From the Institute of Molecular Medicine
of the University of Lübeck
Director: Prof. Dr. Georg Sczakiel

Mechanistic investigation of human Argonautes and *Streptococcus pyogenes* Cas9 substrate interactions

for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

From the Department of Natural Sciences

Submitted by
Zhenhuang Yang
from Shanxi, China

UNIVERSITÄT ZU LÜBECK
STIFTUNGSUNIVERSITÄT
SEIT 2015

Lübeck, 2015

First referee:	Prof. Dr. Tobias Restle
Second referee:	Prof. Dr. Thomas Peters
Date of oral examination:	20.01.2016
Approved for printing:	20.01.2016

Contents

1. Summary

Nucleic acids are one of the most important biological macromolecules, essential for all known forms of life. Nucleic acids include deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). They not only encode genetic information but are also involved in the regulation of gene expression. In this PhD thesis, I focused on nucleic acid and endonucleases and their regulatory functions. The mechanism and functional interaction between nucleic acids and these proteins was studied in two regulatory systems.

The first system comprises Argonaute (Ago) proteins and RNA in RNA interference (RNAi). RNAi is a cellular pathway by which gene expression is regulated at the post-transcriptional level. Argonautes are the central protein components of the RNA-induced silencing complex (RISC) and involved in small RNA (sRNA) silencing pathways. MicroRNAs (miRNAs) and short interfering RNAs (siRNAs) are the two primary categories of non-coding small RNAs which Argonautes bind to. All four human Argonautes (hAgo1–hAgo4) are associated with small RNAs although only hAgo2 has a nuclease function. Here, Ago and RNA interaction was analysed in vitro. The difference in miRNA binding kinetics between hAgo1 and hAgo2 applying steady-state and pre-steady-state techniques was determined. With regard to complex formation, the results indicate hAgo1 binds to mismatched targets more effectively than hAgo2. Furthermore, domain-swapping experiments revealed an important role for the PAZ domain in the miRNA interplay, which is further supported by B factor analysis. The flexible PAZ domain regulates miRNA binding to mismatched targets. An amino acid sequence alignment of Agos in different species showed hAgo1 is evolutionary more conserved than hAgo2. In summary, it is strong evidence that hAgo1 shows a preference for miRNA over siRNA. Moreover, hAgo1 binds to miRNA targets more effective than hAgo2. Combining these findings with sRNA sorting reported in flies and plants suggests miRNA sorting mechanism might also apply to humans, which inherited from lower organisms because of the evolutionary conservation of hAgo1.

The second system addresses DNA interference by CRISPR-Cas9. CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) system is a prokaryotic immune system that confers resistance to foreign genetic elements. Bacterial adaptive immunity uses Cas proteins together with CRISPR transcripts for invading DNA degradation. The enzyme Cas9 is an RNA-guided endonuclease that targets foreign DNA by RNA-DNA hybridisation, and is now developed as a powerful tool for genome engineering. Here, the kinetics of target DNA binding and cleavage by Cas9-sgRNA were determined, and these kinetic data was related to conformational changes during the process. Combing the data obtained in this thesis with other reported biochemical studies and crystal structures it was possible to propose a comprehensive mechanistic model describing DNA interrogation by Cas9. Furthermore, it was found that PAM recognition is essential for target binding and cleavage. Target sequence pairing to guide RNA provides high binding affinity and a sufficient “seed” region pairing is required for initial RNA-DNA heteroduplex formation.

In this thesis, mechanistic details of the complex processes of these two RNA or DNA interference pathways were further unraveled. Both, regulation of gene expression in eukaryotes and host defense in prokaryotes are based on small noncoding RNAs, which guide either RNA or DNA interference. Here, the importance of the seed regions was highlighted, being a good example of convergent evolution; i.e. “seed” mechanism, which is used by both Argonaute and Cas9 proteins. All small guide RNAs identify their target sequences by base pairing in the “seed” region, a 6–12 nucleotide comprising segment, which plays a critical role in the complicated initial target search and further down steam processes. This holds true for target RNA recognition by Argonaute proteins during RNA interference in eukaryotes and for target DNA interrogation by Cas9 during DNA interference in prokaryotes.

Zusammenfassung

Nukleinsäuren gehören zu den wichtigsten biologischen Makromolekülen und sind essentiell für alle Lebensformen. Nukleinsäuren beinhalten Desoxy- (DNA) und Ribonukleinsäuren (RNA). Sie kodieren nicht nur für genetische Informationen sondern sind auch involviert in die Regulation der Genexpression. Die vorliegende Doktorarbeit beschäftigt sich mit Endonukleasen und deren regulatorischer Funktion. Der Mechanismus und die funktionelle Wechselwirkung dieser Proteine mit Nukleinsäuren wurden in zwei regulatorischen Systemen untersucht.

Das erste System beinhaltet Argonauten Proteine und RNA im Kontext der RNA-Interferenz (RNAi). RNAi ist ein zellulärer Stoffwechselweg zur Genregulation auf posttranslationaler Ebene. Argonauten sind die zentrale Proteinkomponente des RNA-induzierten Silencingkomplexes (RISC) und involviert in den durch kleine RNAs (sRNA) vermittelten Silencing Stoffwechselweg. MikroRNAs (miRNAs) und kurze interferierende RNAs (siRNAs) sind die beiden Hauptkategorien von nicht-kodierenden kleinen RNAs an die Argonauten binden. Obgleich nur das humane Argonauten Protein 2 (hAgo2) eine Nukleasefunktion besitzt, binden alle 4 humanen Argonauten Proteine (hAgo1-hAgo4) kleine RNAs. In der vorliegenden Arbeit wurde die Wechselwirkung zwischen Argonauten und RNA *in vitro* analysiert. Unter Anwendung von *steady-state* und *pre-steady-state* Methoden wurden Unterschiede hinsichtlich der Kinetik der Bindung an miRNAs analysiert. Die erhaltenen Ergebnisse zeigen, dass hAgo1 fehlgepaarte Zielmoleküle effektiver bindet als hAgo2. Darüber hinaus zeigen Proteindomänenaustauschexperimente, dass der PAZ-Domäne hierbei eine entscheidende Funktion zukommt. Dies wird durch eine B-Faktoren-Analyse unterstützt. Die flexible PAZ-Domäne reguliert die miRNA Bindung an fehlgepaarte Zielmoleküle. Ein Argonauten Protein Aminosäuresequenzvergleich mit verschiedenen Spezies zeigte, dass hAgo1 evolutionär konservierter ist als hAgo2. Zusammengefasst ergeben sich hieraus eindeutige Evidenzen, dass hAgo1 eine klare Tendenz zeigt miRNA Zielmoleküle effizienter zu binden als hAgo2. Zusammengenommen mit Beobachtungen bei Fliegen und Pflanzen bezüglich einer Sortierung von kleinen RNAs ergibt sich, dass ein derartiger Mechanismus auch im Menschen von Bedeutung sein könnte. Dies könnte in der evolutionären Konservierung von hAgo1 bis hin zu niedrigen Lebensformen begründet sein.

Das zweite System beschäftigt sich mit DNA-Interferenz durch CRISPR-Cas9. Das CRISPR (clustered regularly interspaced short palindromic repeats)-assoziierte System (Cas) ist ein prokaryotisches Immunsystem zur Abwehr fremder genetischer Elemente. Die adaptive bakterielle Immunität beruht auf Cas Proteinen in Verbindung mit CRISPR Transkripten um eindringende DNA zu degradieren. Das Enzym Cas9 ist eine RNA-gesteuerte Endonuklease die fremde DNA durch RNA/DNA-Hybridformation targetiert. Derzeit wird dieses System als äußerst effektives Werkzeug für die Gentechnologie benutzt und weiterentwickelt. In der vorliegenden Arbeit wurde die Kinetik der Ziel-DNA Bindung und Spaltung durch Cas9-sgRNA untersucht und versucht die erhobenen kinetischen Daten strukturellen Umlagerungen während des Prozesses zuzuordnen. Durch Kombination der in dieser Arbeit erhaltenen Daten mit publizierten biochemischen und Röntgen-kristallographischen Daten war es möglich ein erstes detailliertes mechanistisches Modell der Cas9/DNA-Wechselwirkung zu postulieren. Darüber hinaus zeigte sich, dass eine Erkennung der sogenannten PAM-Sequenz essentiell ist für Zielmolekül Bindung und Spaltung. Basenpaarungen zwischen dem Zielmolekül und der sogenannten *guide* RNA gewährleisten eine hohe Bindungsaffinität. Ausgangspunkt für diese RNA/DNA-Heteroduplexausbildung ist eine initiale Wechselwirkung im Bereich der sogenannten *seed* Region.
In der vorliegenden Arbeit wurden mechanistische Details der komplexen Abläufe dieser zwei RNA- bzw. DNA-Interferenz Stoffwechselwege weiter entschlüsselt. Sowohl Regulation der Genregulation in Eukaryoten als auch Wirtsabwehrmechanismen in Prokaryoten beruhen auf kleinen nicht-kodierenden RNAs, die als Steuerungselement für RNA- bzw. DNA-Interferenz fungieren. Die Bedeutung der *seed* Region wurde besonders beleuchtet. Sie ist ein gutes Bespiel für konvergente Evolution. Man spricht in diesem Zusammenhang auch von *seed* Mechanismus, der sowohl für die Argonauten als auch die Cas Proteine von großer Bedeutung ist. Alle kleinen *guide* RNAs erkennen die entsprechenden Ziekmoleküle durch Basenpaarung in dieser *seed* Region, die 6-12 Nukleotide umfasst. Diese Wechselwirkung spielt eine entscheidende Rolle während der initialen Zielmolekülsuche und aller weiteren darauf folgenden Schritte. Das gilt sowohl für die Ziel-RNA Erkennung durch Argonauten Proteine während der RNA-Interferenz in Eukaryonten als auch der Ziel-DNA Abtastung durch Cas9 während der DNA-Interferenz in Prokaryoten.

2. Introduction

2.1. RNA interference

RNA interference (RNAi) is one of the most exciting findings in the past decades. Our knowledge of RNAi has expanded dramatically in the short time since its discovery. However, many new aspects of gene regulation of non-coding RNAs (1) were discovered since Andrew Fire and Craig Mello shared the 2006 Nobel Prize for their work on RNAi in the nematode worm *C. elegans* (2). Two types of small RNA molecules, i.e. microRNA (miRNA) and small interfering RNA (siRNA), are central to RNAi. Although, the overall mechanism of RNAi is well understood, many molecular details are not.

2.1.1. Si- and miRNA RNAi pathways

Small RNAs (sRNAs) as found in plants, animals, fungi, and viruses play crucial roles in gene regulation. They are not translated into proteins; instead, they mediate regulation of a variety of cellular processes by silencing pathways. Non-coding small RNAs occur in different forms. The best known classes are short interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-associated RNAs (piRNAs). SiRNAs arise from cleavage of long double-stranded RNA molecules and are particularly important for modulation of the activity of transposons and combating viral infection (3). MiRNAs are encoded by specific genes and function by repressing mRNA translation (1). PiRNAs are associated with the Piwi subfamily of Argonaute proteins and are essential for the development of germ cells (4).

SiRNAs and miRNAs are the two primary categories of non-coding small RNAs, which regulate gene expression on post-transcriptional levels. For example, in animals, siRNAs direct target mRNA cleavage in highly complementary sites and function as defense mechanism against foreign genetic elements such as viruses and transposons. In contrast, miRNAs guide mRNA binding, as key regulators of development, physiology and disease through a pathway that involves translational repression or destabilization of mRNAs rather than direct Ago-mediated RNA cleavage. This is triggered by binding of miRNAs to a partially

complementary target sequence (5). Figure 2.1 illustrates si- and miRNA pathways and exemplary modes of action in humans.

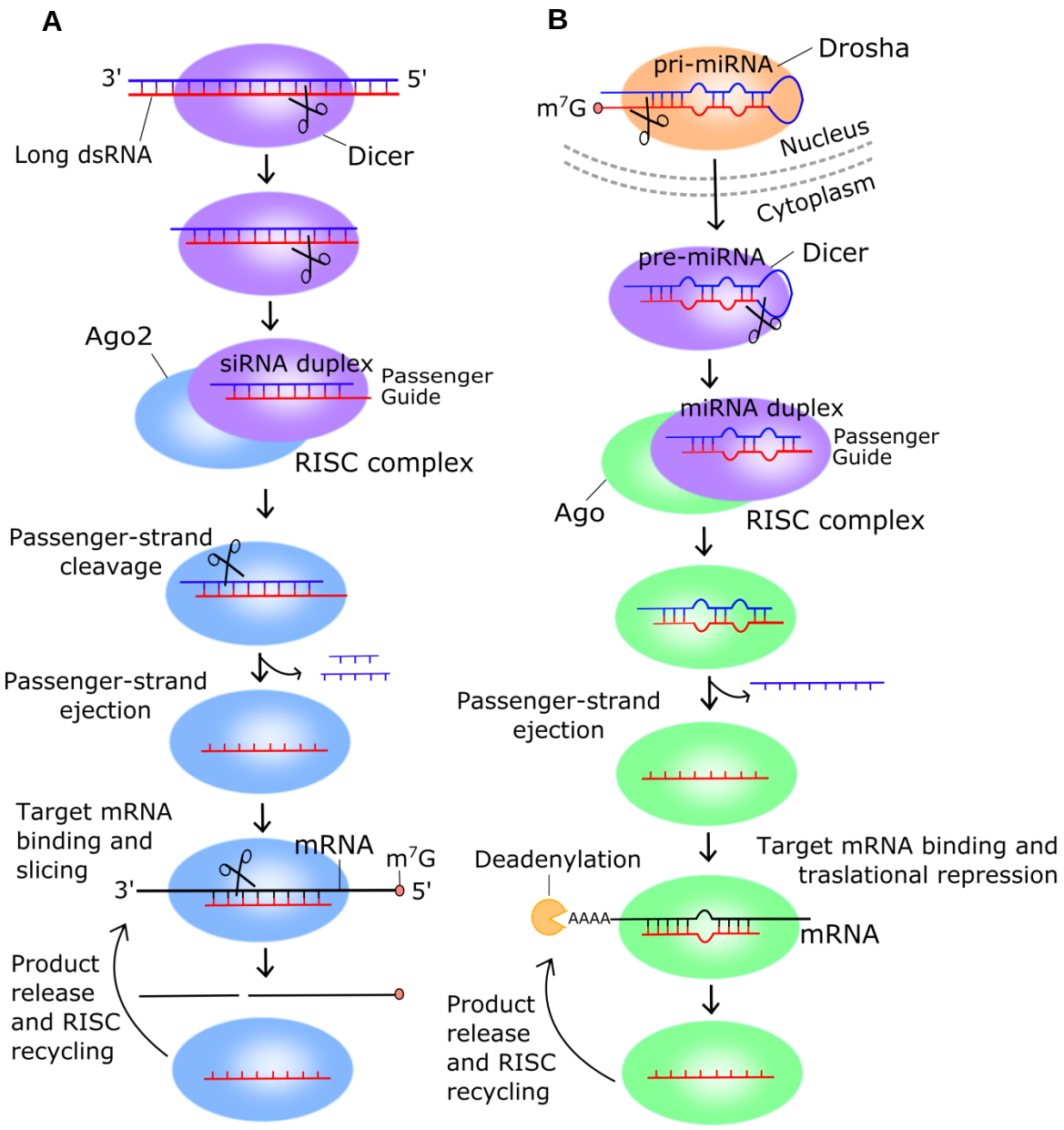


Figure 2.1 Mechanism of siRNA- and miRNA-mediated RNAi. (A) siRNAs mediate RNAi by endonucleolytic cleavage of target RNAs. These siRNAs are derived from long double-stranded RNA (dsRNA) molecules. As a molecular ruler, endonuclease Dicer cleaves the dsRNA at 21–25-nucleotide intervals. After Dicer-mediated cleavage, the guide strand of the siRNA duplex is loaded onto an Argonaute protein, the core of an RNA-induced silencing complex (RISC). For simplicity, here the multiprotein RISC is represented by a single Argonaute protein. Meanwhile, the non-guide (passenger) strand

is cleaved by an Argonaute protein and ejected. Then target RNAs that contain a fully complementary sequence to the guide siRNA are associated with the Argonaute protein followed the slicing. The cleaved target RNA is released after slicing, and the Argonaute is recycled for another round. (B) miRNAs silence target mRNAs without a direct cleavage. Primary transcripts (pri-miRNAs) are transcribed from endogenous miRNA genes, then excised by the Drosha complex to yield precursor miRNAs (pre-miRNA) in the nucleus (for simplicity, the complex is represented just by Drosha). In the cytoplasm, the pre-miRNAs are cleaved to miRNA duplexes by Dicer. The guide strand is loaded onto an Argonaute protein and passenger strand is ejected. The mechanism of miRNA-mediated silencing is unclear but is assumed to occur by translational repression of target mRNA and deadenylation of mRNA poly (A) tails, which eventually leads to mRNA degradation.

