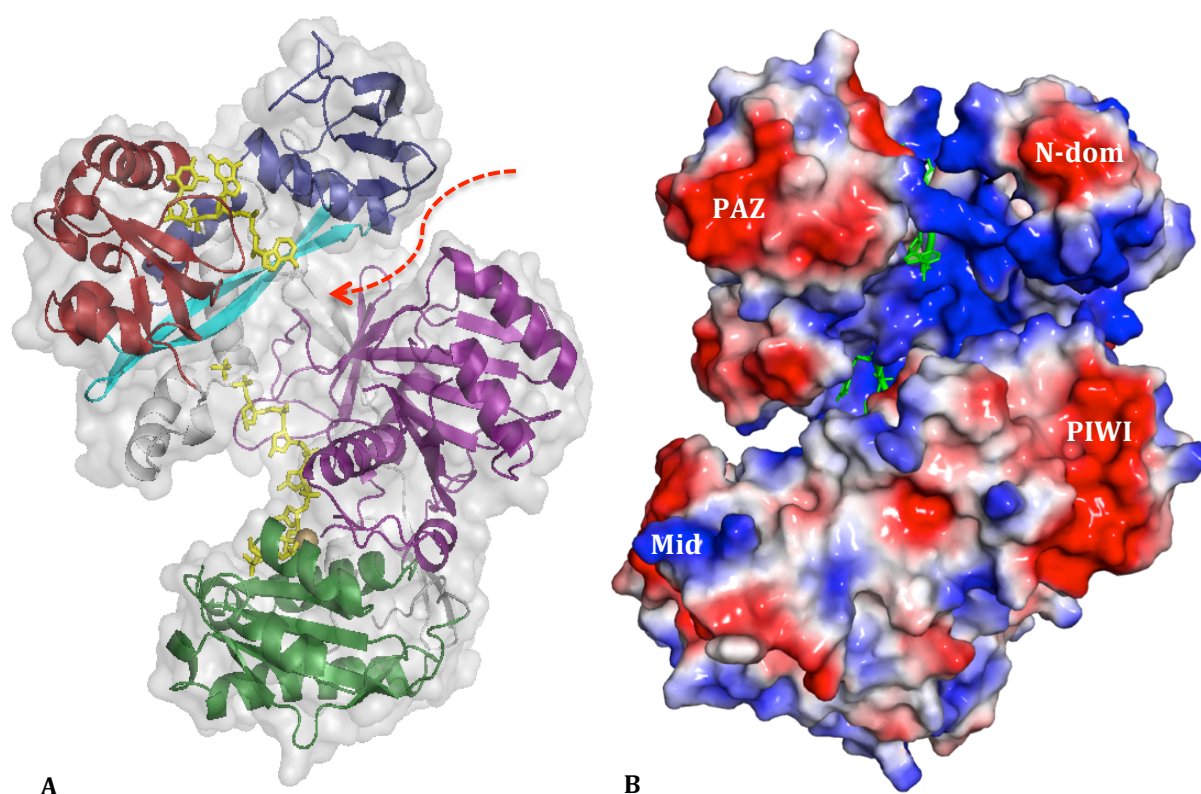
Figure 17: Potential secondary binding channel of MjAgo. A) Crystal structure of the binary complex of MjAgo, overlaid with the semi-transparent surface representation of the protein**.** The protein and DNA is coloured accoring to figure 13 with the red arrow pointing to the potential secondary binding cleft. **B)** Amino acid charges (blue=positive, red=negative) are mapped onto the surface of MjAgo with the guide DNA shown as green sticks.

This hypothesis was confirmed by introducing point mutations in supposed amino acid (Phe572, Gln574, Asn575) which might contribute to target binding. Here, cleavage activity was strongly dependent upon the 5′-nucleotide of the guide DNA which resulted in fully

active (in case of guanine) or complete lost of slicing activity (in case of thymine) [144].

Another reasonable function might be the assistance in harbouring long guide-target molecules since pAgos have been shown to successfully cleave long plasmid DNAs [56, 142]. However, to clearly determine the function of this channel and its possible involvement in target binding and slicer activity, further experiments would be required. These include for example the structural characterization of the ternary complex as well as mutation analysis of critical residues supposed to be involved in slicer activity.

5.4.2.   Crystallization of the ternary complex

Based on the ability of MjAgo to bind DNA/DNA hybrids [55], dsDNA was tested first in order to crystallize the ternary complex. However, adding dsDNA to low or high concentrated protein resulted in aggregation and precipitation of the prepared complex solution (data not shown).

Therefore, the ternary complex was formed in a 2-step binding process.

In order to avoid aggregation, the formation of the binary complex was performed as described in **chapter 5.3.1** *via* mixing MjAgo with a 5′-phosphorylated 21-mer guide DNA using a ratio of 1:2 of protein to DNA. To form the ternary complex, target DNA (**Figure 11**) was added to the concentrated binary mixture using a ratio of 1:1 and binding was performed at 37 °C for several hours. To assess whether the binary complex aggregates or oligomerize upon addition of the target DNA, the sample was centrifuged and analyzed by DLS.

Previous DLS measurements revealed that the binary complex of MjAgo has a hydrodynamic radius of 4.7 nm, respectively (**chapter 5.4.1**). After mixing the binary mixture with the target DNA, two peaks corresponding to a R_H of 5,86 nm and 0.86 nm are apparent (**Table 3**). Furthermore, the sample was monodisperse (polydispersity of 16.5 %) with no evidence of high-order aggregation. Additionally, the measured MW of 127 kDa corresponding to the binary complex increased to 213 kDa, thus indicating rearrangement and binding of the target strand to the protein.

	% PD	MW-R [kDA]	Radius [nm]
1	7.1	87	4.1
2	Multimodal	-	-
3	16.6	127	0.85 4.7
4	16.5	213	0.86 5.86

Table 3: DLS measurements of the apo, binary and ternary complex of MjAgo. 1) Native MjAgo at 7.5 mg/mL. **2)** Binary complex of MjAgo using a ratio of 1:2 (protein to DNA) after incubation at 37 °C for 3-4 h. **3)** Concentrated binary complex after adding DNA (ratio of 1:2 of protein to DNA) to diluted protein (1 mg/mL) with following incubation at RT over night. **4)** Concentrated ternary complex at approximately 7.5 mg/mL. Complex formation was performed by the addition of target DNA to the binary mixture (ratio of 1:1 of protein to DNA) with following incubation at 37 °C for 4 h. Sum of squares (SOS); polydispersity PD; Molecular weight (MW) hydrodynamic radius R_h.

Obtained ternary complex was subsequently screened for crystallization using the sitting drop vapour diffusion method. Therefore, a final drop size of 0.3 µL was utilized with a protein: reservoir ratio of 1:1 and crystallization plates were stored at 4 and 20 °C, respectively. After 2-3 days at 4 °C and 20 °C, serveral hit conditions comprising microcrystals were found **(Figure 18)**.

Figure 18: Protein crystals of the ternary complex of MjAgo. Crystals of MjAgo in complex with guide and target DNA using the sitting drop vapour diffusion method.

However, grown crystals were too small for X-ray diffraction and require further optimization. An ideal starting point for improving morphology and obtaining bigger quality protein crystals represents microseeding, by using smashed microcrystals as nucleation point for crystallization [115, 116].

Additionally, scaling up the crystallization process and varying the drop size can also be utilized in order to increase the size of protein crystals [117-119]. Furthermore, changing variables which alter the solubility - such as pH or precipitant type - have shown to have a great impact of the crystallization behaviour of the target protein [117, 120]. Also using different guide and target strands that vary in length and sequence composition can be tried out in order to obtain bigger quality crystals of the ternary complex.

6. Conclusion

Ago proteins are highly conserved endonucleases present throughout nature and are considered as the key enzymes in all RNA interference pathways. Despite low sequence similarity, all characterized Ago proteins from evolutionarily distant organisms show a strikingly deep conservation in their structural architecture.

A recently new characterized member is the Ago protein from the hyperthermophilic organism *M. jannaschii*, that exclusively shows a DNA guided DNA-slicing mechanism. Interestingly, MjAgo is able to cleave DNA strands out of dsDNA, resulting in multiple cleavage products [55].

In order to expand the existing knowledge, we determined the crystal structure of the free state of MjAgo. Moreover, we were able to crystallize MjAg in complex with a 5'-phosphorylated 21-mer guide DNA. This is, to our knowledge, the first Ago protein that could be crystallized in both its apo and complex bound forms.

Although derived from a hyperthermophilic organism, recombinantly expressed MjAgo was sensitive towards elevated temperatures, resulting in aggregation and precipitation of the protein. This result might be explained by the expression in *E. coli* as well as the high protein concentration required for crystallization. However, after optimization of the expression and purification strategy in combination with an increased NaCl concentration, MjAgo was able to concentrate, monodipserse and free of aggregates.

The performance of a thermal shift assay revealed a high stability of the protein over a broad pH and buffer range. However, testing the buffer comprising the highest melting temperature, no increase in stability and solubility of the target protein could be observed.

Purified protein was subsequently tested for crystallization and crystals appeared after a few days at 20° C. The size and quality of the crystals could be enhanced using microseeding and crystals diffracted to 3 Å, respectively. Furthermore, the diffraction data could significantly be improved by dehydration. Here, an increased PEG concentration in the crystallization condition showed to be most effective, extending the resolution from 3 to 2.35 Å, respectively.

Due to the low sequence conservation among Ago proteins, phases were determined experimentally. Since seleno-methionine expression failed to receive soluble protein, native MjAgo crystals were soaked with selected heavy metal atoms suitable for derivatizetion. Here, by testing several heavy atoms at different concentrations in various crystallization conditions, a Mercury-derivatived crystal was found to be most suitable for solving the phase problem of MjAgo.

As expected, MjAgo adopts the typical Ago fold, consisting of four predominant domains, the N, PAZ, Mid and PIWI. Despite low sequence similarity, the overall structural architecture of MjAgo is most similar to PfAgo, comprising a marginal difference in the relative positions of the two lobes as well as in the secondary structures building up the N and PAZ domains. Although used in the purification buffer, no divalent metal ion could be detected in the active site. However, the PIWI domain of MjAgo displays a high identity in structure, orientation and composition to other characterized Agos. Here, crucial highly conserved residues building up the catalytic tetrad could be identified within the catalytic pocket at equivalent positions as in other active Ago variants. However, to clearly determine catalytic residues involved in slicer activity and metal coordination, further experiments would be required.

In the binary complex, the incorporated 21mer guide DNA is traceable for nucleotides 1-6 and 19-21, whereas the 5'-end is anchored in the Mid domain and the 3'-end in the PAZ domain. Here, four critical, highly conserved residues which are responsible for the proper insertion of the final 5'-end phosphate within the Mid domain could be identified. Furthermore, the identity of the 5'-nucleotide and the consequential interactions with the Mid domain appear to influence the activity and the positioning of the target strand within the protein [144].

The magnesium ion located in the binding cleft within the Mid domain is coordinated by Qln479 and Lys483 as well as the carboxy-terminal carboxylate end (Ile713) of the PIWI domain.

The major contact points with the 3'-end of the guide strand occur predominantly between the final phosphate group and several PAZ domain residues, whereas the final bases are stabilized by stacking interactions, thus anchoring the DNA within the PAZ domain. As within other binary complex structures, no sequence-specific contacts are made to the guide DNA by the protein.

In contrast to other binary complexes, MjAgo lacks the fifth ß-strand building up the central beta-barrel of the PAZ domain. This difference leads, compared to MjAgo, to a perpendicular binding position of the 3′ base within hAgo2 and TtAgo. Furthermore, the PAZ domain of MjAgo is defined by an α-helix, providing a unique specificity for the 3'-nucleotide to the protein [144].

Due to the fact that it was possible to grow MjAgo crystals in its apo and complex bound forms, new information about Ago function could be obtained.

In particular, to incorporate the DNA within the protein, the PAZ domain rotates towards the N domain around a hinge region that is accompanied with an opening of the binding cleft between the Mid and PIWI domain. Additionally, secondary structural elements important for DNA binding and cleavage are rearranged upon formation of the binary complex. These include for example the folding of the hinge region into an extended α-helix and anti-parallel β-sheet as well as the disordering of the so-called nucleotide specificity loop.

Here, cleavge assays revealed that this switch helix seems to influence the position of the guide DNA strand within the protein (depending on the the 5'-nucleotide) [144]. Furthermore, the specificity loop in MjAgo appears not be involved in the readout of the 5'-base identity as observed in other Ago variants but stabilizes and influences the anchoring of a 5'-dT guide DNA within the Mid domain [144].

Moreover, an additional nucleic acid binding channel could be observed in the binary structure of MjAgo. This channel is located between the PIWI and the N-terminal domain and might be involved in target guidance as well as in harbouring of long guide-target duplexes. The importance of this channel could be demonstrated by introducing single mutations in relevant amino acids, resulting in abolishing of cleavge activity by using a particular 5'-terminal nucleotide of the guide strand [144]. However, to clearly determine the exact structure and function of the binding channel, further crystallization experiments would be required.

7. Materials and Methods

7.1. Materials

7.1.1.　　Chemicals

All chemicals unless stated otherwise, were obtained from the following companies:
Fluka (Neu-Ulm, GER), Merck (Darmstadt, GER), Sigma-Aldrich (Steinheim, GER), Serva (Heidelberg, GER), Roth (Karlsruhe, GER).

7.1.2.　　Media

LB medium	Peptone	10 g/L
	Yeast extract	5 g/L
	NaCl	5 g/L
	(Agar)	(20 g/L)
SOC medium	Peptone	20 g/L
	Yeast extract	5 g/L
	NaCl	10 mM
	$MgSO_4$	10 mM
	$MgCl_2$	10 mM
	KCl	2.5 mM
	Glucose	20 mM
2YT medium	Peptone	16 g/L
	Yeast extract	10 g/L
	NaCl	5 g/L
TB medium	Peptone	20 g/L
	Yeast extract	15 g/L
	NaCl	8 g/L
	Glycerol	8 ml/L
	Na_2HPO_4	2 g/L
	KH_2PO_4	1 g/L

48

Minimal medium M9	Na$_2$HPO$_4 \bullet$ 7 H$_2$O	48 mM
	KH$_2$PO$_4$	22 mM
	NaCl	9 mM
	NH$_4$Cl	19 mM
	MgSO$_4$	2 mM
	CaCl$_2$	0.1 mM
	Glucose	0.4 % (w/v)

7.1.3.   Protein standard

Roti-Mark STANDARD (14-200 kDa)   Carl Roth (Karlsruhe, GER)

7.1.4.   DNA oligos

All oligos were purified by HPLC and resuspended in 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 % glycerol and 5 mM MgCl$_2$ to reach a final concentration of 1 mM.

5′TGA GGT AGT AGG TTG TAT AGT   Microsynth (Wolfurt, Austria)

5′TAT ACA ACC TAC TAC CTC GT   Microsynth (Wolfurt, Austria)

5′TAG AGG TAC GTG CTG AGG CTT   Eurofins (Ebersberg, Germany)

7.1.5.   Expression plasmid

The plasmid pET 101 from Invitrogen, harbouring the full length of the Argnaute gene was kindly provided by our collaboration partner and used for recombinant protein expression.

7.1.6.   *E. coli* strains

Table 4: Bacterial strains used for protein expression.