SiRNAs are ca. 21–25 nucleotides in length and processed from long, fully complementary dsRNAs, which are cleaved by the endonuclease Dicer. After Dicer-mediated cleavage, one strand of the siRNA duplex (guide) is loaded onto the Argonaute protein. During loading, the other strand of the siRNA duplex (passenger) is cleaved by Argonaute protein and ejected. Within Argonaute the guide siRNA targets perfectly complementary sequences of RNAs and trigger slicing of such targets (Figure 2.1A). The cleaved target RNA is released, allowing Argonaute to be a multiple turnover enzyme, so RISC may be recycled for further slicing of targets (6). RISC assembly is an ordered multi-step pathway and each sRNA class requires specific accessory factors (7).

Whereas miRNAs are also 21–25 nucleotides in length, they show a mismatch-containing double-strand. Primary transcripts (pri-miRNAs) comprising a stem-loop structure are cut by Drosha leading to precursor miRNA (pre-miRNAs). Then, Dicer cleaves the pre-miRNA, producing a miRNA duplex and the guide strand is loaded onto the Argonaute (Figure 2.1B). Typically, animal miRNAs contain mismatches to sequences in the 3´-untranslated regions of their target mRNAs (or 5´-end seed region of guide strand), and this lack of complementarity leads to a different mechanism for the regulation of target mRNA expression. The mRNA is not cleaved (8, 9). In addition, unlike the siRNA-RISC, some Argonaute proteins involved in the miRNA-RISC lack the catalytic properties required for cleavage (6, 10).

2.1.2. Sorting of small RNAs

Both miRNAs and siRNAs fulfill their regulatory function when they are bound to Argonaute (Ago) proteins, but can be distinguished since they trigger two different RNA interference pathways. Sorting of small silencing RNAs has been discovered in plants and flies and shows that the pathways of siRNA and miRNA diverge in part due to a sorting mechanism that distinct sRNA duplexes directly into specific Ago-RISCs. There is a noticeable high amount of studies dealing with Ago protein using the model system *Drosophila melanogaster* (11–21). Here, miRNAs are loaded into Argonaute1 (Ago1) from miRNA/miRNA* duplexes, whereas siRNAs are loaded into Argonaute2 (Ago2) from guide/passenger duplexes. The sorting is based on sRNA duplex structure and the identity of the first nucleotide (11). A identical sRNA assorting among Ago complexes was found in *Arabidopsis* (AtAgo) and shows both: miRNA duplex structure and the 5′-terminal nucleotide contribute to miRNA sorting (22). AtAgo2 favours RNA duplexes with no middle mismatches, whereas AtAgo1 tolerates or prefers duplexes with central mismatches (23). However, such sorting mechanisms seem to be lost in more complex organisms. Deep-sequencing and cloning of human Argonaute-associated endogenous miRNAs did not reveal clear evidence for miRNA sorting in humans (24–26). Although, besides there are some Ago1-specific sRNAs, Ago may define the length of mature miRNAs (27, 28). It is more likely that all Agos contribute to miRNA silencing in mammals, and individual Agos have overlapping functions in this process (29). hAgo2 is the only active slicer among the four human Argonautes and as such it is essential for siRNA cleavage. The other non-catalytic Ago proteins (hAgo1, 3 and 4) might have at least partially redundant functions (30).

2.2. Argonaute proteins

Small non-coding RNAs were found to be important contributors to gene regulation during the past decade. To conduct their biological functions, Argonaute proteins are a unique class of proteins required for these small RNAs. Ago proteins occur in bacteria, archaea and eukaryotes.

2.2.1. Evolution and Classification

After the first mention of Argonaute in a report describing a mutant in *Arabidopsis thaliana* (31), Ago proteins have been found to be conserved among all domains of life. Lately, it was characterized that prokaryotic Argonaute proteins (pAgos) take part in host defense such as DNA interference (DNAi), while eukaryotic Argonaute proteins (eAgos) guide many processes by RNA interference (32). Ago proteins evolved as protein families and both eAgos and pAgos belong to the PIWI-protein superfamily, containing a PIWI (P element-induced wimpy testis) domain and in few instances a PAZ (PIWI-Argonaute-Zwille) domain. (33).
In eukaryotes, Ago proteins can be classified into three paralogous groups based on phylogenetic analysis: the Ago subfamily, the PIWI subfamily, and the worm-specific WAGO subfamily (13, 34). In *Arabidopsis*, ten different Ago proteins exist, and they show preferences for distinct classes of small RNAs. In *Drosophila,* there are only two Ago proteins named DmAgo1 and DmAgo2 with different binding preferences to sRNAs (30). There are four Ago proteins in humans and only hAgo2 is catalytically active, but all human Ago proteins bind miRNAs (24, 26).

2.2.2. Ago protein structures

The first Ago crystal structures determined from bacteria, i.e. a guide-free pAgos from *Pyrococcus furiosus* (PfAgo) (35), *Aquifex aeolicus* (AaAgo) (36) and short pAgo from *Archaeoglobus fulgidus* (AfAgo) (37) provided information about the overall structural organization of this protein family. Then binary and ternary structures of *Thermus thermophilus* Ago (TtAgo) bound to DNA guide strands and RNA or DNA target strands provided deep insights into the conformational changes in the binding actions of the Ago enzyme (38–40). In 2012, there were reports about crystal structures of full-length human Argonaute-2 bound to RNA guides (41, 42). These findings showed that the overall organization of domains and sRNA binding of hAgo2 is similar to that of pAgo. Lately, a hAgo2 crystal structure of a ternary complex became available (43), showing hAgo2 bound to a defined guide RNA with a short target RNA pairing in g2-g9, which represents the miRNA recognition sites (Figure 2.2). In this ternary complex, guide strand and seed-matched target RNA, as a structural model for miRNA targeting, shows that

seed pairing opens the N-PAZ channel for supplemental pairing after initial target recognition (43). Here, the 3´-end of the guide strand is released from the PAZ domain during target strand loading, according to the two state model for Argonaute protein action which is supported by pre-steady-state and FRET data (44, 45). All these results reveal the function of Ago and the mechanism of RNAi.

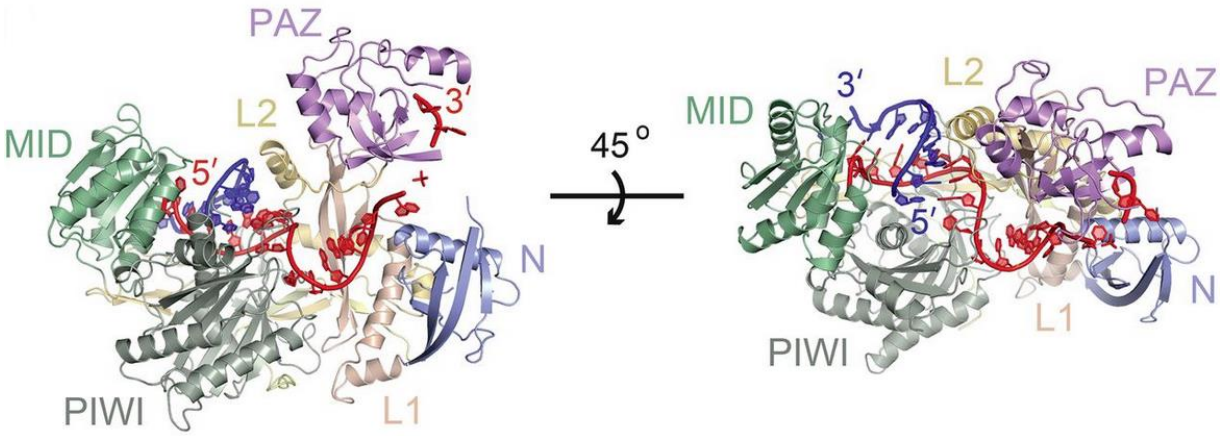


Figure 2.2 Structure of human Ago2. Front and top views of Ago2 bound to guide and target RNAs. Ago2 has the N (navy), PAZ (purple), MID (light green), PIWI (dark green) domains and linkers L1 (bisque) and L2 (gold). The RNA guide is shown in red and RNA target is shown in blue. Ago2 contains a large central cleft between two lobes (N-PAZ and MID-PIWI) connected by two linker domains (L1 and L2). Taken from Schirle, et al. (43), with kind permission of The American Association for the Advancement of Science.

The individual core functions of these four structural domains are highly conserved cross prokaryotes to eukaryotes. All eukaryotic Agos feature an N (N-terminal), a PAZ a MID (middle) and a PIWI domain, together with two domain linkers, L1 and L2 (30, 32). From crystal structures of hAgo2, it is clear that the 3´-end of single-stranded guide RNA is associated to the PAZ domain and the 5´-phosphate is anchored in a binding pocket of the MID domain (41, 42).

The MID domain contains a nucleotide-binding pocket which interacts with the phosphate group at the 5´-end of the guide (46–48). It anchors the 5´-end of the small RNA by providing a binding pocket in which the 5´-terminal base engages in stacking interactions with a conserved tyrosine. To promote the potential target binding, the seed region (nucleotides 2–7 or 2–8) of the guide stacks in a helical conformation (49).

The PIWI domain includes a RNase H-like active site, which contains a catalytic center composed of DEDX motif (here, X is aspartic acid or histidine) (35, 36). The catalytic site is completed by a glutamate residue which located in a mobile PIWI loop (the glutamate finger) (50, 51). Generally, only a subset of the Ago subfamily possesses cleavage activity. For example, only Ago2 of the four Argonautes in mammals is catalytically active and functions as an endonuclease (52, 53).

The N domain is not involved in guide binding but plays an important role in target cleavage and dissociation of cleaved product (54, 55). This critical role of the N domain was discovered by studies of human Ago 3 (hAgo3), which has no cleavage function *in vitro*, even there is an intact catalytic site in hAgo3. However, it can become an active slicer when the N domain from hAgo2 is swapped to it (54). It was also reported that N domain assists in unwinding the small RNA duplex (56).

The PAZ domain binds the 3´-end of the guide by interacting with nucleotides 20 and 21 (38, 46, 50, 57) and PAZ domain is reported to function in protecting the guide from degradation (58).

2.3. Differences between hAgo1 and hAgo2 proteins

All human Argonautes are involved in miRNA pathway and associate with sRNAs and other proteins or chaperones to carry out gene regulation (55), however, hAgo2 is the only active slicer. Structural superposition between hAgo1 and 2 shows very high overall similarity (Figure 2.3).

Individual domains appear to have the same functions (30). MID domain anchors the 5´-end and PAZ domain binds the 3´-end of the guide RNA (39, 47). A primary question in this context is why hAgo2 is the only slicer among the four human Agos. The answer was thought to reside in the difference of the activity site, i.e. Asp-Glu-Asp-His (DEDH) (35, 36). Among the remaining three, hAgo1 has an arginine instead of the active-site histidine, whereas hAgo4 has one of the aspartates missing, this providing an indication of the loss of endonuclease activity. Surprisingly, the catalytic tetrad of hAgo3 is intact, but it is still inactive for

cleavage activity (see sequence alignments in Figure 5.21) (54). This strongly indicates that the active site is not the only essential determinant for slicer activity.

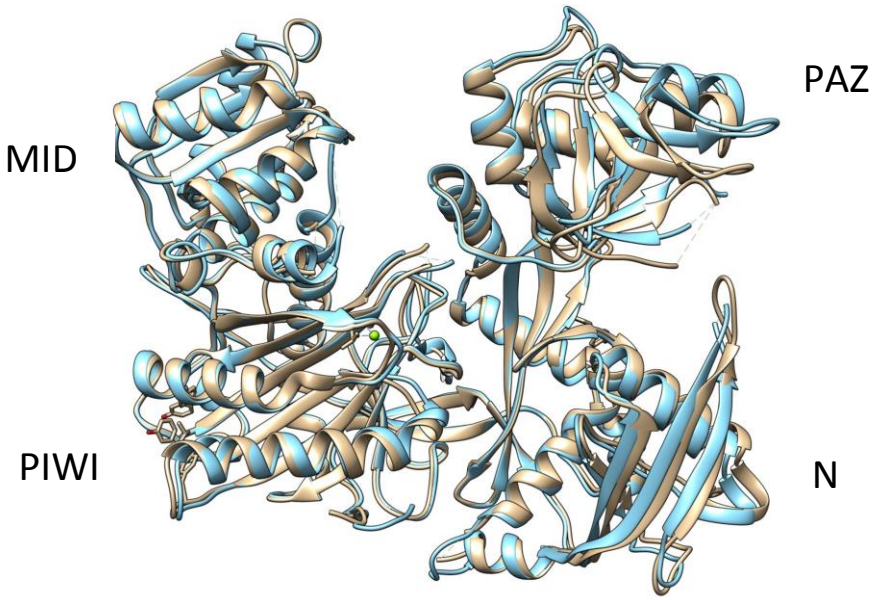


Figure 2.3 Structural superposition of hAgo1 and hAgo2 proteins. Structural resemblance of hAgo2 (PDB ID 4W5N) and hAgo1 (PDB ID 4KRF). The figure was prepared with Chimera.

Continuative studies focused on Ago1 and Ago3, because Ago4 is not expressed in most cell lines (59). In a recent publication structural features between hAgo1 and hAgo2 were compared. The conformational differences between binary guide-RNA-bound hAgo1 and hAgo2 complexes are within insertion element cS7 (aa 669–675) proximal to the catalytic tetrad (60). A reconstitution of the catalytic site was reported with R805H leading to low-level hAgo1 slicing activity. Moreover, the cleavage activity can be substantially augmented by combining R805H with cS7 substitutions (P670S and P675Q) (60). Biochemical studies of cleavage activity between hAgo1 and hAgo2 proteins revealed elements within the N-terminus to be important for the activity (54, 61). The slicer activity of hAgo1 is substantially enhanced by swapping amino acids 1–64 of the N domain of hAgo2, combined with a mutation in a loop adjacent to the active site. So besides mutations in the catalytic center of the PIWI-domain, minimal changes required to impart activity to hAgo1 are swapping of the N-terminal domain (54).

Structural comparison also shows the N domain of hAgo1 forming a convex surface within the nucleic acid binding channel near the inactive catalytic tetrad (62). The regulatory role of the N domain shows slicer activity depends on a sophisticated interaction between the catalytic site and more distant regions of the enzyme. Although there are now crystal structures of hAgo1 and hAgo2 available, it remains unclear why there are four distinct Argonaute proteins in humans. Moreover, the different interplay between Ago and substrates needs to be deciphered.

2.4. DNA interference

Prokaryotes also encode Argonautes, but according to current knowledge none of these are involved in typical RNAi pathways. Recently, it was unraveled that pAgos are involved in host defense (32, 63). However, in contrast to eukaryotes, prokaryotes use DNA rather than RNA. So the characterized eukaryotic Argonaute proteins control a wide range of gene regulation by RNAi, whereas prokaryotic Argonaute proteins participate in host defense by DNAi. Argonaute from a bacterium uses small interfering DNA guides to target foreign DNA (e.g. invading plasmids) during and after transformation. The pAgos can cleave single stranded DNA (ssDNA) or double stranded DNA (dsDNA) (64, 65).

The best studied pAgo in DNA interference is *Thermus thermophilus* Ago (TtAgo), which has been discovered to utilize DNA guides for single stranded RNA (ssRNA), ssDNA and dsDNA target cleavage *in vitro* (38–40, 50, 65). This suggests that TtAgo interferes with invading DNAs directly to lower plasmid transformation and is also able to cleave RNA targets, maybe to interfere with plasmid transcripts. Recent studies demonstrated the TtAgo has a dual-function, both in host defense and in gene silencing, which is similar to eAgos and prokaryotic CRISPR-Cas (65, 66). Moreover, TtAgo stimulates expression of CRISPR loci and associated genes which indicates that TtAgo-mediated DNAi can stimulate CRISPR adaptation (66).