Strain	Genotype	Company
BL21 (DE3)	F- *omp*T $_{hsd}$SB (r$_B$- m$_B$-) *gal dcm* (DE3)	New England Biolabs (Ipswich, USA)
SoluBL21	F- *omp*T $_{hsd}$SB (r$_B$- m$_B$-) *gal dcm* (DE3) *†*	Genlantis (San Diego, USA)
BL21 Star (DE3)	F- *omp*T $_{hsd}$SB (r$_B$- m$_B$-) *gal dcm, rne 131*(DE3)	Invitrogen (Carlsbad, USA)
B834 (DE3)	*gal hsdS$_B$ met ompT*	Novagen (Darmstadt, GER)

† Uncharacterized mutations in the SoluBL21 strain enhance soluble protein expression

7.1.7.   Instruments

Balances

Analytical Balance TE124S Sartorius	(Göttingen, GER)
Precision Balance BP3100 P Sartorius	(Göttingen, GER)
Precision Balance BP4100 P Sartorius	(Gottingen, GER)

Centrifuges

SIGMA 3-30K rotor 12150	SIGMA Laborzentrifugen (Osterode, GER)
SIGMA 3-30K rotor 12154	SIGMA Laborzentrifugen (Osterode, GER)
SIGMA 4K15 rotor 11150	SIGMA Laborzentrifugen (Osterode, GER)
SIGMA 4K15 rotor 13220	SIGMA Laborzentrifugen (Osterode, GER)
SIGMA 6K16 rotor 12500	SIGMA Laborzentrifugen (Osterode, GER)

Electrophoresis

Chamber and tray	Appligene (Watford, UK)
Digital Graphic Printer UP-D897	Sony (Minato, JP)

Electrophoresis Power Supply EPS 600	Pharmacia Biotech (Uppsala, SE)
Gel documentation system G:BOX	Syngene (Cambridge, US)
Mini PROTEAN® Cell BioRad	(Hercules, US)
PowerPAc Basic Power Supply	BioRad (Hercules, US)

Electroporation

2 mm electroporation cuvette	PeqLab (Erlangen, GER)
Gene Pulser with Pulse Controller	BioRad (Hercules, USA)

Protein purification

ÄKTAprime™ plus	GE Healthcare (Chalfont St. Giles, UK)
ÄKTApurifier™	GE Healthcare (Chalfont St. Giles, UK)
Superdex 200 16/60	GE Healthcare (Chalfont St. Giles, UK)
Ni Sepharose™ 6 Fast Flow	GE Healthcare (Chalfont St. Giles, UK)
Econo-Chromatography Column	BioRad (Hercules, US)

Protein crystallography

Art Robbins Instruments Intelli-Plates	Dunn Labortechnik (Asbach, DE)
Art Robbins Instruments Phoenix	Dunn Labortechnik (Asbach, DE)
Oryx8 Protein Crystallization Robot	Douglas Instruments (Berkshire, UK)
Protein Crystallization Screening Suites	QIAGEN (Hilden, DE)
Heavy Atom Screens	Hampton (Aliso Viejo, US)
CrystalCap HTTM für CryoLoopTM	Hampton (Aliso Viejo, US)
CrystalCap HTTM Vial	Hampton (Aliso Viejo, US)
CrystalWand MagneticTM	Hampton (Aliso Viejo, US)
Cooled Incubator Series 3000	RUMED®RubarthApparate (Laatzen, DE)
Magnetic Caps, Pins and Vials	Molecular Dimensions (Newmarket, UK)
MICORLAB® STARlet	Hamilton (Reno, US)
Micro Tool Box	Molecular Dimensions (Newmarket, UK)
Vial clamp	Molecular Dimensions (Newmarket, UK)
Mounted CryoLoopTM	Hampton (Aliso Viejo, US)
Foam dewar Spearlab	(San Francisco, US)
Siliconized Glass Cover Slides	Hampton (Aliso Viejo, US)

Glass Cover Slides	Hampton (Aliso Viejo, US)
SuperClear Pregreased 24 Well Plate	Crystalgen (New York, US)
Zoom Stereo Microscope SZX10/KL1500LCD	Olympus (Tokio, JP)

Additional instruments

DynaProNanoStar	Wyatt Technology Europe (Dernbach, GER)
Ultrasonic cell disruption	Branson Ultrasonics (Danbury, USA)
CFX Connect Real-Time PCR	BioRad (Hercules, US)
Incubator	Binder (Tuttlingen, DE)
Infors HT Multitron 2 Cell Shaker	INFORS HT (Bottmingen/Basel, CH)
InoLab pH 720 pH-Meter	WTW (Weilheim, GER)
Thermomixer comfort	Eppendorf (Hamburg, GER)
Vortex Genie 2	Scientific Industries (New York, USA)

7.2. Methods

7.2.1.   Cultivation and storage of *E. coli*

In order to prepare *E. coli* cells for long-term storage, a single colony from agar plate was inoculated over night at 37 °C in 2YT medium supplemented with the appropriate antibiotic. Cells were harvested at 5 000 x *g* for 10 min, resuspended in 1 ml 2YT medium containing 50 % (v/v) glycerol and stored at –80 °C.

7.2.2.   Isolation of DNA

Plasmid DNA was purified and isolated using the peq GOLD Cycle-PureKit (Peqlab, Erlangen, GER) as described in the instruction manual. DNA was eluted using 50 µL of ddH$_2$O and stored at –20 °C.

7.2.3. Transformation of electrocompetent *E. coli* cells

1 µL of plasmid DNA was added to 40 µL thawed, electrocompetent cells and electroporation was performed in a 2 mm cuvette at 2.5 kV. Cells were straight resuspended in 450 µL SOC medium and incubated at 37 °C for 1 h under shaking. Afterwards, 30 µL of cells were plated on selective LB Agar plates with following incubation overnight at 37 °C.

7.2.4. SDS polyacrylamide gel electrophoresis

Proteins were analyzed using SDS polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli *et al*., 1970 [121]. Therefore, protein samples were mixed with SDS loading dye and heated up to 95 °C for 5 min before loading on a 10 % SDS- PAGE (buffer formulations see table 5). Electrophoresis was performed at 160 V for 60 min. The SDS-gels were stained in Coomassie staining solution for at least 30 min and destained using 10 % acetic acid.

Table 5: SDS-PAGE buffer solutions.

	Separating gel (10 %)	Stacking gel (4 %)
ddH$_2$O	4.8 ml	3,75 ml
Acrylamide solution	2.5 ml	0,75 ml
Separating gel buffer	2.5 ml	-
Stacking gel buffer	-	1,5 ml
APS (20 % (v/v))	100 µL	100 µL
TEMED (1 % (v/v))	10 µL	10 µL
Acrylamide solution	Acrylamide	39 % (w/v)
	Bisacrylamide	1.2 % (w/v)
Separating gel buffer	Tris-HCl, pH 8.8	1.5 M
	SDS	0.4 % (w/v)
Stacking gel buffer	Tris-HCl, pH 6.8	0.5 M
	SDS	0.4 % (w/v)
Running Buffer	Tris-HCl, pH 8.2	25 mM
	Glycine	192 mM
	SDS	0.1 % (w/v)

Loading buffer	Tris-HCl, pH 6.8	60 mM
	Glycerol	30 % (v/v)
	Saccharose	10 % (w/v)
	SDS	5 % (w/v)
	Bromphenol blue	0.02 % (w/v)
	β-Mercaptoethanol	3 % (v/v)
	Glycine	192 mM
	SDS	0.1 % (w/v)
Staining solution	Coomassie Brilliant Blue R250	0.05 % (w/v)
	Isopropanol	25 % (v/v)
	Acetic acid	10 % (v/v)

7.2.5.   Determination of protein concentration

In order to determine the protein concentration, the absorbance at a wavelength of 280 nm was measured and calculated using the Lambert-Beer law. The required molar extinction coefficient ε was calculated using the ExPASy ProtParam tool (http://web.expasy.org/protparam) with the full length of protein sequence ε (MjAGO) = 122.8 cm L/mmol

7.2.6.   Protein production

The expression plasmid pET101, harboring the full length Ago protein (MJ_1234) from *M. jannaschii* (Gene ID: 1452130), was kindly provided by our cooperation partner Dr. Grohmann from the University of Braunschweig.

A fresh overnight culture of *E. coli* BL21 (DE3) pET101 was cultivated in 2YT medium containing the corresponding antibiotic (100 mg/L ampicillin) and incubated at 37 °C. The pre-culture was added to 1 L selective TB-medium (ratio 1:100) and grown while shaking at 37 °C until an optical density of 1 was reached. Cells were subsequently cooled down to 4 °C. The expression was induced by adding 1 mM isopropyl-β-D-thiogalactoside (IPTG) and continued under shaking at 37 °C for 17 h. Cells were harvested by centrifugation at 5 000 x *g* for 35 min and stored at –80 °C.

7.2.7. Protein purification

Overexpressed cells were lysed by sonication in buffer containing 50 mM Tris-HCl, pH 7.4, 500 mM NaCl, 10 % glycerol, 5 mM $MgCl_2$ and 20 mM imidazole (lysis buffer) supplemented with a complete protease inhibitor cocktail (Roche). The cell debris were removed by centrifugation for 45 min at 21 000 rpm. The supernant was added to Ni Sepharose™ 6 Fast Flow beads, which were equilibrated in lysis buffer and incubated at room temperature on a shaker at low speed for between 1 to 2 h. The sample was subsequently loaded on gravity-flow column and contaminating proteins were removed by washing the resin extensively with twenty column volumes of lysis buffer. The recombinant protein was eluted with 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 % glycerol, 5 mM MgCl and 250 mM imidazole and supplemented with Tobacco Etch Virus (TEV) protease to reach a final ratio of 1:50. The resulting mixture was dialyzed against 50 mM Tris-HCl, pH 7.4 containing 300 mM NaCl, 10 % glycerol and 5 mM $MgCl_2$ (dialysis buffer) for 4 h at room temperature. For His-tag removal, the sample was incubated with equilibrated Ni resin in dialysis buffer for 30 min and the non-bound fraction was collected. Final purification was achieved by size-exclusion chromatography on a Superdex 200 16/60 performed in 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 % glycerol and 5 mM $MgCl_2$. The fractions corresponding to the major elution peak were combined and concentrated at 2 000 x g using an Amicon Ultra Centrifugal Filter Device with a molecular weight cut-off of 30 kDa (Millipore, Billerica, USA) to reach a protein concentration of between 7 and 8 mg/mL. The protein concentration was measured using the theoretical extinction coefficient ε = 122,84 M-1 cm-1 by UV/V is spectroscopy at 280 nm and concentrated protein was stored at –80 °C for further use.

7.2.8. Selenomethionine expression

The expression of SeMet-labeled protein was performed as described by van Duyne et al. [122].

Therefore, the expression plasmid pET101, harboring the full length of the Ago gene of *M. jannaschii* was transformed in the following *E. coli* strains: BL21 (DE3), BL21 star (DE3), SoluBL21 and B834 (DE3). From resulting colonies, a fresh overnight culture was prepared

in selective 2YT or minimal medium M9 and incubated under shaking at 37 °C. The pre-culture was added in a ratio of 1:100 to the selective minimal medium M9, enriched with vitamins as well as trace elements and incubated shaking at 37 °C until at an optical density of 0.6 was reached. After adding a mixture of amino acids containing 0.5 g/L SeMet (Calbiochem, San Diego, USA), leucine, isoleucine, valine and 1.0 g/L lysine, threonine and phenylalanine, cells were further incubated for 15 min at 37 °C under shaking.

The expression was induced *via* adding of 1 mM IPTG and the cell cultures were incubated at 20 °C or 37 °C over night while shaking. Cells were harvested at 5 000 x *g* for 30 min and stored at -80°C before further use.

7.2.9.   Thermal shift assay (Thermoflour assay)

7.2.9.1.   Buffer screen

In order to improve protein stability, the thermofluor buffer screen of EMBLE (Hamburg, GER) was utilized and performed as described in the instruction manual.

Therefore, 2 µL of protein (2.5 mg/mL, 30 µM) was added to each buffer formulation and mixed with 2 µL of a 1:80 dilution of SYPRO Orange dye (Invitrogen) in water. Plates were centrifuged at 2 000 x *g* for 1 min and the fluorescence was monitored in regular 0.5 °C intervals over a temperature range from 25 °C to 95 °C using a RT-PCR machine (excitation and emission at 470 and 569 nm).

7.2.9.2.   DNA binding

Recombinant MjAgo at 1.5 mg/mL was mixed with DNA in a ratio of 1:2 (DNA in excess) and incubated at increasing temperatures (35 °C to 60 °C) for 15 min. Samples were spun down at 21 000 x *g* for 10 min and 20 µM of protein was mixed with 2 µL of a 1:80 dilution of SYPRO Orange dye (Invitrogen) in water and buffer containing 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 % glycerol, 5 mM $MgCl_2$ to reach a final volume of 20 µL. Fluorescence was monitored in regular 0.5 °C intervals over a temperature range from 25 °C to 95 °C using a RT- PCR machine (excitation and emission at 470 and 569 nm).

7.2.10.   Dynamic ligth scattering

Pure, homogeneous protein is the most challenging and crucial factor for successful protein crystallization. Therefore, protein samples should be analysed precisely before setting up crystallization trails. Next to size-exclusion chromatography and circular dichroism (CD), dynamic light scattering (DLS) represents a powerful tool to measure the size distribution, polydispersity as well as the aggregation state of proteins in solution (96).

In general DLS uses the laser light, which is scattered by molecules or particles in solution. Due to Brownian motion, molecules are in motion and fluctuations of the scattering intensity can be observed. Based on measuring in very short intervals and the fact that large molecules move slower than small molecules, a defined correlation function can be calculated. After working out the diffusion coefficient (D), the hydrodynamic radius (R_H) of the particles and molecules can be calculated using the Stokes-Einstein equation (Equation 1) [122, 123].

$$R_H = \frac{K_B \cdot T}{6\pi\eta D}$$

Equation 1: Stokes-Einstein equation: D (diffusion coefficient), K_B (Boltzmann constant), T (temperature), η (solvent viscosity) and R_H (hydrodynamic radius).

The typical DLS experiment takes less than a few minutes, requires no sample information, is non-destructive and desires only a minimum amount of protein (typically < 1 mg). Additionally, purified protein can be tested in various buffers and analysed in the presence of ligands, inhibitors or cofactors. [124]. Furtermore, the correlation between the enhanced ability to form crystals and monodisperse samples as measured by DLS, could clearly be demonstrated by crystallization of several proteins [95, 125, 126]. Hence, DLS was used to analyse the quality and homogeneity of concentrated protein prior to setting up crystallization trails. Therefore, samples were centrifuged at 21 000 x *g* for 10 min and 20 µL of the supernant was selected and transferred into a microcuvette. Afterwards, the protein solution was recovered and utilized for crystallization.

7.2.11.   Protein crystallization

Initial crystallization screens were realized using the sitting- drop vapor diffusion method with crystallization buffer kits from Qiagen (Hilden, DE) using a final drop size of 0.3 μL with a protein:reservoir ratio of 1:1. Crystals appeared within 3-5 days at 20 °C in 0.1 M tri-sodium citrate pH 5.6, 20 % isopropanol, 20 % PEG 4000 as well as 0.2 M tri-sodium citrate, 0.1 M HEPES-HCl pH 7.5 and 20 % isopropanol. In order to optimize crystallization, native crystals with a final drop size of 2 μL were grown using the hanging- drop vapor diffusion method, with a 1:1 (v/v) ratio of protein to mother liquor (0.5 ml well volume).

The shape and size of the crystals were significantly improved by microseeding. Therefore, 0.5 μL of seed stock was added to a 2 μL drop containing equal amounts of protein and reservoir and the crystals appeared after 3 days at 20 °C in 0.1 M tri-sodium citrate, 10 % isopropanol, 8 % PEG 4000 as well as 0.1 M HEPES-HCl, pH 7.5, 12 % isopropanol and 0.05 M tri-sodium citrate. All crystals were soaked in the cryosolution containing 30 % (v/v) MPD in addition to the mother liquor before freezing and storing in liquid nitrogen.