2.4.1. CRISPR-Cas system

Clustered regularly interspaced short palindromic repeats (CRISPR) associated (Cas) systems provide bacteria and archaea with adaptive defense systems

against invading viruses and plasmids (67–69). As one of the most remarkable findings in recent years, the CRISPR/Cas systems enable the systematic interrogation of mammalian genome function. It was developed to be a versatile tool with various applications, even beyond genome engineering technologies (70–74).

CRISPR was first described in 1987 as a series of short direct repeats interspaced with short sequences in the genome of *Escherichia coli* (75), but the investigation of DNA repair or gene regulation began not before the mid-2000s (76, 77). In 2007, infection experiments firstly provided evidence of CRISPR-Cas mediated adaptive immunity (78). A year later, mature CRISPR RNAs (crRNAs) were shown to serve as guides in a complex with Cas proteins in *E. coli* (79), and the DNA targeting activity of the CRISPR-Cas system was reported in the pathogen *Staphylococcus epidermidis* (80). At the same time, it could be shown that viruses are constantly evolving and obtain the property to evade CRISPR mediated attenuation (81). All these findings led to a clearer understanding of natural CRISPR-Cas systems working as adaptive immunity defense.

The three CRISPR-Cas system types (I, II and III) use distinct molecular mechanisms to achieve nucleic acid recognition and cleavage. The type I and type III systems use a large complex of Cas proteins for crRNA-guided targeting (79, 82–84). In 2011 and 2012, the type II system was shown to require only a single Cas9 protein for RNA-guided DNA recognition and cleavage; a property that proved to be extremely useful for genome engineering applications (85–87).
Adaptive immunity occurs in three stages: (i) insertion of a short sequence of the invading DNA as a spacer sequence into the CRISPR array, (ii) transcription of precursor crRNA (pre-crRNA) that undergoes maturation to generate individual crRNAs (each composed of a repeat portion and an invader targeting spacer portion), and (iii) crRNA-directed cleavage of foreign nucleic acids by Cas proteins at sites complementary to the crRNA spacer sequence (72, 88).

2.4.2. CRISPR-Cas9

Biochemical studies showed that recombinant Cas9 protein can be guided by CRISPR RNAs (crRNAs) to cleave target DNA *in vitro* (87). Furthermore, a single

guide RNA (sgRNA) containing the targeting guide sequence was constructed by fusing a crRNA to a transactivating crRNA (tracrRNA) that facilitates DNA cleavage by Cas9 *in vitro* (86, 87). 6 month later, a couple of studies simultaneously demonstrated that type II CRISPR systems can be successfully engineered to facilitate genome editing in mammalian cells (89, 90). Interestingly, sgRNAs can direct DNA cleavage by Cas9 and multiple guide RNAs can be used to target more than one gene at once on a genome wide scale (91). After these initial studies, Cas9 technology has been used to facilitate a great variety of targeted genome editing applications in different experimental models (70).

Bioinformatic analyses identified Cas9 as a massive multifunctional protein with two distinctive nuclease domains, HNH and RuvC-like (76, 92, 93). Furthermore, it was found that the HNH domain is used to cleave the DNA strand which is complementary to the 20 nt long target sequence of the crRNA, whereas catalytic activity arising from the RuvC-like domain of Cas9 cleaves the other DNA strand opposite the complementary one (86, 87). Either a mutated HNH or a mutated RuvC-like domain will generate an enzyme with single-stranded DNA cleavage activity (nickase); if both domains are mutated this will result in a deactivated Cas9 (dCas9) (87), a RNA-guided DNA binding protein. The Cas9 nickases are used to reduce off-target effects, because nickases create a single stranded break (SSB), and SSBs are repaired via the high-fidelity base excision repair pathway, so the double-nicking strategy can improve on-target specificity (94, 95). The catalytically deactivated version of Cas9 has been used for targeted gene screens and regulation on a genome-wide scale, because of its programmable property. For example, various RNAs or proteins can be tethered to Cas9 to change transcription states of targeted genomic loci (96). More and more results demonstrated that dCas9 can be used as a modular and flexible DNA-binding platform for target DNA sequence recognition within the mammalian genome (97, 98), and it is useful for the discovery of therapeutic targets (99). As a powerful new tool, it has caused an explosion in visionary applications (Figure 2.4) (100).

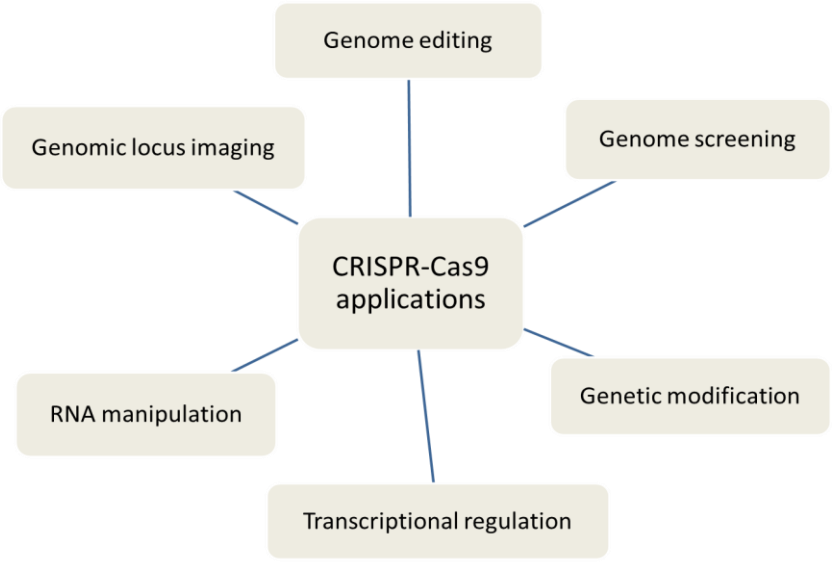


Figure 2.4 Exemplary diverse applications of CRISPR-Cas9. RNA-guided Cas9, originally developed for genome editing, were expanded to many other applications because of the programmable character of Cas9.

2.4.3. Off target effects

Despite the high amount of potential applications, some reports suggest that CRISPR-Cas9 causes frequent off-target editing events (101), so the CRISPR technology may ultimately be limited by its lack of specificity. Because of the permanent genome modifications by CRISPR genome editing, the specificity of Cas9 nuclease targeting is of particular concern, especially for clinical applications e.g. gene therapy (70). Previous studies suggested the first 8–12 protospacer adjacent motif (PAM)-proximal nucleotides in the guide RNA (so called “seed” region, see section 2.5.3) determines specificity of Cas9 (87, 89, 102). Furthermore, these results demonstrated that Cas9 tolerates mismatches within the guide sequence. Meanwhile it was found that Cas9 is sensitive to the number and position of these mismatches (103, 104). Besides using paired Cas9 nickases, variant approaches have been used for better targeting: dCas9-FokI fusion nucleases (105, 106), 5´-truncated sgRNAs (94) and CRISPR scaffold RNAs (107) have been reported recently to reduce off-targeting effects and to further improve Cas9 specificity. However, all these strategies only improve the fidelity of target recognition, but do not eliminate off-target effects. Future improvements in Cas9

specificity may be achieved via structure and function analyses of Cas9, e.g. via rational design or directed evolution.

2.5. Cas9 protein

As described above, Cas9 is the only endonuclease in Type II CRISPR-Cas system which has been widely used in genetic engineering. Cas9 functions in conjunction with crRNAs and a tracrRNA or a sgRNA. The usage of sgRNA simplify the applications, so most biochemical and structural studies are based on the model Cas9-sgRNA (100).

2.5.1. Structure analyses of Cas9 protein

Crystal structures of Cas9 protein are available since 2014 (108–110), especially the Cas9-sgRNA complex bound to a ssDNA target provides detailed insights into the mechanism of Cas9 upon binding guide RNA and complementary target DNA (Figure 2.5).

These structural studies show, Cas9 adopts an auto-inhibited conformation in the absence of nucleic acids. A conformational rearrangement of Cas9 protein triggered by guide RNA binding, leads to a central channel formation for target DNA recognition. It suggests Cas9 interrogates the potential target DNA by searching for PAM motifs throughout the genome. Further conformational changes trigger PAM binding, followed by dsDNA unwinding and sgRNA target DNA pairing (111).

However, the mechanistic details of underlying conformational changes remain unclear. For example, does the large structural rearrangement contribute to DNA duplex unwinding when Cas9 does not harbor energy-dependent helicase activity (87, 112)?

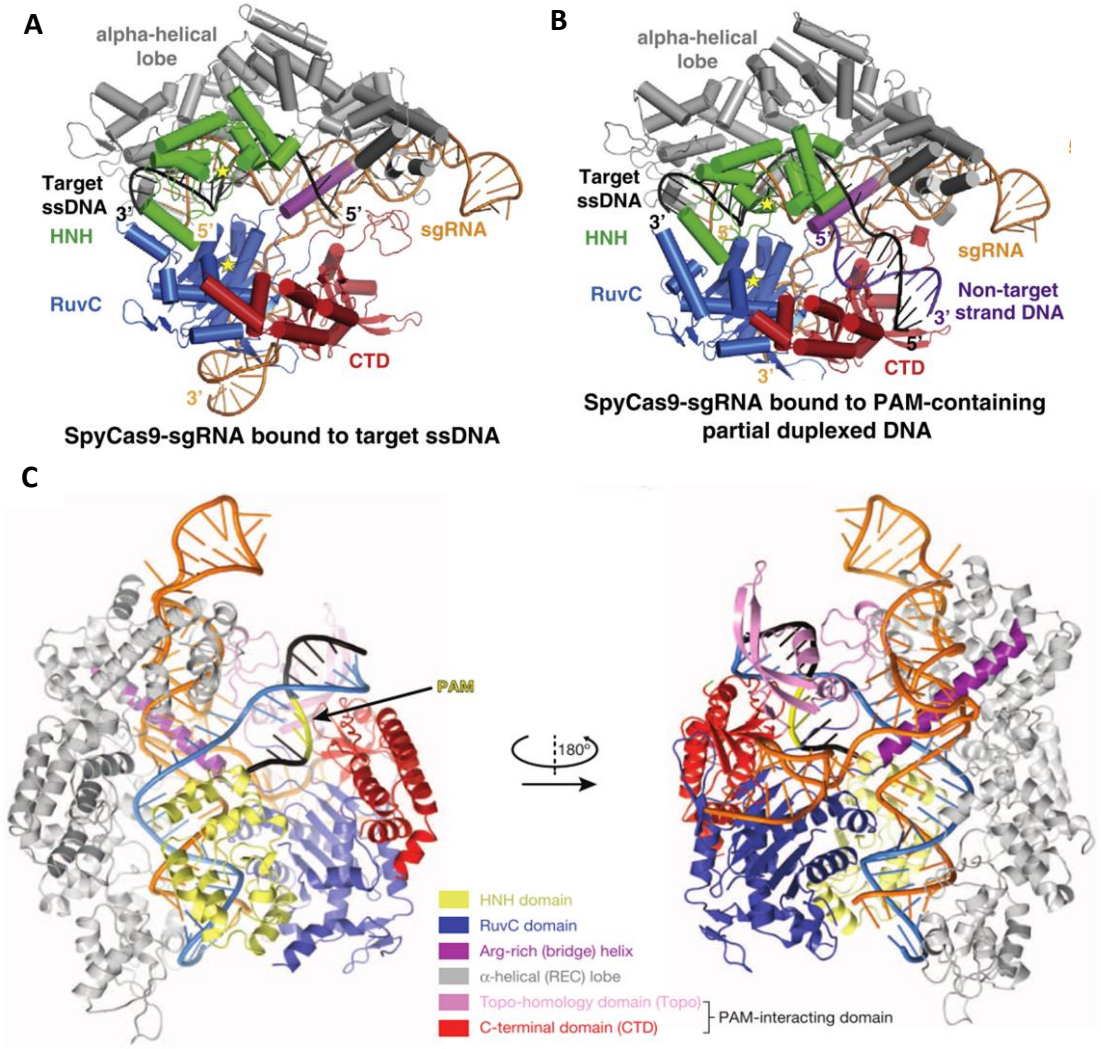


Figure 2.5 Crystal structure of Cas9 protein. (A) Ribbon representation of the *S. pyogenes* Cas9 (SpyCas9)-sgRNA-ssDNA ternary complex (PDB ID: 4OO8). (B) Ribbon representation of the SpyCas9 viewed in the same orientation as in (A), but in complex with sgRNA and a partially duplexed target DNA containing a PAM sequence (PDB ID: 4UN3). (C) Front and rear views of the SpyCas9-sgRNA-dsDNA complex (PDB ID: 4UN3). The guide RNA is colored orange, target DNA strand in light blue and non-target DNA strand in black. The PAM trinucleotide in the non-target strand is highlighted in yellow. Modified from F. Jiang and C. Anders with kind permission of Nature Publishing Group and Elsevier (110, 111).

2.5.2. PAM motif

PAM (protospacer adjacent motif) is a short sequence motif 5´-NGG-3´ for target DNA recognition (Figure 2.6). The target recognition strictly depends on the presence of PAM. In the absence of PAM, even target sequences perfectly

complementary to the guide RNA sequence are not recognized by Cas9 (112). The molecular mechanism of PAM recognition was studied by X-ray crystallography of Cas9 protein in complex with a single-molecule guide RNA and a target DNA containing a canonical PAM (110). It reveals that the GG dinucleotide in the PAM motif of the non-complementary strand is read out via interactions with conserved arginine residues of the carboxy-terminal domain of Cas9. The phosphodiester group at position +1 in the target DNA strand interacts with a groove of the duplexed PAM, possibly resulting in local target strand separation, the so-called R-loop, immediately upstream of the PAM. PAM affects primarily the R-loop association rates observed in single-molecule experiments (113). These results suggest formation of Cas-RNA/DNA complexes starts with PAM recognition and expands toward the distal protospacer end, whereas some PAM-distant protospacer elements influence primarily R-loop stability. Together with biochemical experiments using mutated target DNAs (112), a mechanism can be proposed whereby target DNA melting starts at the time of PAM recognition, resulting in directional R-loop formation expanding toward the distal protospacer end and concomitant RNA strand invasion and RNA-DNA hybrid formation.

Figure 2.6 Schematic diagram of guide and target nucleic acids. The sgRNA is colored orange, target DNA strand in blue and non-target DNA strand in green. The PAM trinucleotide in the non-target strand is shown in yellow. Sequences are from Anders *et al.* (110).

PAM recognition is not only critical for target DNA binding but also for subsequent Cas9 catalytic activity. After the RNA-DNA hybrid duplex formation is initiated at the PAM, when the target has a matching sequence, R-loop expansion propagates and Cas9 directionally process towards the distal end of the target sequence. Furthermore, PAM regulates scission of double-stranded DNA by triggering Cas9 catalytic activity (112). In the study by Sternberg *et al*. (112), the binding affinity of target DNA and Cas9-sgRNA was determined by post-steady-state and equilibrium binding measurements. The dissociation of Cas9-sgRNA and target DNA shows faster dissociation (k_{off}) of partial complementarity targets as compared to completely paired ones. When the target lacks a nearby PAM even sequences fully complementary to the guide RNA do not bind to Cas9-sgRNA.

2.5.3. The “seed” region

From RNAi to CRISPR, both, regulation of gene expression in eukaryotes and host defense in prokaryotes, are based on small noncoding RNAs, which guide RNA or DNA interference. All small guide RNAs identify their target sequences by base pairing within a so-called “seed” region, a 6–12 nucleotide long segment, which plays a critical role in the complicated target search process (49).

The term “seed” was first used to describe positions 2–7 or 2–8 of miRNAs because it is the minimal length required to be hybridized to the target to cause translational repression (114). Furthermore, Wee *et al.* (115) showed the RNA guides can be divided into different domains. According to this study, the guide consists of an anchor (position 1), a central region (position 9–12), a 3´-supplementary region (position 13–17), and a tail region (position 18–21), each with its own distinct functions and RNA-binding properties. In CRISPR, a seed was reported in early 2015 (116). Here, crystal structures of Cas9 bound to sgRNA showed the Cas9-sgRNA complex is pre-organized for target DNA recognition. It revealed distinct conformations from both the apo and the DNA-bound states. A 10 nt long “seed” sequence is required for initial DNA interaction which is preordered in an A-form conformation. This indicated a convergent evolution of this “seed” mechanism which is also used by Argonaute proteins in RNAi (116).