For dehydration, crystals were soaked in dehydration solution containing 8 % PEG 4000, 10 % isopropanol and 100 mM tri-sodium citrate enriched with either 20 % PEG 4000, 15 % MPD, 15 % ethylenglycol or 15 % glycerol and equilibrated against 1 ml reservoir for 17 h at 4 °C. All crystals were cryoprotected by the addition of 3 μL reservoir solution supplemented with 50 % (v/v) MPD and incubated for at least 2 min prior to flash- freezing in liquid nitrogen.

7.2.12.   Crystallization of the binary and ternary complex

For the formation and crystallization of the binary complex, MjAgo was concentrated to 1 mg/mL or between 8-9 mg/mL in either 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 % glycerol, 5 mM MgCl$_2$, or 25 mM Tris-HCl, pH 7.4, 150 NaCl, 10 % glycerol and 5 mM MgCl$_2$. The higher concentrated protein solution was mixed with several guide DNAs in a molar ratio at 1:1, 1:1.2 or 1:2 (DNA in excess) and incubated at 37 °C for 4 h, whereas the lower concentrated MjAgo solution was mixed with guide DNA at a molar ratio of 1:2 and incubated at RT over night. After the binary complex formation, possible aggregates were

removed via centrifugation at 21 000 rpm for 15 min and samples were concentrated to 8-9 mg/mL before setting up crytsallization trails.

For the crystallization of the ternary-complex, MjAgo at a concentration of 1 mg/mL was mixed with a 5′-phosphorylated 21-mer guide DNA at a molar ratio of 1:2 and incubated over night at RT. The binary complex was concentrated to 8–9 mg/mL and the target DNA was added at a molar ratio of 1:1. The ternary complex was formed by incubating the sample for 4 h at 37 °C. Initial crystallization screens for the binary as well as for the ternary complex were set up in 96-well plates by sitting drop vapour diffusion with a final drop size of 0.3 µL containing a protein:reservoir ratio of 1:1. Crystallization plates were stored at 4 °C, 20 °C and 37 °C respectively.

7.2.13.   Heavy metal derivatization

Due the amino acid composition of the protein, following heavy metals were selected:
Ammonium tetrachloroplatinate (II) ($(NH_4)_2PtCl_4$), Gold(III) chloride ($AuCl_3$,) Potassium hexachloroplatinate (IV) (K_2PtCl_6)**,** Silver nitrate ($AgNO_3$), Mercury (II) acetate ($Hg(OOCCH_3)_2$), Mercury (II) chloride ($HgCl_2$) and Yttriumchlorid ($YtCl_3$).

Native crystals were grown by the hanging drop vapor diffusion as described in chapter 7.2.11 using following crystallization conditions: 12 % isopropanol, 0.05 M tri-sodium citrate in either 0.1 M HEPES-HCl, pH 7.5, HEPES-HCl pH 6.5 or MES-NaOH, pH 6.5 as well as 8 % PEG 4000, 0.1 M tri-sodium citrate with 10 % isopropanol. For derivatization, grains of heavy atoms were either added directly to the crystallization drop containing MjAgo crystals or a freshly prepared soaking solution (100 mM in ddH_2O) was added to reach a final drop concentration ranging from 1 to 10 mM. The crystals with the heavy atoms were incubated within the drop at 4 °C for between 4 and 16 h, after which the crystals were back-soaked into mother liquor for around 1 to 2 min. All crystals were cryoprotected by the addition of reservoir solution supplemented with either 50 % MPD or 50 % ethylenglycol and incubated for 1 min prior to flash-freezing in liquid nitrogen.

7.2.14.   X-ray data collection and structure determination

Datasets were recorded using synchrotron radiation at beamlines of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France (ID23-1, ID23-2, ID29), the Swiss Light Source (SLS) Villigen Switzerland (PX I) and the Deutsches Elektronen Synchrotron (DESY), Hamburg, Germany (PETTRA IV). The obtained diffraction data were further processed with XDS [125] and DIALS [128] and crystals of MjAgo belonged to the space group $P2_12_12_1$ whereas crystals of the binary complex belonged to the space group $P4_32_12$, respectively. Experimental phasing was perfomed with CRANK2 [129-134] utilizing the collected data set on a mercury derivatized crystal. Initial Model building was carried out manually using COOT [127], before refinement in REFMAC5 [132], PHENIX [130, 131] or BUSTER [138] was achieved against the native data. The model has further been improved by successive rounds of manual building and refinement and the amounts and coordinates of TLS groups were determined utilizing the TLSMD server [139, 140].

The crystal structure of the binary complex was determined by molecular replacement using the program PHASER [141]. Further model building was carried out in COOT and the obtained structure was refined with REFMAC5/PHENIX/BUSTER. Structure validation was performed utilizing MOLPROBITY [142]. Diffraction data and refinement statistics are shown in **Table 2**. All structural figures were created with PyMol (Delano Scientific) whereas structural superpositions were performed with the program cealign.

8. Abbreviations

%	percent
Å	angstrom
A	absorbance
Ago	Argonaute
AMP	ampicillin
AS	amino acids
bp	base pair
BSA	bovine serum albumin
°C	degree Celsius
CRISPR	clustered regulatory interspaced short palindromic repeats
C. elegans	*Caenorhabditis elegans*
Da	Dalton
ddH_2O	double distilled water
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acid
ssDNA	single stranded deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
et al.	*et alia*
h	hour
HCl	hydrochloric acid
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
His_6	hexahistidin
HPLC	high-performance liquid chromatography

IMAC	Immobilized-metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kbp	kilo base pairs
kDa	kilo Dalton
L	liter
L1	linker 1
L2	linker 2
LB	Lysogeny broth
μ	micro
M	molar
mA	milliampere
MES	2-(N-morpholino) ethanesulfonic acid
min	minute
Middle	Mid
miRNA	microRNA
mRNA	messenger RNA
M. jannaschii	*Methanocaldococcus jannaschii*
MPD	2-Methyl-2,4-pentanediol
nt	nucleotide
OD_{600}	optical density at 600 nm wavelength
PAZ	PIWI Argonaute and Zwille
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	polydispersity
PDB	Protein Data Bank
PEG	polyethylene glycol
P. furiosus	*Pyrococcus furiosus*

pH	potentia hydrogenii
piRNA	PIWI-interacting RNA
PIWI	P-element-induced wimpy testis
pre-miRNA	precursor miRNA
R_{free}	free R-factor
R_h	hydrodynamic radius
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
RT	room temperature
R_{work}	crystallographic R-factor
SAD	single-wavelength anomalous diffraction
SDS	sodium dodecyl sulfate
sec	second
SeMet	Selenomethionine
siRNA	small interfering RNA
ss	single strand (DNA/RNA)
SOC	super optimal broth with catabolite repression
TB	terrific broth medium
TEV	tobacco etch virus
Tris	tris(hydroxymethyl)aminomethane
UV/Vis	ultraviolet/visible
v/v	volume per volume
w/v	weight per volume
MW	molecular weight (g/mol or Da)
ε	molar extinction coefficient
λ	wavelength

9. References

(1) C. Napoli, C. Lemieux, R. Jorgensen, *Plant Cell* **1990**, 2, 279–289.

(2) S. Covey, N. Al-Kaff, A. Lángara, D. Turner, *Nature* **1977**, 385, 781–782.

(3) J. A. Lindbo, L. Silva-Rosales, W. M. Proebsting, W. G. Dougherty, *Plant Cell* **1993**, 5, 1749–1759.

(4) A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, *Nature* **1989**, 391, 806–811.

(5) E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **2001**, 409, 363–366.

(6) A. Grishok, A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, C. C. Mello, *Cell* **2001**, 106, 23–34.

(8) Y. S. Lee, K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer, R. W. Carthew, *Cell* **2004**, 117, 69–81.

(9) S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **2000**, 404, 293–296.

(10) P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **2000**, 101, 25–33.

(11) Z. Mourelatos, J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Mann, G. Dreyfuss, *Genes Dev.* **2002**, 16, 720–728.

(12) G. Meister, M. Landthaler, L. Peters, P. Y. Chen, H. Urlaub, R. Lührmann, T. Tuschl, *Curr. Biol.* **2005**, 15, 2149–2155.

(13) J. Martinez, A. Patkaniowska, H. Urlaub, R. Lührmann, T. Tuschl, *Cell* **2002**, 110, 563–574.

(14) R. F. Ketting, *Adv. Exp. Med. Biol.* **2011**, 700, 1–14.

(15) L. Joshua-Tor, G. J. Hannon, *Cold Spring Harb. Perspect. Biol. 3,* **2011**, a003772.

(16) K. Tiemann, J. J. Rossi, *EMBO Mol. Med.* **2009**, 1, 142–151.

(17) K. Gavrilov, W. M. Saltzman, *Yale J. Biol. Med.* **2012**, 85, 187–200.

(18) G. Hutvagner, P. D. Zamore, *Curr. Opin. Genet. Dev.* **2002**, 12, 225–232.

(19) B. R. Cullen, *Nat. Immunol.* **2002**, 3, 597–599.

(20) M. Ghildiyal, P. D. Zamore, *Nat Rev Genet* **2009**, 10, 94–108.

(21) A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M Zavolan, T. Tuschl, *Nature* **2006**, 442, 203-7.

(22) A. Girard, R. Sachidanandam, G. J. Hannon, M. A. A. Carmell, *Nature* **2006**, 442, 199-202.

(23) N. C. Lau, A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel, R. E. Kingstone, *Science* **2006**, 313, 363-7.

(24) J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, *Cell* **2007**, 128, 1089–1103.

(25) L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, M. C. Siomi, *Science* **2007**, 315, 1587–1590.

(26) C. C. Mello, D. Jr. Conte, *Nature* **2004**, 431, 338–342.

(27) Z. Lippman, R. Martienssen, *Nature* **2004**, 431, 364–370.

(28) D. E. Golden, V. R. Gerbasi, E. J. Sontheimer, *Mol. Cell* **2008**, 31, 309–312.

(29) D. C. Swarts, K. Makarova, Y. Wang, K. Nakanishi, R. F. Ketting, E. V. Koonin, D. J. Patel, J.van der

Oost, *Nat. Struct. Mol. Biol.* **2014**, 21, 743–753.

(30) S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, T. Tuschl, *Nature* **2001**, 411, 494–498.

(31) S. M. Elbashir, J. Martinez, A. Patkaniowska, W. Lendeckel, T. Tuschl, *EMBO J.* **2001**, 20, 6877–6888.

(32) F. V. Rivas, N. H. Tolia, J. J. Song, J. P. Aragon, J. Liu, G. J. Hannon, L. Joshua-Tor, *Nat. Struct. Mol. Biol.* **2005**, 12, 340–349.

(33) I. J. Macrae, E. Ma, M. Zhou, C. V. Robinson, J. A. Doudna, *Proc. Natl. Acad. Sci.* **2008**, 105, 512–517.

(34) Y. Tomari, P. D. Zamore, *Genes Dev.* **2005**, 19, 517–529.

(35) Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim, V. N. Kim, *Nature* **2003**, 425, 415–419.

(36) M. T. Bohnsack, K. Czaplinski, D. Gorlich, *RNA* **2004**, 10, 185–191.

(37) E. Lund, S. Guttinger, A. Calado, J. E. Dahlberg, U. Kutay, *Science* **2004**, 303, 95–98.

(38) R. Yi, Y. Qin, I. G. Macara, B. R. Cullen, *Genes Dev.* **2003**, 17, 3011–3016.

(39) D. P. Bartel, *Cell* **2004**, 116, 281–297.

(40) C. Matranga, Y. Tomari, C. Shin, D. Bartel, P. Zamore, *Cell* **2005**, 123, 607–620.

(41) R. I. Gregory, T. P. Chendrimada, N. Cooch, R. Shiekhattar, *Cell* **2005**, 123, 631–640.

(42) J. S. Parker, E. A. Parizotto, M. Wang, S. M. Roe, D. Barford, *Mol. Cell* **2009**, 33, 204–214.

(43) J. Liu, M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J. J. Song, S. M. Hammond, L. Joshua-Tor, G. J. Hannon, *Science* **2004**, 305, 1437–1441.

(44) N. H. Tolia, L. Joshua-Tor, L. *Nat. Chem. Biol.* **2007**, 3, 36–43.

(45) C. D. Kuhn, L. Joshua-Tor, *Trends in Biochemical Sciences* **2013**, 38, 263-271.

(46) Y. Tomari, P. D. Zamore, *Genes Dev.* **2005,** 19, 517–529.

(47) N. Rajewsky, *Nat. Genet.* **2006**, 38 Suppl:S8–13.

(48) B. P. Lewis, C. B. Burge, D. P. Bartel, *Cell* **2005**, 120, 15–20.

(49) W. Filipowicz, S. N. Bhattacharyya, N. Sonenberg, *Nat. Rev. Genet.* **2008**, 9, 102-14.

(50) M. Jinek, J. A. Doudna, *Nature* **2009**, 457, 405-412.

(51) R. W. Carthew, E. J. Sontheimer, *Cell* **2009**, 136, 642–655.

(52) K. S. Makarova, Y. I. Wolf, J. van der Oost, E. V. Koonin, *Biol. Direct* **2009**, 4, 29.

(53) E. Kaya, K. W. Doxzen, K. R. Knoll, R. C. Wilson, S. C. Strutt, P. J. Kranzusch, J. A. Doudna, *Proc. Natl. Acad. Sci. USA* **2016**, 113, 4057-4062.

(54) D. C. Swarts, J. W. Hegge, I. Hinojo, M. Shiimori, M. A. Ellis, J. Dumrongkulraksa, R. M. Terns, M. P. Terns, J. van der Oost, *Nucleic Acids Research* **2015,** 43, 5120–5129.

(55) A. Zander, P. Holzmeister, D. Klose, P. Tinnefeld, D. Grohmann, *RNA Biol.* **2014**, 11, 45–56.

(56) I. Olovnikov, K. Chan, R. Sachidanandam, D. K. Newman, A. A. Aravin, *Mol. Cell* **2013**, 51, 594–605.

(57) F. Gao, X. Z. Shen, F. Jiang, Y. Wu, C. Han, *Nat. Biotechnol.* **2015,** 7, 768-773.

(58) C. Ender, G. Meister, *Journal of Plant Science* **2010**, 123, 1819-1823.

(59) K. Bohmert, I. Camus, C. Bellini, D. Bouchez, M. Caboche, C, Benning, C. *EMBO J.* **1998**, 17, 170-180.

(60) N. H. Tolia, L. Joshua-Tor, *Nat. Chem. Biol.* **2007**, 3, 36-43.

(61) T. Sasaki, A. Shiohama, S. Minoshima, N. Shimizu, *Genomics* **2003**, 82, 323-330.

(62) G. Hutvagner, M. J. Simard, Nat. *Rev. Mol. Cell Biol.* **2008**, 9, 22–32.

(63)   M. A. Carmell, Z. Xuan, M. Q. Zhang, G. J. Hannon, *Genes Dev.* **2002**, 16, 2733-2742.

(64)   G. Hutvagner, M. J. Simard MJ, *Nat Rev Mol Cell Biol.* **2007**, 9, 22-32.

(65)   G. Meister, M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng T. Tuschl, *Mol. Cell* **2004**, 15, 185-197.

(66)   L. Jidong, M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J. Song, S. M. Hammond, L. Joshua-Tor, G. J Hannon, *Science* **2004**, 305, 1437-41.

(67)   J. J. Song, S. K. Smith, G. J. Hannon, L. Joshua-Tor, L., *Science* **2004**, 305, 1434–1437.