RNAi-associated small RNAs, mainly miRNA and siRNA are involved in different pathways regulating gene expression in eukaryotes. CRISPR RNA (crRNA), is the guide of an adaptive immune system in bacteria and archaea, which protects microbial cells from invading nucleic acids. Although these two classes of small guide RNAs are unrelated, they show similarities in the way of their target recognition (Table 2.1). All initial recognitions occur by Watson-Crick base pairing in a seed region between guide and targets.

Table 2.1 Comparison of small guide RNA properties

	miRNA/siRNA	crRNA (Type II)
Length of seed region	6–7 nt	8–12 nt
Length of small RNA	20–25 nt	35–45 nt
Length of guide section	6–25 nt	20 nt
Position of seed	Position 2–7 or 2–8	Position 9–20
Contiguous seed	yes	yes
RNA 5´-end	P	Not described
RNA 3´-end	OH, 2´-CH_3	OH
Target nucleic acid	mRNA	Invader DNA
Associated protein	Argonaute	Cas9
References	(41–43, 51, 114, 117, 118)	(86, 87, 116, 119, 120)

RNAi-based small RNAs are relatively uniform and carry a 6–8 nt seed which is pre-ordered in an A-form helix. The RNA is tightly associated with Argonaute and target RNA binding is sequential and starts at the seed. The extent of guide–target pairing and the type of Argonaute determines the further processing of the target. The crRNA has a seed of 8–12 nt in length which is possibly pre-ordered in a structure favoring duplex formation with target DNA. The DNA target is tightly associated with Cas proteins. The PAM in the target DNA and the seed cooperate in a sequential manner to maximize the efficiency of the target search process. To achieve effective recognition of target nucleic acids, small guide RNAs make use

of seed sequences of approximately 6–12 nt. By pairing short regions, small guide RNAs can rapidly associate with and dissociate from potential targets. Moreover, the short length of the seed enables a single miRNA, siRNA, or sRNA to regulate a multitude of genes with conserved seed matching sites, reducing the number of small RNAs required to regulate a high number of genes (121, 122).

All small RNA systems are highly sensitive to mismatches within the seed. Whereas miRNAs and siRNAs may overcome seed mismatches using compensatory pairing, crRNAs lose their function when mismatches occur at seed positions (123–127). In CRISPR-Cas, instead of directly interfering with this class of mutated invaders, seed mutants trigger priming; a process in which new spacers against the same target are integrated in the CRISPR array to restore immunity of the host (126).

2.6. Aim of the study

In this PhD study, I focused on the interaction of non-coding small nucleic acids and associated proteins. The two systems investigated are part of two distinct cellular pathways. The first project described here, addresses mechanistic aspects of hAgo-protein substrate interactions. There are four Ago proteins in mammals. However, the details of their individual functions are not known. It was reported that Argonaute 2 protein in *Drosophila melanogaster* (DmAgo2) is involved in the siRNA pathway while DmAgo1 is associated with the miRNA pathway. The question is whether there are comparable scenarios for human Ago-proteins. So this study focused on potentially differential properties among hAgo1 and hAgo2. First, hAgo1 and hAgo2 had to be tested for binding properties concerning different RNA oligonucleotides, especially miRNA targets. It was intended to complement and compare these data with kinetic studies of hAgo1-hAgo2 chimeras which had to be developed, expressed and investigated in this thesis. By domain swapping, distinct functions of these domains had to be unraveled. Combining these data with published results from others, both, a fine tuning of hAgo1 and hAgo2 binding models, and of Argonaute and miRNA pathways had to be achieved.

The second project was supposed to address mechanistic details of the CRISPR system, focusing on the type II “CRISPR-Cas9 system”. The aim was to determine the affinities and binding kinetics of *Streptococcus pyogenes* Cas9-sgRNA/DNA target complexes. Furthermore, it was intended to establish a minimal mechanistic model for binding of Cas9 protein to nucleic acids. To obtain these data, three strategies were supposed to be used: (1) usage of different magnesium concentrations in binding and cleavage assays, (2) usage of a cleavage-incompetent variant of Cas9 protein (dCas9) for comparable studies, and (3) investigation of sequence effects, especially of several PAM and seed sequences. The results were intended to refine existing models of Cas9-nucleic acid interactions as well as drawing conclusions from the biochemical data to catalytic mechanisms.

3. Materials

3.1. List of manufacturers

Abcam (Cambridge, UK); Amersham Pharmacia (Freiburg, DE); Applied Biosystems (Freiburg, DE); Applied Photophysics (Leatherhead, UK); Ascenion GmbH (München, DE); Bandelin (Berlin, DE); Braun Melsungen AG (Melsungen, DE); Beckmann Coulter (Krefeld, DE); Becton DEickinson (Heidelberg, DE); Berthold (Bad Wildbad, DE); Biomers (Ulm, DE); Biometra (Göttingen, DE); Biorad (München, DE); Biotec-Fischer (Reiskirchen, DE); Biosciences (St. Louis, USA); Biozym Scientific (Hessisch Oldendorf, DE) Bosch (Gerlingen-Schillerhöhe, DE); Brand (Wertheim, DE); Canon (Krefeld, DE); Dako (Hamburg, DE); Eppendorf (Hamburg, DE); Erithacus Software (Surrey, UK); Feather (Osaka, JP); Fermentas (St. Leon-Rot, DE); Finnzymes (Espoo, Fin); Fluka (Buchs, DE); Forma Scientific (Marietta, USA); Fuji (Düsseldorf, DE); GE Healthcare (München, DE); Gerhardt (Königswinter, DE); Gibco (Karslruhe, DE); Gilson (Bad Camberg, DE); Greiner (Frickenhausen, DE); H + P Labortechnik (Oberschließheim, DE); Heidolph (Schwalbach, DE); Hellma (Jena, DE); Heraeus Instruments (Hanau, DE); Hettich (Tuttlingen, DE); Hirschmann (Eberstadt, DE); Hoefer (San Francisco, USA); Horiba Jobin Yvon (Unterhaching, DE); IBA Lifesciences (Göttingen, DE); Infors (Bottmingen, CH); Invitrogen (Karlsruhe, DE); Jackson Immuno-Research (Suffolk, UK); Jahnke & Kunkel (Staufen, DE); Jena Bioscience (Jena, DE); Kern & Sohn GmbH (Balingen-Frommern, DE); KnF Neuberger (Freiburg, DE); Kodak (Rochester, USA); Life technologies (Darmstadt, DE); LKB (Bromma, SE); Macherey & Nagel (Düren, DE); Merck (Darmstadt, DE); Micromath (St. Louis, USA); Millipore (Schwalbach, DE); Miltenyi (Bergisch Gladbach, DE); Molecular Deynamics (Sunnyvale, USA); neoLab (Heidelberg, DE); New Brunswick Scientific (Edison, USA); New England Biolabs (Ispwich, USA); Peqlab (Erlangen, DE); Perkin-Elmer (Boston, USA); Phase (Lübeck, DE); Pierce (Rockford, IL, USA); Promega (Mannheim, DE); Qiagen (Hilden, DE); Renner GmbH (Dannstadt, DE); Roche (Basel, CH); Roth (Karlsruhe, DE); Sarstedt (Nümbrecht, DE); Sartorius (Göttingen, DE); Savant (Midland, USA); ScalTec (Göttingen, DE); Sci Ed Central (Cary, USA); Schleicher & Schüll (Dassel, DE); Schott (Darmstadt, DE); Scotsman

(Vernon Hills, USA); Serva (Heidelberg, DE); Sigma-Aldrich (Deisenhofen, DE); Sörk-Tronik (Medingen, DE); Spectrum (Breda, NL); Systec (Wettenberg, DE); Takara Clontech (Saint-Germain-en-Laye, FR); Thermo Fisher Scientific (Rockford, USA); UCSF Resource for Biocomputing, Visualization and Informatics (San Francisco, USA); Vilber Lourmat (Eberhardzell, DE); Waters (Eschborn, DE); Werner Hassa GmbH (Lübeck, DE); Whatman (Dassel, DE); Wyatt Technology (Santa Barbara, USA).

3.2. Chemicals and devices

3.2.1. Chemicals

Name	Supplier
Acetic acid	Roth
Acetone	Roth
Agar	Sigma-Aldrich
Agarose	Biozym
Ammonium peroxodisulphate (APS)	Fluka
Ammonium sulphate [$(NH_4)_2SO_4$]	Gerbu Biochemicals
Ampicillin (Amp)	Sigma-Aldrich
β-Mercaptoethanol	Sigma-Aldrich
Boric acid	Roth
Bromophenol blue	Sigma-Aldrich
Calcium chloride ($CaCl_2$)	Roth
Chloroform/isoamyl alcohol (24:1,v/v)	Merck
Coomassie brilliant blue R-250	Sigma-Aldrich
Dithiothreitol (DTT)	Sigma-Aldrich
dNTPs	Jena Biosciences
Ethanol	Fisher Scientific
Ethidium bromide (EtBr)	Serv
Ethylenediaminetetraacetic acid (EDTA)	Merck
Ficoll® 400	Fluka
Formamid (entionisiert)	Sigma-Aldrich
Glutathion (reduziert)	Thermo Fisher Scientific
Glycerol	Merck

Glycine	Roth
Hydrochloric acid	Merck
Imidazol	Merck
Isopropanol	Merck
Isopropyl-β-D-thiogalactopyranosid (IPTG)	Roth
Kanamycin	Roth
Magnesium chloride ($MgCl_2$)	Sigma-Aldrich
Magnesium sulphate ($MgSO_4$)	Sigma-Aldrich
Methanol	Roth
N,N,N´,N´,-tetramethylethylenediamine (TEMED)	Roth
NTPs	Jena Biosciences
Phenol/Chloroform/3-Methyl-1-butanol	Roth
Phenylmethylsulfonylfluorid (PMSF)	Sigma-Aldrich
Potassium chloride (KCl)	Roth
Potassium dihydrogen phosphate (KH_2PO_4)	Roth
Rotiphorese gel 40 (acrylamide/bisacrylamide; 19/1)	Roth
Rotiphorese gel 30 (acrylamide/bisacrylamide; 37.5/1)	Roth
Sigmacote®	Sigma-Aldrich
Sodium acetate ($NaC_2H_3O_2$-$3H_2O$)	Merck
Sodium chloride (NaCl)	Roth
Sodium dodecylsulphate (SDS)	Roth
Sodium phosphate dibasic (Na_2HPO_4)	Merck
Stains-All	Sigma-Aldrich
Tris(hydroxymethyl)aminomethane (Tris)	Roth
Trichloroacetic acid (TCA)	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Tween-20	Roth
Urea	Roth
Xylene cyanol (XC)	Roth
Yeast extract	Roth
[γ-32P]-ATP (3000 Ci/mmol)	Perkin-Elmer

3.2.2. Consumables

Name	Supplier
Amicon® Ultra Centrifugal Filter 100,000 MWCO	Millipore
Cryotubes (1.5 ml, 2 ml)	Greiner
Centrifuge tubes (15 or 50 ml)	Greiner
Disposable cuvettes	Brand
Disposable pipettes (1 ml, 2 ml, 5 ml, 10 ml, 25 ml, 50 ml)	Greiner
Gel filtration NICK columns (G-50)	GE-Healthcare
PCR tubes (200 µl)	Biozym Scientific
Pipette tips (10, 200, 1000 µl)	Sarstedt
Polyvinyl difluoride (PVDF)	Millipore
Membrane filters (0.45 µm/0.2 µm)	Whatman
Scalpels	Feather
Syringes (1 ml, 10 ml, 20 ml, 50 ml)	Becton Dickinson
3 MM paper	Whatman®

3.2.3. Columns and column materials

Name	Supplier
Glutathion Sepharose High Performance	GE Healthcare
NICK columns Sephadex-G50	GE Healthcare
Ni-NTA Superflow	QIAGEN
TALON Superflow	Takara Clontech
Superdex 200 HiLoad 16/30	GE Healthcare
Superdex 200 HiLoad 26/60	GE Healthcare
Superformance 50-10 column	Merck
XK 16/20 column	GE Healthcare

3.2.4. Devices

Name	Supplier
Biotrap Electro-separation system	Schleicher & Schuell
Blotting device	Hoefer
Centrifuge Microfuge® R	Beckman
Centrifuge 5415 C	Eppendorf
Centrifuge Avanti™ J-25	Beckman
Centrifuge Rotixa 120 R	Hettich
Electroporation apparatus	Bio-Rad
Electrophoresis Power Supply EPS 600/3500	Amersham Pharmacia
Fluorescence cuvettes	Hellma
Fluorescence Spectrometer FluoroMax® -3	Horiba Jobin Yvon
FPLC Advanced Protein Purification System 650E	Waters
Freezer -20°C	Liebherr
Freezer -80°C	Forma Scientific
Fridge +4°C	Bosch
Gel chamber for agarose gels Sub-Cell® GT	Bio-Rad
Gel chamber for Polyacrylamid gels	Biometra
Gel chamber for SDS-Gel (10 x 10.5 cm)	Hoefer
Gel chamber for sequencing gels Sequi-Gen® GT	Bio-Rad
Gel chamber for sequencing gels Model S2	Hoefer
Handheld UV Lamp	Fisher Scientific
Incubator (bacteria)	Heraeus
Liquid scintillation counter Wallac 1409	Wallac
Microwave	Bosch
NanoDrop spectrophotometer ND-1000	Peqlab
PCR-Block UNO II	Biometra
pH meter Lab 850	Schott

PhosphorImager Typhoon™ 9500	GE-Healthcare
FUSION SOLO S	Vilber Lourmat
Pipette (2 µl, 10 µl, 20 µl, 100 µl, 200 µl, 1000 µl)	Eppendorf or Gilson
Pipette Accu-Jet	Brand
Spectrophotometer DU® 640	Beckmann
Stopped flow SX20	Applied Photophysics
Storage Phosphor Screen	Fuji
Thermomix 5436	Eppendorf
Water bath W22	Sörk-Tronik
Water Purification System	Millipore

3.3. Buffers and solutions

3.3.1. Buffers

Name	Components
Ago binding buffer (10x)	100 mM Tris (pH 7.4), 1 M KCl
Ago cleavage buffer (10x)	100 mM Tris (pH 7.4), 1 M KCl, 20 mM $MgCl_2$
Ago elution buffer	50 mM Tris (pH 8.3), 20 mM Glutathion
Ago wash buffer	50 mM Tris (pH 7.4), 1 mM EDTA, 10 mM DTT
Cas cleavage buffer (5x)	100 mM Tris (pH 7.5), 500 mM KCl, 25% glycerol, 2.5 mM EDTA (pH 8.0), 10 mM $MgCl_2$, 5 mM DTT
Cas elution buffer for IMAC	20 mM Tris (pH 8.0), 250 mM NaCl, 250 mM imidazole
Cas SEC buffer	20 mM Tris (pH 7.5), 500 mM KCl, 1 mM DTT
Coomassie blue staining solution	40% (v/v) Methanol, 10% (v/v) Acetic acid, 0.01%(w/v) Coomassie brilliant blue R-250
Denaturing loading buffer (PAGE)	7.0 M Urea, 0.1% (w/v) Bromophenol blue, 0.1% (w/v) Xylene cyanol, in 1x TBE buffer pH 8.0