(68)   Y. Wang, G. Sheng, S. Juranek, T. Tuschl, D. J. Patel, *Nature* **2008**, 456, 209-13.

(69)   Y. Wang, S. Juranek, G. Sheng, T. Tuschl, D. J. Patel, *Nature* **2008**, 456, 921-6.

(70)   N. T. Schirle, I. J. MacRae, *Science* **2012**, 336, 1037–1040.

(71)   K. Nakanishi, D. E. Weinberg, D. P. Bartel, D. J. Patel, *Nature* **2012**, 486, 368-374.

(72)   E. Elkayam, C. D. Kuhn, A. Tocilj, A. D. Haase, E. M. Greene, G. J. Hannon, L. Joshua-Tor, *Cell* **2012**, 150, 100–110.

(73)   J. Hauptmann, A. Dueck, S. Harlander, J. Pfaff, R. Merkl, G. Meister, *Nat. Struct. Mol. Biol.* **2013**, 7, 814-817.

(74)   P. B. Kwak, Y. Tomari, *Nat. Struct. Mol. Biol.* **2012**, 19, 145–151.

(75)   J. K. Hur, M. K. Zinchenko, S. Djuranovic, R. Green, *J. Biol. Chem.* **2013**, 288, 7829–7840.

(76)   B. Simon, J. P. Kirkpatrick, S. Eckhardt, M. Reter, E. A. Rocha, M. A. Andrade-Navarro, P. Sehr, R. S. Pillai, T. Carlomagno, *Structure* **2011**, 19, 172-80.

(77)   J. S. Parker, S. M. Roe, D. Barford, *EMBO J.* **2004**, 23, 4727–4737.

(78)   J. S. Parker, E. A. Parizotto, M. Wang, S. M. Roe, D. Barford, *Mol. Cell* **2009**, 33, 204–214.

(79)   M. Jinek, J. A. Doudna, *Nature* **2009**, 457, 405–412.

(80)   M. Nowotny, S. A. Gaidamakov, R. J. Crouch, W. Yang, *Cell* **2005**, 121, 1005–1016.

(81)   G. Sheng, H. Zhao, J. Wang, Y. Rao, W. Tian, D. C. Swarts, J. van der Oost, D. J. Patel,. *Proc. Natl. Acad. Sci. USA* **2014**, 111, 652–657.

(82)   R. A. Neumuller, J. Betschinger, A. Fischer, N. Bushati, I. Poernbacher, K. Mechtler, S. M. Cohen, J. A. Knoblich, *Nature* **2008**, 454, 241–245.

(83)   F. Frank, N. Sonnenberg, B. Nagar, *Nature* **2010**, 465, 818–822.

(84)   J. S. Parker, S. M. Roe, D. Barford, *Nature* **2005**, 434, 663–666.

(85)   A. Boland, F. Tritschler, S. Heimstadt, E. Izaurralde, O. Weichenrieder, *EMBO Rep.* **2010**, 11, 522–527.

(86)   A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nat. Struct. Mol. Biol.* **2004**, 11, 576–577.

(87)   J. B. Ma, Y. R. Yuan, G. Meister, Y. Pei, T. Tuschl, D. J. Patel, *Nature* **2005**, 434, 666–670.

(88)   E. Elkayam, C. D. Kuhn, A. Tocilj, A. D. Haase, E. M. Greene, G. J. Hannon, and L. Joshua-Tor, *Cell* **2012**, 150, 100–110.

(89)   W. J. Jones, J. A. Leigh, F. Mayer, C. R. Woese, R. S. Wolfe, *Archives of Microbiology* **1983**, 136, 254-261.

(90)   C. J. Bult, O. White, G. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, *et. al., Science* **1996**, 273, 1058–1073.

(91)   D. Baulcombe, *Nature* **2004**, 431, 356–363.

(92)   W. Filipowicz, *Cell* **2005,** 122, 17–20.

(93)   T. M. Rana, *Nature Rev. Mol. Cell Biol.* **2007,** 8, 23–36.

(94)   D. H. Kim, J. J. Rossi, *Nature Rev. Genet.* **2007**, 8, 173–184.

(95)   A. De Fougerolles, H. P. Vornlocher, J. Maraganore, J. Lieberman, *Nature Rev. Drug. Discov.* **2007**, 6, 443–453.

(96)   G. E. O. Borgstahl, *Methods Mol. Biol.* **2007**, 363, 109-29.

(97)   S. Boivin, S. Kozak, R. Meijers, *Protein Expression and Purification* **2013,** 9, 192–206.

(98)   U. B. Ericsson, M. Hallberg, G. T. DeTitta, N. Dekker, P. Nordlund, *Analytical Biochemistry* **2006**, 357, 289–298.

(99)   D. M. Salunke, B. Veerapandian, R. Kondandapani, M. Vijayan, *Acta Cryst.* **1985**, 41, 431–436.

(100)   M. Frey, *Acta Cryst.* **1994**, 50, 663–666.

(101)   C. Abergel, *Acta Cryst.* **2004**, 60, 1413–1416.

(102)   B. Heras, J. L. Martin, *Acta Cryst.* **2005**, 61, 1173–1180.

(103)   B. Heras, M. A. Edeling, K. A. Byriel, A. Jones, S. Raina, J. L. Martin, *Structure* **2003**, 11, 139–145.

(104)   C. Abergel, *Acta Crystallogr. D Biol. Crystallogr*. **2013**, 69, 2167–2173.

(105)   J. Agniswamy, M. G. Joyce, C. H. Hammer , P. D. Sun, *Acta Cryst.* **2008**, 64, 354–367.

(106)   M. G. Joyce, S. Radaev, P. D. Sun, *Acta Cryst.* **2010**, 66, 358–365.

(107)   E. Garman, J. W. Murray, *Acta Crystallogr. D Biol. Crystallogr*. **2003**, 59, 1903-1913.

(108)   P. D. Sun, S. Radaev, *Acta Crystallogr. D Biol. Crystallogr*. **2002**, 58, 1099-1103.

(109)   P. D. Sun, S. Radaev, M. Kattah, *Acta Crystallogr. D Biol. Crystallogr*. **2002**, 58, 1092-1098.

(110)   S. R. Jordan, T. V. Whitcombe, J. M. Berg, C. O. Pabo, *Science* **1985**, 230, 1383-1385.

(111)   S. Rhee, R. G. Martin, J. L. Rosner, D. R. Davies, *Proc. Natl. Acad. Sci. USA* **1998,** 95, 10413-10418.

(112)   W. Yang, J. Y. Lee, M. Nowotny, *Mol. Cell* **2006**, 22, 5-13.

(113)   P. A. Rice, S. Yang, K. Mizuuchi, H A. Nash, *Cell* **1996**, 87, 1295-1306.

(114)   N. E. Chayen, E. Saridakis, *Nature Methods* **2008**, 5, 147-153.

(115)   A. D. Arcy, F. Villard, M. Marsh, *Acta Crystallogr. D Biol. Crystallogr.* **2007**, 63, 550–554.

(116)   G. Obmolova,T. J. Malia, A. Teplyakov, R. Sweet, G. L. Gilliland, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 927–933.

(117)   J. B. Glusker, *Crystal Growth & Design* **2003**, 3, 109–109.

(118)   E. L. Forsythe, D. L. Maxwell, M. Pusey, *Acta Crystallogr. D Biol. Crystallogr.* **2002**, 58, 1601–1605.

(119)   T. S. Walter *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* **2005**, 61, 651–657.

(120)   A. McPherson, *Methods* **2004**, 34, 254–265.

(121)   U. K. Laemmli, *Nature* **1970**, 227, 680-685.

(122)   G. D. van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, J. Clardy*, J. Mol. Biol*. **1993**, 229, 105-124.

(123)   K. Ahrer, A. Buchacher, G. Iberer, D. Josic, A. Jungbauer, *J. Chromatogr. A* **2003**, 1009, 89–96.