Destaining solution for Coomassie	40% (v/v) Methanol, 1.0% (v/v) Acetic acid
Hybridization buffer	15 mM HEPES (pH 7.4), 50 mM CH_3COOK, 1 mM $Mg(CH_3COO)_2$
Lysis buffer for *E. coli* cells	50 mM Tris (pH 7.4), 1 mM PMSF, 1 mM EDTA, 10 mM DTT
Loading buffer (6x) for native PAGE and Agarose gel	0.25% (w/v) Bromophenol blue, 0.25% (w/v) Xylene cyanol, 25% (w/v) Ficoll 400, in 1x TAE buffer pH 7.4
Loading buffer for SDS-PAGE (6x)	300 mM Tris (pH 6.8), 12% (w/v) SDS, 60% (w/v) Glycerol, 600 mM DTT, 0.06% (w/v) BPB
Loading buffer denaturing (6x)	95% Formamid, 0.025%(w/v) SDS, 0.5 mM EDTA, 0.025% (w/v) BPB, 0.025% (w/v) XC
Running buffer (SDS-PAGE) (10x)	0.25 M Tris (pH 8.3), 2.50 M Glycine, 10% (w/v) SDS
Phosphate buffer (10x)	720 mM K_2HPO_4, 170 mM KH_2PO_4
Stains-All-Solution	15 mM Tris (pH 8.8), 0.005% (w/v) Stains-All, 5% (v/v) Formamid, 25% (v/v) Isopropanol
Transcription buffer (10x)	400 mM Tris (pH 8.3), 0.1% NP-40, 10 mM Spermidine, 500 mg/ml PEG 8000
Tris-buffered saline buffer (TBS) (10x)	0.20 M Tris (pH 7.6), 1.5 M NaCl, 3.8% (v/v) HCl
TBS-T buffer	1x TBS buffer, 0.1% (v/v) Tween-20
Transfer buffer for Western blotting	25 mM Tris (pH 8.3), 0.19 M Glycine, 20% (v/v) Methanol
Tris-acetate-EDTA buffer (TAE) (10x)	0.40 M Tris (pH 8.5), 0.01 M EDTA, 0.20 M Acetic acid
Tris-borate-EDTA buffer (TBE) (10x)	0.89 M Tris (pH 8.0), 0.02 M EDTA, 0.89 M Boric acid

3.3.2. Enzymes and antibodies

Name	Supplier
Afl II	Fermentas
Age I	Fermentas
CIP (Calf Intestine Phosphatase)	NEB
DNaseI	NEB
RiboLock™ RNase Inhibitor	Fermentas
RNase T1	Fermentas
Sac I	Fermentas
Spe I	NEB
T4 DNA Ligase	Fermentas
polynucleotide kinase	Fermentas
Taq Polymerase	NEB
T7 RNA polymerase	Prepared by T. Restle
TEV protease	Invitrogen
Monoclonal rat anti-hAgo1 (4B8)	Ascenion
Monoclonal rat anti-hAgo2 (11A9)	Ascenion
GST Antibody (B-14)	Santa Cruz Biotechnology
Polyclonal goat anti-mouse IgG HRP*-conjugated	Dako
Polyclonal goat anti-rat IgG HRP-conjugated	Jackson Immuno-Research

* Horseradish Peroxidase-Conjugated Antibodies (HRP)

3.3.3. Kits and size markers

Name	Supplier
AcTEV™ Protease	Invitrogen
Bradford Assay Kit	Biorad
Enhanced chemiluminescence (ECL)^{TM}	Pierce
GenElute Plasmid Miniprep Kit	Sigma-Aldrich
Gene Ruler™ DNA Ladder 1 kb Plus	Fermentas

Gene Ruler™ DNA Ladder High Range	Fermentas
Gene Ruler™ DNA Ladder Low Range	Fermentas
PageRuler™ Prestained Protein Ladder	Fermentas
PageRuler™ Plus Prestained Protein Ladder	Fermentas
Prestained Protein Ladder, Broad Range (10-230 kDa)	NEB
Phusion High Fidelity Polymerase Kit	NEB
RiboMAX™ Large Scale RNA Production System	Promega
Wizard® SV Gel and PCR Clean-Up Kit	Promega

3.4. Nucleic acids

3.4.1. Oligonucleotides

All RNA and DNA oligonucleotides were purchased from IBA or Biomers, and all of them are PAGE or HPLC purified.

RNA	**Sequence (5´ – 3´)**
as2b	uag agg uac gug cug agg cTT
as2b^{FAM}	uag agg uac gug c**T***g agg cTT
s2b	gcc uca gca cgu acc ucu aTT
s2b^{FAM}	gcc uca gca cgu acc ucu aT**T***
s2b-m1	gcc uca gac agu acc ucu aTT
s2b-m2	gcc gag gca cgu acc ucu aTT
s2b-m3	cug uca gca cgu acc ucu aTT
s2b-m4	gac gca gac agu acc ucu aTT
s2b-sm	cug cgg ugc gau acc ucu aTT
s2b-nsm	gcc uca gca gcg gga aua uTT

DNA	Sequence (5´ – 3´)
target^{FAM}	GCG CAA TAC CAT TTT TTA CAA ATT GAG TTA **T***
non-target	ATA ACT CAA TTT GTA AAA AAT GGT ATT GCG C
nt-8	GTA AAA AAT GGT ATT GCG C
nt-12	ATT TGT AAA AAA TGG TAT TGC GC
nt-16	CTC AAT TTG TAA AAA ATG GTA TTG CGC
non-target^{FAM}	ATAACTCAATTTG**T***AAAAAATGGTATTGCGC
target	GCG CAA TAC CAT TTT TTA CAA ATT GAG TTA T
t8	GCG CAA TAC CAT TTT TTA C
t12	GCG CAA TAC CAT TTT TTA CAA AT
t16	GCG CAA TAC CAT TTT TTA CAA ATT GAG
target-noPAM^{FAM}	GCG CAA TAA ATT TTT TTA CAA ATT GAG TTA **T***
nt-noPAM	ATA ACT CAA TTT GTA AAA AAT TAT ATT GCG C
sgRNA template	GGA UAA CUC AAU UUG UAA AAA AGU UUU AGA GCU AGA AAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AAC UUG AAA AAG UG
T7 promoter	TAA TAC GAC TCA CTA TAG GG

Boldface letters indicate the position of the fluorescence label: Asterisk marks position of C6-linked 5/6-FAM; Uppercase letters represent deoxynucleotides.

3.4.2. Primers

Name	Sequence (5´ – 3´)
hAgo1-fwd AflII	CCC CTT AAG ATG GAA GCC GGC CCC TCC
hAgo1-rev AgeI Stopp	CCA CCG GTT CAA GCA AAG TAC ATG GTG CGC AG
hAgo2-fwd AflII	TGA CTA GCT TAA GAT GTA CTC GGG AGC CGG CC
hAgo2-rev AgeI Stopp	TCA GTC AAC CGG TTC AAG CAA AGT ACA TGG TGC GCA GAG
N-Ago2-rev	TGC CTT GTA AAA CGC TGT TGC TGA CAC
PAZ-Ago1-fwd	GCA ACA GCG TTT TAC AAG GCA CAG
PAZ-Ago1-rev	GTT ACA GAC CTC CAG GGG AAG GTA GGT
MID-PIWI-Ago2-fwd	CTG GAG GTC TGT AAC ATT GTG GCA GGA CAA
N_ MID-Ago2-rev	CAG GAT GTT GTT CAC GCC TCC CAG
PIWI-Ago1-fwd	AAC AAC ATC CTG GTC CCC CAC CAG
N-ago1-rev	GAA GAA GGA GCG GCC CAC GGG
PAZ_PIWI-ago2-fwd	CGC TCC TTC TTC ACC GCG TCC GAA
PAZ-1-rev	ATT GCA GAC ACG CTA CTT CCT
PAZ-2-fwd	TAC CGT GTC TGC AAT GTG ACC

3.4.3. Plasmids

Name	Supplier
pET41b(+)-TEVThroligo-GST-hAgo2	Dr. A. Deerberg (Institut für Molekulare Medizin, Universität zu Lübeck)
pMJ806	Addgene (Depositing Lab: Jennifer Doudna)
pMJ841	Addgene (Depositing Lab: Jennifer Doudna)

3.5. Bacterial culture

3.5.1. Bacterial strains

Strain	Genotype
E. coli BL21(DE3)	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])
E. coli DH5a	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, $hsdR17(r_K^- m_K^+)$, λ–
E. coli BL21-Gold(DE3)	E. coli B F- dcm+ Hte ompT $hsdS(r_B- m_B-)$ gal λ (DE3) endA Tet^r
E. coli BL21 Rosetta(DE3)	F^- ompT $hsdS_B(R_B^- m_B^-)$ gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cam^R)

3.5.2. Bacterial culture media

Name	Components
Luria-Bertani (LB)-medium (1x, pH 7.4)	0.5% (w/v) Yeast extract, 1.0% (w/v) Tryptone, 1.0% (w/v) NaCl
SOB-medium	0.5% (w/v) Yeast extract, 2% (w/v) Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$
SOC-medium	SOB Medium, 20 mM Glucose
TFB-medium (1x, pH 7.4)	0.4% (w/v) Glycerin, 2.4% Yeast extract, 1.2% (w/v) Tryptone, 1x Phosphate buffer
2xTY medium	1.0% (w/v) Yeast extract, 1.6% (w/v) Tryptone, 0.5% (w/v) NaCl
LB-Agar plates	LB medium, 18 g/l Agarose

3.6. Software

Name	Source
Clone Manager 7.04	SciEd Central
DataMax 2.20	Horiba Jobin Yvon
Fusion 15.09	Vilber Lourmat, Eberhardzell
GATC Viewer	Biotech AG
GraFit 5.0.6	Erithacus Software
Image Quant 5.2	Molecular Dynamics, Sunnyvale
Inkscape 0.91	https://inkscape.org
ProData SX 2.0.3	AppliedPhotophysics
ProData Viewer 4.0.17	AppliedPhotophysics
Scientist	Micromath
Chimera 1.10.1	UCSF
Mfold	RNA Institute, University at Albany (NY, USA)
ExPASy	Swiss Institute of Bioinformatics

4. Methods

4.1. Determination of nucleic acid or protein concentration

4.1.1. Nucleic acids

Nucleic acids were quantified by measuring the absorbance at 260 nm (A_{260}) using a BECKMAN DU640 spectrophotometer or a NanoDrop spectrophotometer (allows the analysis small volumes up to 1 µl of the sample). The concentration of nucleic acids is determined by Lambert-Beer's law as described below:

$$c = \frac{A_{260}}{\in \cdot d}$$

In this formula, "A" represents the absorbance, ∈ is the molar extinction coefficient, *c* is the concentration and d the optical path of the cuvette (for BECKMAN DU640: quartz cuvettes with 1 cm width were used; for NanoDrop: no cuvette used but software provides normalization of data according to an assumed width of 1 cm.). Therefore, for nucleic acids with known sequence and accordingly known molar extinction coefficient the concentration was calculated by using the derived equation. Alternatively, the following approximate conversions were used:

Nucleic acids: 1 A_{260} Unit=50 µg/ml dsDNA or 33 µg/ml ssDNA or 40 µg/ml ssRNA

Moreover, the purity of DNA or RNA solutions was evaluated by considering the ratio A_{260}/A_{280} that was ≥ 1.8 upto ≥ 2.0.

4.1.2. Proteins

The concentration of proteins was measured by using the same method but absorbance at 280 nm. The absorbance at 280 nm of a protein is almost totally a function of the content of the aromatic rings of the amino acids Tyrosine and especially Tryptophan (the aromatic ring of Phenylalanine absorbs well at 260 nm, but not 280 nm). Calculation of the absorbance of a 1 M solution of a protein, the total number of Tryptophan residues was multiplied with the absorbance of a M solution of Tryptophan. This number was added to the quantity of Tyrosine residues multiplied by the absorbance of a one molar solution of Tyrosine, whereas the sum of both is the absorbance of 1 M of the whole molecule:

Extinction coefficient = (Number of Tryptophan x 5500) + (Number of Tyrosine x 1490)

Pure protein was measured at 280 nm absorbance and the measured value was divided by the absorbance value obtained from a reference solution with a concentration of 1 mg/ml. the final unit retained by this calculation is mg/ml.

4.2. Isolation of plasmids from *E. coli* cells

E. coli cells from 5 ml LB medium were harvest after overnight cultivation and the GeneEluate Plasmid Miniprep Kit was used to isolate the plasmid DNA according to manufacturer's instructions. The concentration of purified plasmids was measured by NanoDrop 1000 and the plasmid was identified by using PCR or analysis of excised inserts.

4.3. Polymerase chain reaction (PCR)

4.3.1. Standard PCR

For standard PCR *Taq* or Phusion® DNA Polymerase was used. All reaction components were assembled on ice and mixed gently, the liquid was collected to the bottom of the tube by a quick spin, then PCR tubes were transferred to a PCR cycler with the block preheated to 95°C and thermocycling was started.

PCR using *Taq* Polymerase 50 µl

Component	Concentration
Template Plasmid	10 ng
Forward Primer	200 nM
Reverse Primer	200 nM
dNTPs	200 µM
Standard *Taq* buffer	1x
Taq Polymerase	1.25 u

Thermocycling conditions for a *Taq* PCR

Step	Temp	Time
Initial Denaturation	95°C	30 seconds
30 Cycles	95°C 60-68°C 68°C	30 seconds 30 seconds 1.5 min (1 minute/kb)
Final Extension	68°C	5 minutes
Hold	4°C	

PCR using Phusion® High-Fidelity Polymerase

Component	Concentration
Template Plasmid	10 ng
dNTPs	200 µM
Forward Primer	200 nM
Reverse Primer	200 nM
5x Phusion HF Buffer	1x
DMSO	3%
Phusion DNA Polymerase	1.25 u

Thermocycling conditions for a Phusion PCR

Step	Temp	Time
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C 65-72°C 72°C	10 seconds 20 seconds 2 min 30 s (15-30 seconds per kb)
Final Extension	72°C	10 minutes
Hold	4°C	

4.3.2. Generation of chimeras and overlap extension PCR

Overlap extension PCR was used to insert specific mutations at specific positions in a sequence or to combine smaller DNA segments to a larger polynucleotide. Here, it was used to swap different domains from hAgo1 to hAgo2. For primers used see section 3.4.2.

Frist round PCR

Component	Concentration
Template Plasmid	10 ng
dNTPs	200
Forward primers or chimeric forward primer	200 nM
Reverse primer or chimeric reverse primer	200 nM
5x Phusion HF Buffer	1x
Phusion DNA Polymerase	1.25 u

Step	Temp	Time
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C 65-72°C 72°C	10 seconds 20 seconds 2 min 30 s (15-30 seconds per kb)
Final Extension	72°C	10 minutes
Hold	4°C	

Second round PCR

Component	Concentration
purified segments from the 1st round	10 ng
dNTPs	200 µM
Forward primer	200 nM
Reverse primer	200 nM
5x Phusion HF Buffer	1x
Phusion DNA Polymerase	1.25 u

Step	Temp	Time
Annealing without primers 5 Cycles	98°C 65°C 72°C	10 seconds 20 seconds 2 min
forward and reverse primers Added		
Initial Denat ration	98°C	30 seconds
30 Cycles	98°C 65-72°C 72°C	10 seconds 20 seconds 2 min 30 s (15-30 seconds per kb)
Final Extension	72°C	10 minutes
Hold	4°C	

4.4. Nucleic acid electrophoresis

4.4.1. Agarose gel electrophoresis

To analyze plasmids, PCR or purified DNA segments, the agarose gel electrophoretic technique applying non-denaturing conditions was used. The samples were mixed in DNA loading buffer (see section 3.3.1) and loaded onto an agarose gel that was previously assembled by dissolving the required amount of agarose in 1x TAE buffer followed by heating for about 5 min at 800 W in a

microwave oven. The amount of agarose and the relative volume of buffer were determined depending on the percentage of the agarose gel required for the size separation of dsDNA segments to be performed, as described in Table 4.1.

Table 4.1 Agarose concentration and relative size range of nucleic acids

Agarose (%, w/v)	Linear dsDNA sizes (bp)
0.8	800 - 10,000
1.0	500 - 10,000
1.2	400 - 7,000
1.5	200 - 3,000

The agarose solution was poured into a gel casting chamber containing 0.004% (w/v) EtBr that is required for band detection, since it intercalates into the double strand of DNA or binds to helical secondary structures of RNA. The solution solidified for about 30 min at room temperature. Electrophoresis was done horizontally in 1x TAE buffer at 10V/cm for 20-60 min depending on the gel size. The nucleic acid bands were visualized under ultraviolet (UV) light irradiation (312 nm).

4.4.2. Polyacrylamide gel electrophoresis (PAGE)

The PAGE technique was used to allow a fine-resolution separation of nucleic acids. The solution for the polyacrylamide gel was filtered by using a 0.2 µm filter applying low pressure and stored at 4°C or immediately used after the addition of APS for polymerization and TEMED as catalyst. The solution was quickly poured into a vertical gel chamber (20 x 40 cm or 15 x 13 cm) which was assembled before by using two glass plates separated by a 0.4 or 1 mm spacer. The gel polymerized at RT for 20 min to.