(124)   A. D. Hanlon, M. I. Larkin, R. M. Reddick, *Biophys. J.* **2010**, 98, 297–304.

(125)   A. R. Ferrh-D'Amare, S. K. Burley, *Structure* **1994**, 2, 357-359.

(126)   A. R. Ferrh-D'Amare, P. Pognonec, R. G. Roeder, S. K. Burley, *EMBO J.* **1994** 13, 180-189.

(127)   A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff, E,B, S. K.Burley, *Nature* **1993**, 363, 38-45.

(128)   W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 133-144.

(129)   R. J. Gildea R. Gildea, D. G. Waterman, J. M. Parkhurst, D. Axford, G. Sutton, D. J. Stuart, N. K. Sauter, G. Evans, G. Winter, *Acta Crystallogr. D Biol. Crystallogr.* **2014**, 70, 2652-2666.

(130)   K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2000**, 56, 1612-1621.

References

(131) K. Cowtan, K. Fitting molecular fragments into electron density. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, 64, 83-89.

(132) G. N. Murshudov *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 67, 355-367.

(133) G. M. Sheldrick, *Acta Crystallogr. D Biol. Crystallogr.* **2008**, A 64, 112-122.

(134) P. Skubak, N. S. Pannu, *Nat Commun* **2013,** 4, 2777.

(135) P. Skubak, W. J. Waterreus, N. S. Pannu, *Acta Crystallogr. D Biol. Crystallogr*. **2010**, 66, 783-788.

(136) P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 486-501.

(137) P. D. Adams, et al., *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 213-221.

(138) P. V. Afonine, P. V. et al., *Acta Crystallogr. D Biol. Crystallogr.* **2012**, 68, 352-367.

(139) E. Blanc et al., *Acta Crystallogr. D Biol. Crystallogr.* **2004**, 60, 2210-2221.

(140) J. Painter, E. A. Merritt, *J. Appl. Crystallogr.* **2006**, 39, 109-111

(141) M. D. Winn, G. N. Murshudov, M. Z. Papiz, *Methods Enzymol* **2003**, 374, 300-321.

(141) A. J. McCoy *et al.*, *J Appl Crystallogr.* **2007**, 40, 658-674.

(142) V. B. Chen *et al.*, *Acta Crystallogr. D Biol. Crystallogr*. **2010**, 66, 12-21.

(143) D. C. Swarts, J. W. Hegge, I. Hinojo, M. Shiimori, M. A. Ellis, J. Dumrongkulraksa, R. M. Terns, M. P. Terns, J. van der Oost, *Nature* **2014**, 507, 258-261.

(144) S. Willkomm, C. A. Oellig, A. Zander, T. Restle, R. Keegan, D. Grohmann, S. Schneider, *Manuscript under review*

10. Appendix

10.1.   Sequence of MjAgo

Amino acid-Sequence

```
MVLNKVTYKI NAYKIKEEFI PKEVHFYRIK SFVNEAFNFY RFVNFYGGMI
INKKDKSFVL PYKVDNKVLK YKDGNNEIPI DIEYIKSLKL EYVKPEIAEK
LVRGYLKSVH KIEPELSRII KNIRKHKVVE NIKVESYCEY EVKKHDGDYY
LILNFRHTAS ITKHLWDFVN RDKALLEEYV GKKIIFKPNP KVRYTISLVD
APNPQKIEEI MSHIIKYYKW SEDMVKSTFG EIDYNQPIMY CEEILEPFAP
QFCNLVFYMD ELDSYILKEL QSYWRLSNEN KGKIINEIAK KLRFIDNTPK
ELEFMKFNNT PLLVKDVNKN PTKIYSTNTL FTWIYNQNAK IYLPYDVPEI
IRNKNLLTYI LIDEEIKDEL KAIKDKVNKM FRNYNKIANK TELPKFNYAN
RWKYFSTDDI RGIIKEIKSE FNDEICFALI IGKEKYKDND YYEILKKQLF
DLKIISQNIL WENWRKDDKG YMTNNLLIQI MGKLGIKYFI LDSKTPYDYI
MGLDTGLGIF GNHRVGGCTV VYDSEGKIRR IQPIETPAPG ERLHLPYVIE
YLENKANIDM ENKNILFLRD GFIQNSERND LKEISKELNS NIEVISIRKN
NKYKVFTSDY RIGSVFGNDG IFLPHKTPFG SNPVKLSTWL RFNCGNEEGL
KINESIMQLL YDLTKMNYSA LYGEGRYLRI PAPIHYADKF VKALGKNWKI
DEELLKHGFL YFI
```

10.2.   DNA sequences

1st Guide DNA   5´ p-TGAGGTAGTAGGTTGTATAGT 3´

2nd Guide DNA   5´ p-AGAGGTACGTGCTGAGGCTT 3´

 Target DNA   5´ TATACAACCTACTACCTCGT 3´
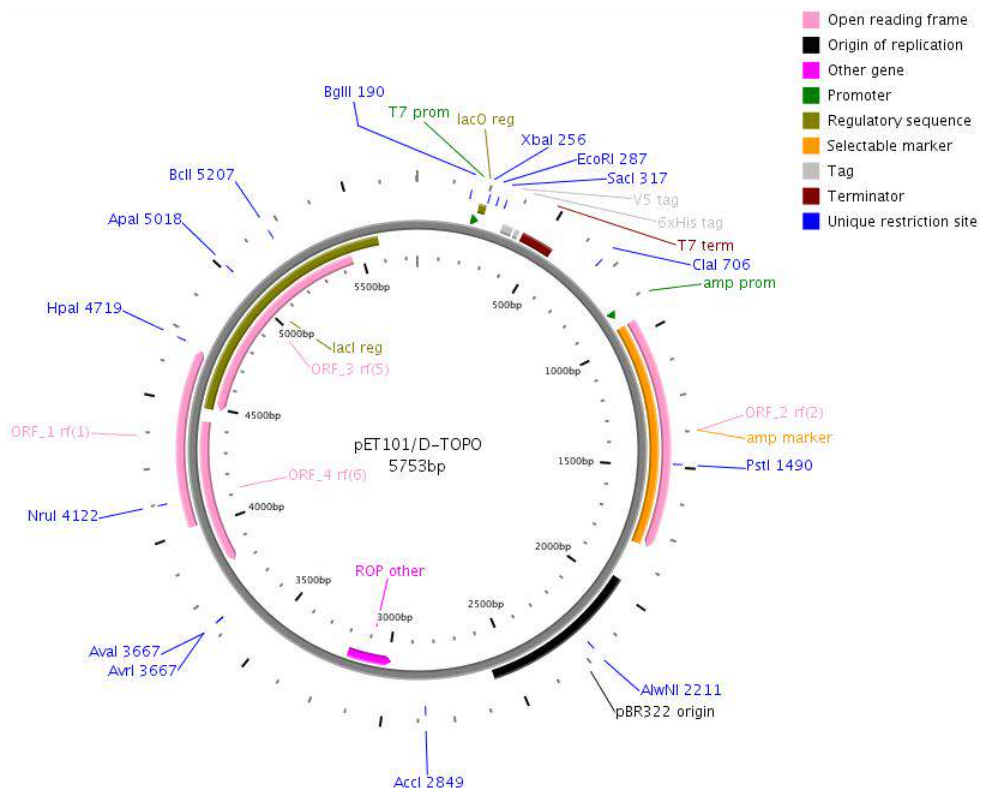
10.3.   Plasmid Map

Figure 10.2.1: Plasmid map of the plasmid used for recombinant expression of MjAgo.

11. Publications

Parts of this thesis will be published:

Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein.
S. Willkomm*, C. A. Oellig*, A. Zander, T. Restle, R. Keegan, D. Grohmann, S. Schneider.
(in preparation)

* these authors contributed equally to this work

12. Declaration

I, Christine A. Oellig, hereby declare that I independently prepared the present thesis, using only the references and resources stated. This work has not been submitted to any examination board yet. Parts of this work will be published in scientific journals.

Munich, September 2016