4.4.2.1. Denaturing PAGE

The analysis of single strand oligonucleotides was carried out by using denaturing PAGE. To prepare the denaturing gel, a volume of 20 ml (for gel measures 15 x 13

cm) or 40 ml (for gel measures 20 x 40 cm) solution was required and it was made by 1x TBE, 7 M Urea and an appropriate volume of acrylamide/bisacrylamide (ratio: 19/1) to achieve the concentration suitable for separating segments of specific length as described in Table 4.2. Then the solution was mixed and APS and TEMED were added to a concentration of 0.05% (w/v) and 0.1% (v/v), respectively. Electrophoresis of small gels was done at 200 V and for sequencing gels 70 W at 50°C was applied.

Table 4.2 Correlation of concentration of acrylamide and relative nucleic acid segment length

Gel Concentration (% w/v)	Effective range of separation (bp)	Xylene Cyanol* (bp)	Bromophenol Blue* (bp)
8.0	60-400	160	45
12.0	40-200	70	20
15.0	25-150	60	15
20.0	6-100	45	12

***** XC or BPB relative to the molecular weight of DNA segments

4.4.2.2. Non-denaturing PAGE

Non-denaturing PAGE was used to analyze annealing of DNA or RNA oligonucleotides. The gel solution was prepared as described in 4.4.2.1 denaturing PAGE except for the addition of urea, then the samples were loaded including native loading buffer (see section 3.3.1) and electrophoresis was done using 200 V and 4°C for ca. 2 h.

4.4.2.3. PAGE purification of RNA

Denaturing PAGE for sequencing was used to purify RNA products produced by *in vitro* transcription reaction. After large scale *in vitro* transcription (5-10 ml), the reaction solution was mixed with loading buffer and loaded to a 12% denaturing polyacrylamide gel. Using 1xTBE electrophoresis buffer PAGE was done applying

300 V for 15 h. The bands were visualized by UV shadowing on a Thin-layer chromatography (TLC) plate using a hand-held UV light. Subsequently, the band of interest was cut out and electroeluted with a Schleicher & Schuell Biotrap electroseparation system (TBE buffer, 200 V). The isolated nucleic acids were collected every hour and the concentration was determined.

4.5. Protein electrophoresis and Western blot

4.5.1. SDS-PAGE

Proteins were analyzed by SDS-PAGE composed of 4% (v/v) stacking gel and a resolving gel whose percentage of acrylamide was choosen according to the size of the proteins to be analyzed (Tabel 4.3). In this study 8% (v/v) SDS-PAGE was performed on a vertical gel casting system in 0.75 mm thick gels. Cell pellets were resuspended in 1x PBS and 2x loading buffer at a ratio 1:1 (v/v).

Tabel 4.3 Gel composition for preparing SDS-PAGE

	Stacking gel 4% (w/v)	Resolving gel 8% (w/v)
Rotiphorese gel 30% (v/v)	1.4 ml	1.3 ml
1.5 M Tris (pH 8.8)	/	1.3 ml
1.0 M Tris (pH 6.8)	250 µl	/
10 % (w/v) SDS	20 µl	50 µl
H_2O	1.4 ml	2.3 ml
TEMED	2 µl	5 µl
10 % (w/v) APS	20 µl	50 µl

The samples were heated at 95°C for 5 min to ensure protein denaturation and loaded on the gel. An adequate size standard without denaturation (see section 3.3.3) was loaded as well. The gel was initially run at a constant voltage of 80 V for 30 min to let the proteins reach the interface between the stacking and resolving gel and followed by electrophoresis with the resolving gel at 120 V for 90 min at RT by using 1x running buffer.

4.5.2. Western blot analysis

Western blotting uses specific antibodies to identify proteins. It was used to identify recombinant Argonaute proteins and chimeras. Proteins were transferred from the gel to a PVDF (polyvinylidene fluoride) membrane by using the “semi-dry” method. The membrane was further processed with specific antibodies, and visualized using secondary antibodies and detection reagents.

First the Immobilon-P membrane was activated in methanol for 1 min followed by washing in Millipore water for 5 min. Subsequently, the membrane and the gel were equilibrated in 1x transfer buffer and the blotting unit was assembled as follows:

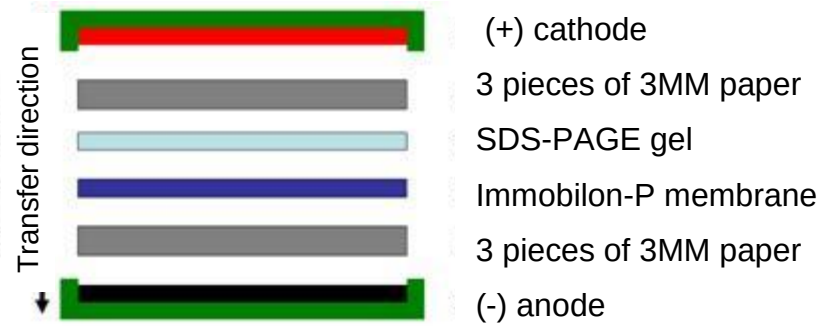


Figure 4.1 Assembling unit for Western blot

The blotting was performed at 0.8 mA/cm^2 for 90 min at RT. The transfer time is related to the MW of proteins. Prior to the detection of proteins, the membrane was blocked by incubating with 5% (w/v) milk powder in 1x T-TBS for 1-2 h at RT to avoid unspecific binding of the antibodies to the PVDF membrane. The membrane was washed two times with 1x T-TBS and then incubated overnight with the specific primary antibody in blocking buffer. Three wash steps were required to remove unbound antibody and the membrane was further incubated for 1 h at RT under shaking with a corresponding secondary antibody conjugated with horseradish peroxidase (HRP). An enhanced chemiluminescent (ECL) system was used for the detection of the secondary antibody according to manufacturer’s

instructions. The detection and imaging of ECL chemical signal was performed by using FUSION SOLO from Vilber Lourmat.

4.6. Molecular cloning

Cloning by restriction enzyme cut and T4 DNA ligase mediated ligation is a simple and easy way of moving a segment of double stranded DNA from one plasmid to another.

4.6.1. Restriction enzyme digestion

Restriction enzymes (RE) function by cutting double stranded DNA at specific 4 to 8 base pair inverted repeat recognition sequences within the target DNA. The products of DNA cleavage are either blunt ended or contain 5´- or 3´-overhangs. The insert DNA (PCR product) and vector are both digested by the same two restriction enzymes and then purified by agarose gel electrophoresis to remove residual nicked. *Afl* II and *Age* I were used for cloning Argonaute.

4.6.2. Ligation

The next step is the ligation of the insert into the linearized vector. This involves the formation of phosphodiester bonds between adjacent 5´-phosphate and 3´-hydroxyl residues, which can be catalyzed by T4 DNA ligase. The DNA concentration to obtain good results is approx. 1 nM of pET vectors, and the molar ratio between vector and insert DNA is 1:3 or 1:5.

4.6.3. Preparation of chemically competent cells

The appropriate *E. coli* strain was incubated in 5 ml LB medium for overnight culture at 37°C, then the overnight culture was added to 500 ml SOB medium and incubated at 25-30°C until the absorbance at 600 nm was approx. 0.5. Then the culture was put on ice to chill for at least 10 min, the cell suspension was centrifuged for 10 min at 2,500 g, and the pellet was resuspended gently in 100 ml ice-cold CCMB80 buffer. After the cell suspension incubated on ice for 10 min, it was centrifuged again for 10 min at 2,500 g, then the pellet was resuspended in 10

ml ice-cold CCMB80 buffer and the cell suspension was aliquoted to 200 tubes (50 µl per tube) and stored at -80°C.

4.6.4. Transformation of *E. coli* cells

Plasmid DNA or ligation products were transformed into *E. coli* DH5α or BL21 by using the heat shock method. It consists of inserting a foreign plasmid or ligation product into bacteria. Briefly, add ~200 ng of plasmid DNA to 50 µl chemically competent cells and incubate on ice for 20 min. Heat-shock cells by incubation at 42°C for 30-60 seconds and then place back in ice. Then SOC media is added and the transformed cells are incubated at 37°C for 60 min in a shaking incubator. Transformed cells carried the gene encoding for Kanamycin resistance, therefore LB agar plates containing 30 µg/ml of Kanamycin were used to plate 200 µl of transformed cell suspension. Bacteria were then incubated at 37°C overnight to allow the formation of colonies.

4.6.5. Selection of clones

To analyze the clones from the transformation, enzymatic restriction and PCR were performed. Firstly, pick one colony from the agar plate to inoculate 50 ml LB medium containing 50 µg/ml Kanamycin and incubate the pre-culture at 37°C in a shaking incubator (~200 rpm) for a minimum of 4–5 h or overnight. Then extract plasmid from the cultured clones, use restriction digestion to check the restriction sites and PCR to confirm the insert DNA. The vector then will be sent for sequencing to make sure no mutation in the insert.

4.7. Protein expression and purification

4.7.1. Protein expression

The selected plasmid extracted from DH5α was transformed into chemically competent BL21 cells for protein expression (see section 4.5.4). Cells were grown to A_{600} of 1.0 at 37°C in TFB medium, induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and let grow for another 2 h at 37°C. In case of Cas9, cells were grown in 2xTY medium at 37°C. At an OD_{600} of ~0.8, the

temperature was reduced to 18°C, and protein expression was induced by the addition of 200 nM IPTG, followed by shaking at 18°C overnight (12–16 h). The cells were harvested by centrifugation for 10 min at 3,000 g, and the cell pellets were lysed by sonication using ice-chilled lysis buffer.

4.7.2. Protein purification

The purification was performed by using a FPLC system at 4°C. For columns, column materials and buffers used see section 3.2.3 and 3.3.1. All purification steps were analyzed by SDS-PAGE.

Since recombinant Argonaute protein carries a glutathione S-transferase (GST) tag, it was purified by using Glutathione Sepharose chromatography (128). The *E. coli* lysate was loaded to a Glutathione Sepharose High Performance column overnight (at 0.1 ml/min). Next, the column was washed with 5-10 fold of column-volume, then the GST fusion proteins were eluted by using elution buffer containing 20 mM reduced glutathione. The eluted fractions were collected and analyzed by 8% SDS-PAGE.

For Cas9, the fusion construct contained an N-terminal hexahistidine-maltose binding protein (His_6-MBP) tag, followed by a peptide sequence containing a tobacco etch virus (TEV) protease cleavage site. Thus, the protein was purified by a combination of affinity and size exclusion chromatographic steps. The lysate was bound to Ni-NTA agarose packed in a XK 16/20 column (0.5 ml/min). The resin was washed with wash buffer ~50 ml at 1.5 ml/min and the bound protein was eluted in elution buffer containing 200 mM imidazole (0.5 ml/min). The concentration of purified protein was estimated by measuring the absorbance at 280 nm, then TEV protease was added (0.5 mg TEV protease per 100 mg of protein) and incubated overnight at 4°C to remove the His_6-MBP affinity tag. Then the protein was concentrated by using a 100,000 MWCO centrifugal concentrator and loaded on a Superdex 200 26/60 column in SEC buffer (1 ml/min). The peak fractions were collected and analyzed by SDS-PAGE. So the cleaved Cas9 was separated from the fusion tag by size exclusion chromatography.

4.8. Nucleic acid preparation

4.8.1. *In vitro* transcription

Firstly, DNA template and T7 promoter oligonucleotide were annealed at 90°C for 5 min and the mixture was slowly cooled down to room temperature. Then the transcription reaction was set up by mixing the reagents according to Table 4.4. For sgRNA, a 5 ml reaction volume was chosen, which is usually sufficient for 1-2 mg of pure RNA.

Table 4.4 *In vitro* transcription reaction

	Stock concentration	Final concentration	Volume (ml)
H_2O	/	/	1.95
transcription buffer	10x	1x	1.00
ATP	100 mM	5 mM	0.25
UTP	100 mM	5 mM	0.25
GTP	100 mM	5 mM	0.25
CTP	100 mM	5 mM	0.25
DTT	1 M	10 mM	0.05
hybridized template	10 µM	1 µM	0.50
T7 RNA polymerase	1 mg/ml	0.1 mg/ml	0.50
Total volume			5.00

The reaction was incubated at 37°C for 1.5 h and then stopped by adding EDTA. Centrifuge the mixture for 5 min at 3000 g and 4°C to pellet magnesium pyrophosphate and remove the supernatant containing transcribed RNA. Add 5 ml 2x RNA loading dye to the sample. Purify the RNA by electrophoresis on a pre-warmed 8% polyacrylamide, 7 M urea denaturing gel in 0.5 x TBE until the bromophenol blue dye reaches the lower quarter of the gel. Subsequent steps of purification are described in section 4.4.2.3.

4.8.2. Nucleic acid modifications

DNAs and RNAs were 5´-phosphorylated using T4 Polynucleotide Kinase (PNK), either with [γ-^{32}P] ATP or unlabeled ATP. Nucleic acids with a 5´-phosphate can be dephosphorylated by Calf Intestinal Phosphatase (CIP). The labeling or removing of 5´-phosphate follows the protocol of PNK or CIP kit manual. In all cases, modified RNAs were purified by using phenol/chloroform extraction, Sephadex G50 columns and ethanol precipitation. The ^{32}P labeled oligonucleotide can be detected by using autoradiography.

4.8.3. Phenol/chloroform extraction

Phenol/chloroform extraction is used to remove proteins from nucleic acid samples and can be carried out in a manner that is very close to quantitative. Nucleic acids remain in the aqueous phase and proteins separate into the organic phase or lie at the phase interface. Add 400 µl prepared phenol/chloroform/isoamyl alcohol (PCI) solution (25:24:1) to 100 µl reaction solution, vortex vigorously to mix the phases and spin in a microfuge at top speed for 1–2 min to separate the phases. Remove the aqueous phase which contains the nucleic acids to a new tube or continue purification by using Sephadex G50 columns.

4.8.4. Ethanol precipitation

Ethanol precipitate nucleic acids by adding 40 µl 3 M $NaCH_3COOH$ and 1000 µl 100% ethanol to 400 µl DNA or RNA solution, vortex, and place the sample at -20°C for 2 h or -80°C for 30 min. Spin the mixture at the highest speed setting in a centrifuge at 4°C for 15 min. Carefully withdraw the supernatant and discard. Wash the pellet once with ice cold 70% ethanol. Allow the remaining liquid to evaporate.

4.8.5. Nucleic acids hybridization

For pre-annealed DNA or RNA duplexes, both strands were incubated at equimolar amounts for 5 min at 90 °C, followed by slowly cooling down to 25°C in 15 mM Hepes, pH 7.4, 50 mM KCH_3COOH, and 1 mM $MgCH_3COOH$. Hybridization was tested by running native PAGE at 4 °C.

4.9. Cleavage assays

4.9.1. hAgo-RNA cleavage assay

Ago slicer assays were performed in cleavage buffer containing 2 mM $MgCl_2$ and 1 µg/mL RiboLock RNase Inhibitor. Unless specified, otherwise 3.7 µM Ago protein was incubated with 100 nM ss-guide RNA (as2b) for 10–15 min at room temperature, and the cleavage reaction was started by adding 2.5 nM radiolabeled target RNA to the reaction mix and incubated at 37°C for up to 2h. Samples were removed at different time points and reactions were stopped by addition of 1 vol. RNA loading dye, followed by sequencing PAGE analysis and autoradiography.

4.9.2. Cas9 cleavage assay

The guide RNA was annealed by heating to 90°C for 5 min and slowly cooled down to room temperature. The reaction was set up according to Table 4.5. Cleavage reaction was started by adding the DNA target (pre-annealed) to the reaction mix and incubated at 37°C. 10 µl aliquots were removed at different time points and quenched by adding DNA loading buffer. Cleavage products were analyzed by 16% denatured PAGE and the gels were scanned by using Typhoon FLA9500.

4.9.2.1. Single-turnover assay

Cas9 (25 nM) was pre-incubated for 10 min at 37°C in cleavage buffer with pre-annealed sgRNA and the reaction was started by adding target DNA duplex (5 nM). The reaction mix was incubated at 37°C. At defined time intervals, samples were withdrawn from the reaction and stopped by adding DNA loading buffer (containing 1% SDS and 20 mM EDTA). Cleavage was monitored by 16% denaturing PAGE gel electrophoresis and scanned by using Typhoon FLA9500. Percentage of cleavage was analyzed by FAM fluorescence and the average of three independent experiments was plotted against time. The data were fitted to a nonlinear single exponential equation and the cleavage rate constant was calculated by using the program GraFit 5: Cleavage % = Limit * (1-exp(-*k**t)), where Limit is the maximal percentage of cleavage, *k* is the cleavage rate of single turnover, and t is the time.

Table 4.5 Endonuclease activity assay

	Stock	Final concentration	Volume (µl)
DEPC–H_2O			42
Cleavage buffer	5x	1x	20
Cas9	30 µM	5 µM	16
Guide RNA	50 µM	5 µM	10
Incubate for 10–15 min at room temperature. Then add annealed DNA duplex.			
Duplex substrate	400 nM	20 nM	12
Total volume			100 µl

4.9.2.2. Multiple-turnover assay

Cas9 (1 nM) was pre-incubated 10 min at 37°C in cleavage buffer with pre-annealed sgRNA. The reaction was started by addition of target DNA duplex (5 nM). At defined time intervals, samples were withdrawn and the reaction was stopped by adding DNA loading buffer. The cleavage reaction was resolved by 16% denatured PAGE gel electrophoresis, scanned by using Typhoon FLA9500 and the percentage of cleavage was analyzed by FAM fluorescence. The data were fitted to a nonlinear double exponential equation and the cleavage rate constants were calculated by using the program GraFit 5: Cleavage % = A1 * (1- exp(-k_1 * t)) + A2 * (1-exp(-k_2 * t)), where A1 and A2 are the amplitudes corresponding to the first and second phase, k_1 and k_2 are the cleavage rates of single turnover and multiple turnover, and t is the time.

4.10. Protein/nucleic acid binding analysis

4.10.1. Equilibrium fluorescence titrations

A FluoroMax-3 fluorescence spectrometer (Figure 4.2) was used to determine the binding affinity. The binding affinities of binary or ternary complexes of protein and nucleic acids were determined in a 700 µL cuvette in binding buffer. The data were recorded by the program DataMax 2.20 and analyzed by GraFit 5.

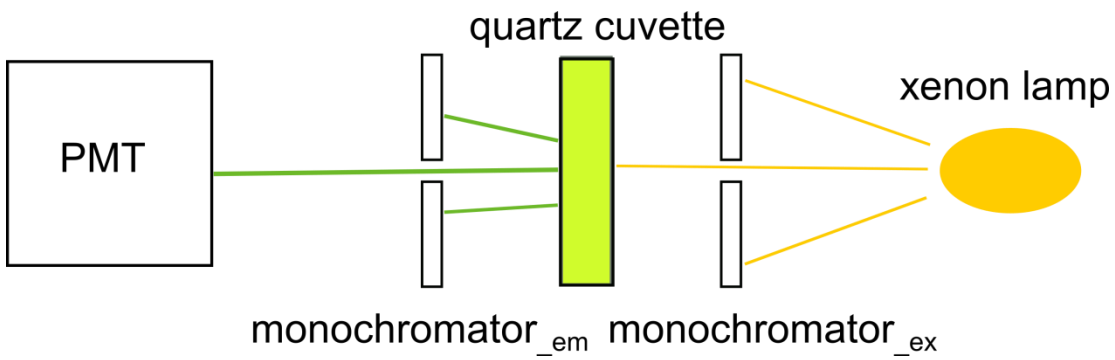


Figure 4.2 Set-up of the fluorescence spectrometer Fluoromax-3. A xenon lamp is using to excite the fluorophor. The emitted fluorescence is detected with a photomultiplier tube (PMT). The reaction takes place in the cuvette and a change in fluorescence is caused by addition of the binding partner.

For Ago/RNA binary complex formation, FAM-labeled guide RNA as2b^{FAM} (20 nM) was titrated with increasing concentrations of Ago. For ternary complexes, pre-incubated Ago and guide [600 nM Ago/20 nM 5´-phosphorylated guide RNA (as2b^{FAM})] were titrated with increasing concentrations of target RNA.

In case of Cas9, FAM-labeled DNA targets (20 nM) were titrated with increasing concentrations of Cas9-sgRNA (pre-incubated binary complex of Cas9 and sgRNA). The binding buffer contained 0 mM or 2 mM $MgCl_2$.

The samples were excited at 492 nm, the emission intensity was measured at 520 nm, and the change in fluorescence signal was recorded. The experimental data were fitted to a quadratic equation by using the program GraFit 5: Fluorescence = F_{max} −((c[guide] + L + K_d) − sqrt((sqr(c[guide] + L + K_d)) − 4*c[guide]*L))*(F_{max} − F_{min})/ ((2*c[guide])), where F_{max} is the maximum fluorescence, F_{min} is the minimum fluorescence, c[guide] is the concentration of substrate, L is the concentration of titration partner, and K_d is the equilibrium dissociation constant.

4.10.2. Pre-steady-state stopped-flow measurements

The association and dissociation of protein and nucleic acid complex formation were measured by fluorescence signal change in binding buffer by using the

stopped-flow instrument SX20 (Figure 4.3). Excitation of the FAM-labeled RNAs was at 492 nm with slits set to 1 nm (equivalent to a wavelength bandwidth of 4.65 nm) and detection was performed through a filter with a cutoff at 530 nm.

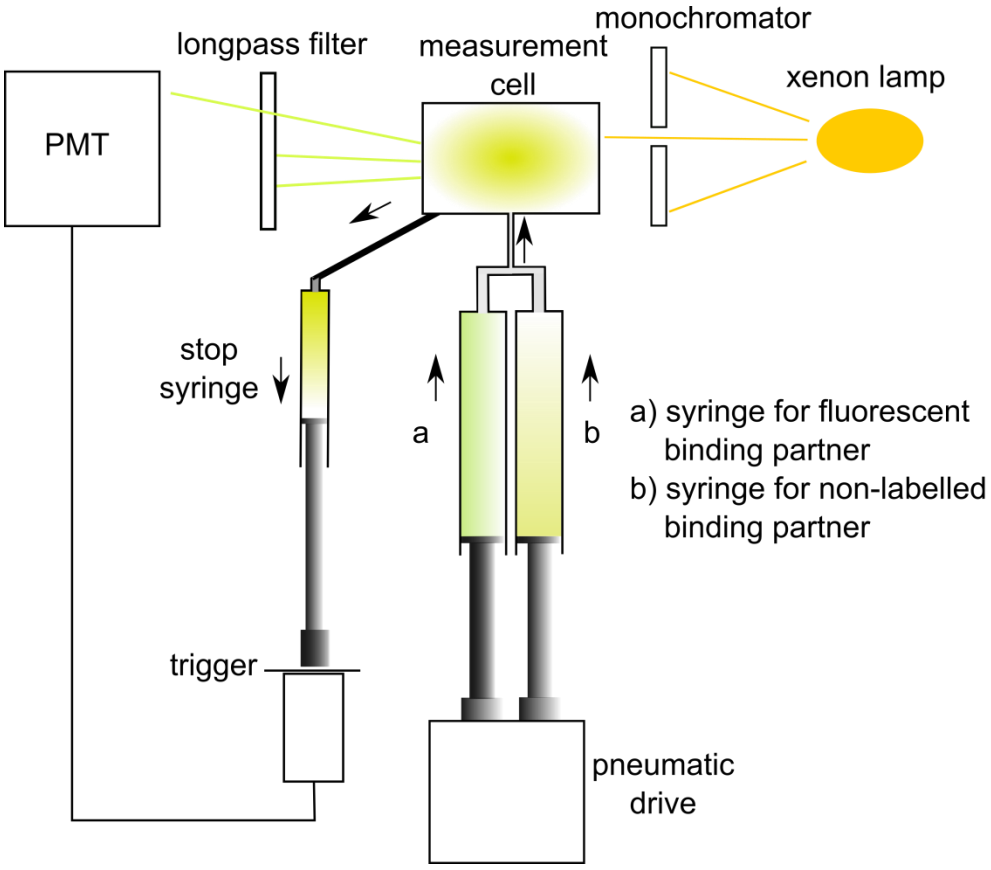


Figure 4.3 Set-up of the stopped-flow instrument. The binding partners are isolated in two different syringes. One of the binding partners carries a fluorophor. The fluorescence is excited by the light of a xenon lamp. When the syringes are pushed up, a defined volume (approx. 50 µl) enters the measurement cell. The reaction is initiated by rapid mixing of the samples, while the stop syringe ensures only a small amount of solution enters the measurement cell and triggers fluorescence detection by the photomultiplier tube (PMT). In principle, a fluorescent probe is employed to follow the course of the reaction by recording changes in the amplitude of the spectroscopic signal as a function of time. A exemplary limit for reaction rates that can be accurately measured with a stopped-flow is in the range of > 500 s^{-1} with a dead-time of > 2ms. Modified with kind permission of Dr. S. Willkomm (128).

In the case of binary complexes of Ago and RNA, 20 nM FAM-labeled guide RNA or ds RNA was rapidly mixed with Ago (300–800 nM) in binding buffer (contained

0.5 mM $MgCl_2$). For dissociation of the binary complexes, 20 nM FAM-labeled guide RNA was pre-incubated with 300 nM hAgo2 and then rapidly mixed with a 30 or 100-fold excess of non-labeled competitor RNA. For the analysis of ternary complexes (Ago and guide/target RNA substrates), 20 nM 5´-phosphorylated guide RNA was pre-incubated with 600nM Ago in binding buffer and then rapidly mixed with target RNA (20–60 nM). For dissociation of the ternary complexes, 20 nM FAM labeled guide RNA, 600 nM Ago and 60 nM target RNA were pre-incubated and then rapidly mixed with a 33-fold excess of nonlabeled competitor guide RNA.

For Cas9 complex formation, 20 nM FAM-labeled DNA substrate was rapidly mixed with Cas9-sgRNA (50–300 nM) in binding buffer (containing either 0 or 2 mM $MgCl_2$). In the case of dissociation, 20 nM FAM labeled target DNA and 60 nM Cas9-sgRNA were pre-incubated and then rapidly mixed with 30 or 100-fold excess of nonlabeled DNA as competitor.

The change in fluorescence signal was recorded over time by using either the stopped-flow instrument SX20 or the FluoroMax-3 fluorescence spectrometer. In all cases, reaction rates k_{+n} and k_{-n} were calculated by fitting the experimental data to an exponential equation by using the program GraFit 5: Fluorescence = $\sum A_n*\exp(-k_n*t)$, where A_n is the amplitude corresponding to the observed phase, k_n is the rate constant of the observed phase, and t is the time.

4.11. Sequence and structure analysis

4.11.1. Multiple alignments of protein sequences

DNA or protein sequence alignments were used to identify functional and evolutionary relationships of Argonautes. The similarity of different Agos was analyzed by using BLAST and ExPASy. The sequence data were obtained from Genbank and Protein database of NCBI.

4.11.2. B factor analysis

B factor is also known as temperature factor or Debye-Waller factor. It is the displacement of the atomic position from its mean position. B factor is an indicator of the relative vibrational motion or the disorder of an atom in a protein crystal. A low B factor implies that the atom is in well-ordered parts of the structure, while a large B factor generally suggests a very high flexibility of this atom.

The B factors obtained from PDB files were normalized as described using the equation $B_{normalized}$ = (Cα B factor − B_{mean})/σ(B), where B_{mean} and σ(B) are the mean value and standard deviation, respectively, of the distribution of observed thermal factors (129). All crystal structures were obtained from Protein Data Bank (PDB) and B-factors were analyzed and presented with Chimera 1.10.1.

5. Results

5.1. Preparation of hAgo1

5.1.1. Plasmid construction and expression of hAgo1 and chimeras.

For cloning and expression of recombinant hAgo1 in *E. coli*, the pET System (Novagen) was used. The sequence of hAgo1 was amplified via high-fidelity PCR from hAgo1 pBluescript II SK (+) (MWG operon). After restriction with *Afl* II and *Age* I, full length DNA sequence of hAgo1 was ligated into a pET41b (+) vector (Figure 5.1).

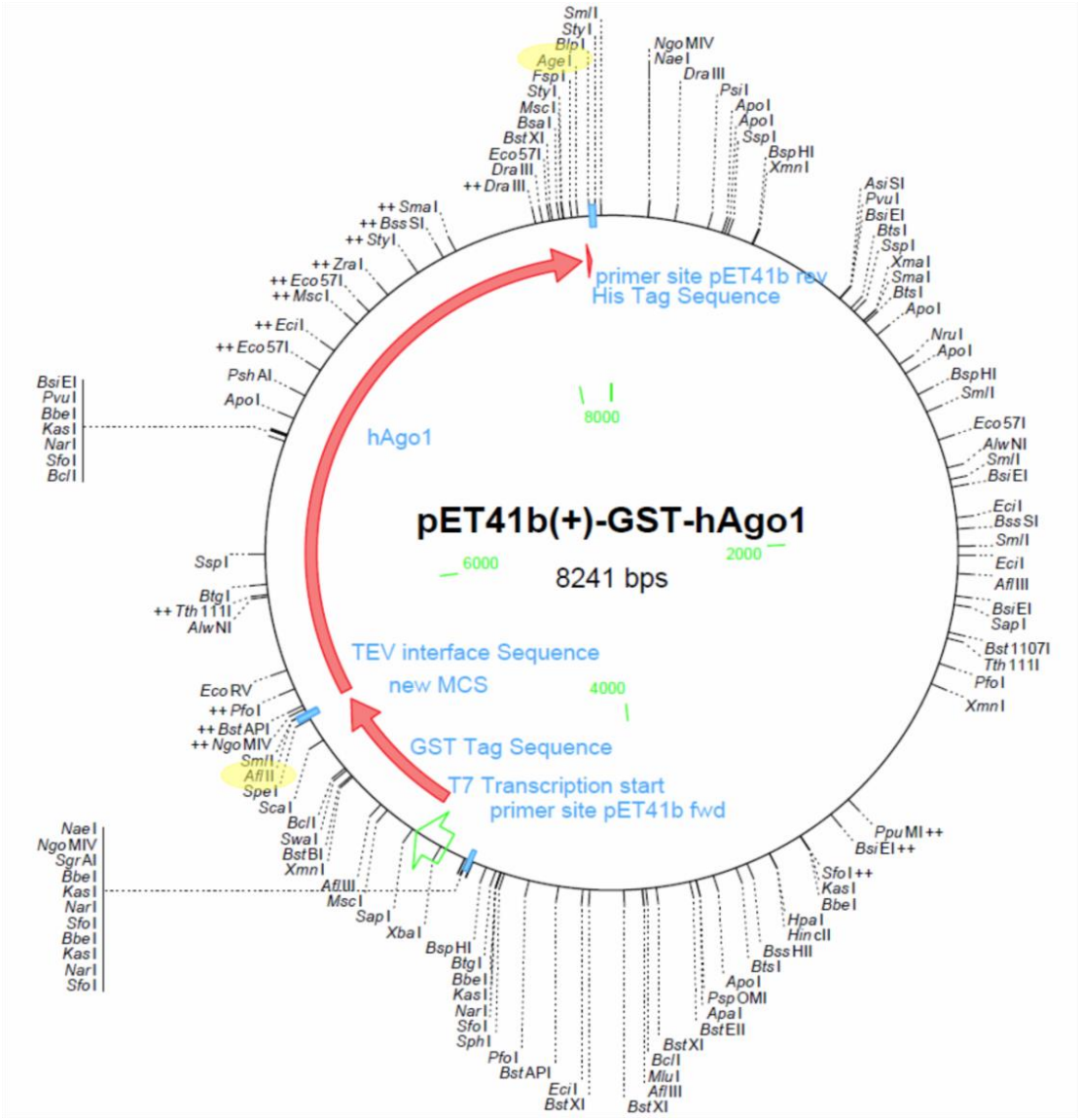


Figure 5.1 Vector map of pET41b-hAgo1. Restriction cutting sites are highlighted in yellow. The figure was prepared with Clone Manager.

The functions of individual domains between hAgo1 and hAgo2 are rarely known. To investigate functions of different domains in hAgo, six chimeras of Ago1 and Ago2 were generated according to our current understanding. Briefly, domains from hAgo1 were swapped to hAgo2 as shown in Figure 5.2. First the sequences were amplified via high-fidelity PCR from plasmids of hAgo1 or hAgo2, and then the full length chimeras were amplified by overlapping PCR (for details see section 4.3.2). In total six chimeras were designed: $Ago2^{Ago1\_PIWI}$, $Ago2^{Ago1\_N/PAZ}$, $Ago2^{Ago1\_N}$, $Ago2^{Ago1\_PAZ}$, $Ago2^{Ago1\_PAZ1}$ and $Ago2^{Ago1\_PAZ2}$.

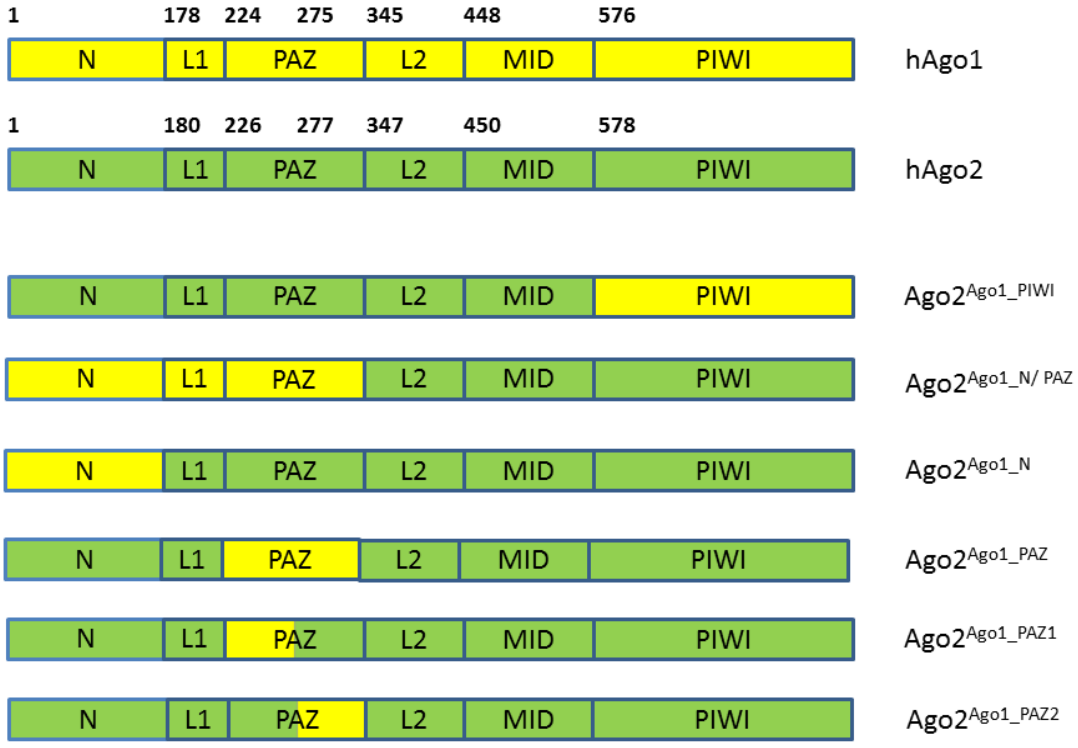


Figure 5.2 Schematic representation of the generation of human Ago1/Ago2 chimeras by domain swapping. Human Ago proteins consist of four domains, indicated as N, PAZ, MID and PIWI, along with two domain linkers, L1 and L2. Sequences of hAgo1 are colored in yellow and of hAgo2 in green. The first amino acids of each domain or linker are indicated above the diagram. The chimeras below consist of the sequence combination of hAgo1 and hAgo2. For example, $Ago2^{Ago1\_PIWI}$ consists of N to MID domain (aa 1-577) from hAgo2 and PIWI domain (aa 576-856) from hAgo1. For amino acid sequences of human Agos see Figure 5.21.

Full length DNA sequences of hAgo1 (GenBank accession no. NP_036331) and the chimeras were ligated into pET41b (+) vectors and expressed in *E. coli* BL21 (DE3), whereas successful insertion of target DNAs to the plasmids were confirmed by DNA sequencing (GATC Biotech) before.

5.1.2. Recombinant protein purification

The GST fusion proteins were purified as described in the method chapter (see 4.6.2. protein purification). The recombinant hAgo1 protein was analyzed by SDS-PAGE (Figure 5.3A) and Western blot (Figure 5.3B).

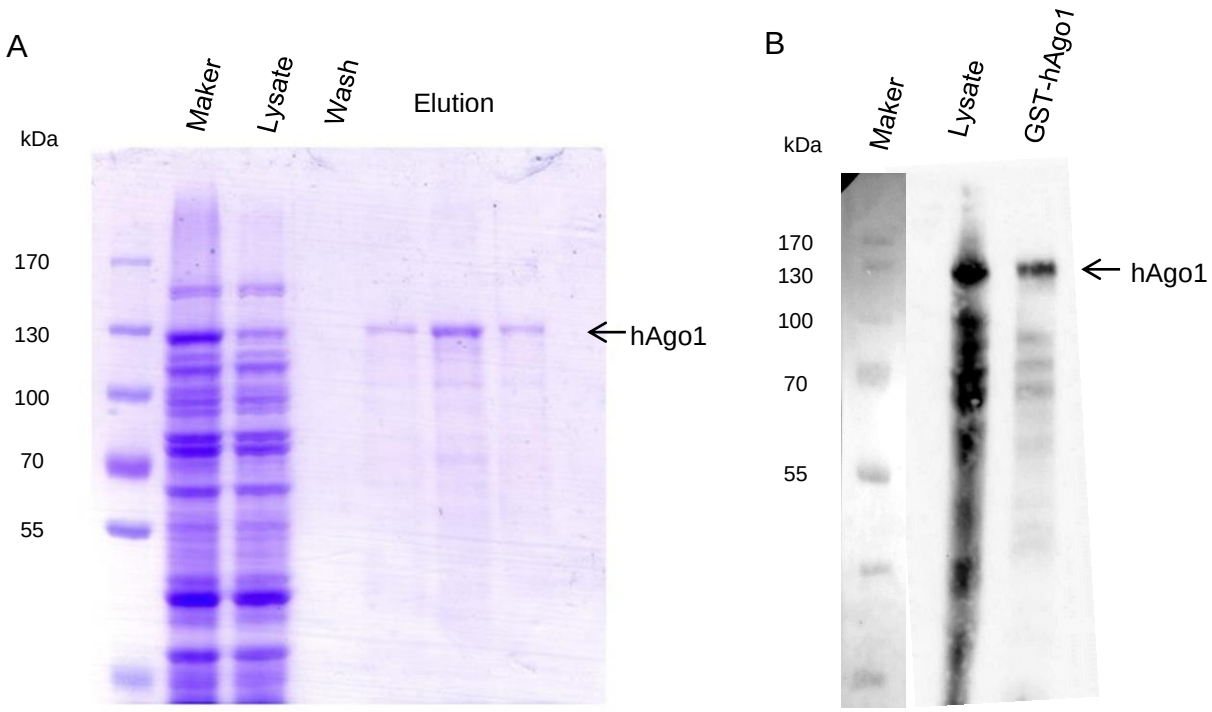


Figure 5.3 SDS-PAGE and Western blot analysis of purified hAgo1. (A) The progress of purification and isolation of hAgo1 was analyzed using 8% SDS-PAGE. Lanes from left to right: marker (PageRuler™ Prestained Protein Ladder), lysate of *E. coli* cells, flow through (FT) of column loading solution, flow through of column wash solution and eluted fractions. (B) Western blot analysis using rat anti-hAgo1 antibody (methods see section 4.5.2).

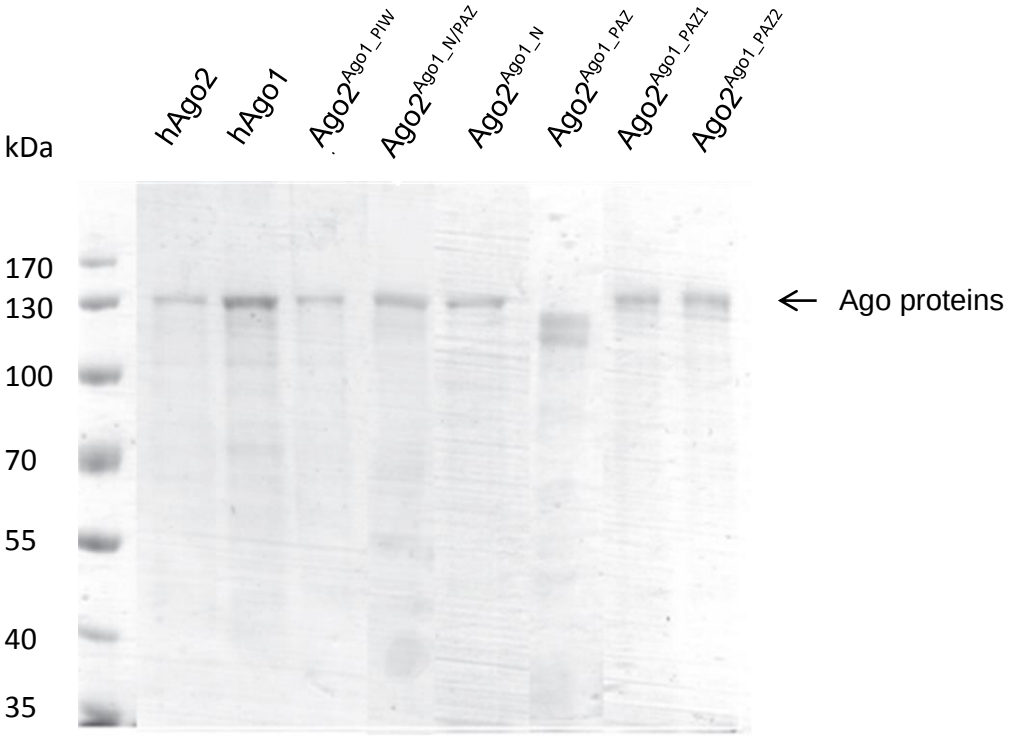


Figure 5.4 SDS-PAGE analysis of purified recombinant Agos and chimeras. All purified proteins were analyzed by 8% SDS-PAGE and showed the same size (ca. 124 kDa) except Ago2^{Ago1_PAZ}.

The validity of the expression of hAgo1/2 protein and the corresponding chimeras is strongly indicated due to the listed facts:

1. hAgo1 protein was successfully and reproducibly expressed and isolated using GST tag (Figure 5.3A and Figure 5.4).
2. According to the protein marker the obtained hAgo1 and hAgo2 proteins display the known molecular weights of approximately 124 kDa (45, 128).
3. The chimeras were designed to have the same length as wild type Ago, protocols for Ago and chimeras expression and purification were comparable. The same molecular weight was apparent for all Ago chimeras (Figure 5.4) except for Ago2^{Ago1_PAZ} (see below for detailed analyses).
4. It was possible to detect hAgo1/2 and chimeras using the specific anti-hAgo1 or anti-hAgo2 antibodies, indicating that no unspecific bacterial protein having the molecular weight of Ago protein was expressed here.
5. As described above on DNA level sequence identity of each construct was verified (see Appendix for DNA sequence). For ongoing studies sequence

analysis e.g. via mass spectrometry or Edman degradation may be used for final verifications.

5.1.3. Ago2^{Ago1_PAZ} protein analysis

All recombinant Ago proteins have a molecular weight of ca. 124 kDa, except for Ago2^{Ago1_PAZ} which showed two lower bands (Figure 5.4). Western blot analysis was applied using an anti-GST antibody to target the N-terminal GST-tag and an anti-hAgo2 antibody to target the C-terminal part of hAgo2 protein. The double bands were recognized by both antibodies (Figure 5.5) and there was a degradation product highly recognized by the α-GST antibody, suggesting those two bands were degradation products of Ago2^{Ago1_PAZ} and the degradation happened most likely in the N-terminal part of the protein.

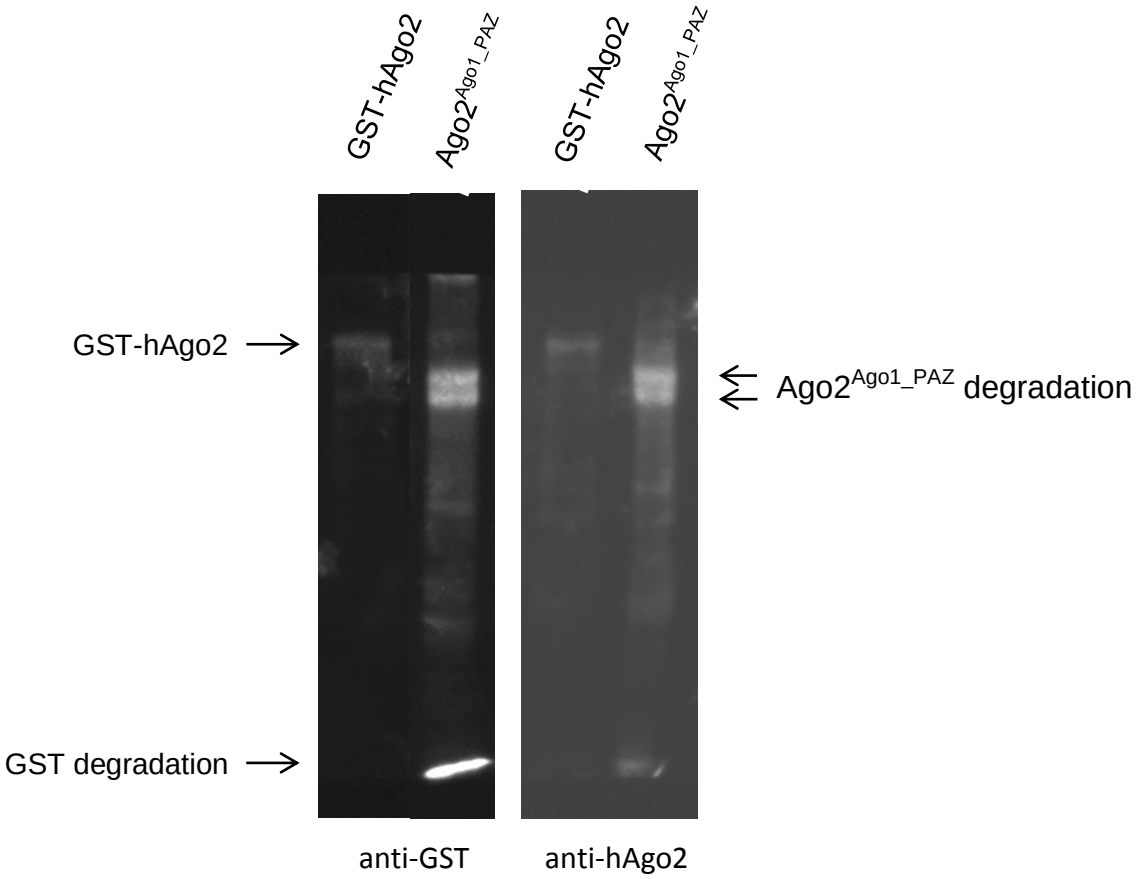


Figure 5.5 Western blot analysis of Ago2^{Ago1_PAZ}. The hAgo2 (as a positive control) and Ago2^{Ago1_PAZ} samples were incubated with either α-GST or α-hAgo2 antibody.

Additionally, *E. coli* strains, BL21-Gold (DE3) and Rosetta™ (DE3) were tested to improve the expression yields and to avoid the N-terminal degradation, but this approach did not prove to be successful. Then the PAZ domain was split in two parts swapping PAZ1 (aa 224-274) and PAZ2 (aa 275-344) of hAgo1 to hAgo2 to generate $Ago2^{Ago1\_PAZ1}$ and $Ago2^{Ago1\_PAZ2}$ (Figure 5.2). Both of them were expressed displaying a molecular weight of ca. 124 kDa (Figure 5.4).

5.2. RNA binding kinetics of hAgo1 protein

5.2.1. hAgo1 shows similar complex formation properties as hAgo2

In a previous study of Prof. Restle's group, a mechanistic model for the hAgo2-catalyzed RNA slicing reaction was established (45). Here, hAgo2/siRNA binding was studied by steady-state and pre-steady-state techniques, and transient kinetic data were related to conformational changes, leading to a mechanistic model for hAgo2 target RNA binding and cleavage.

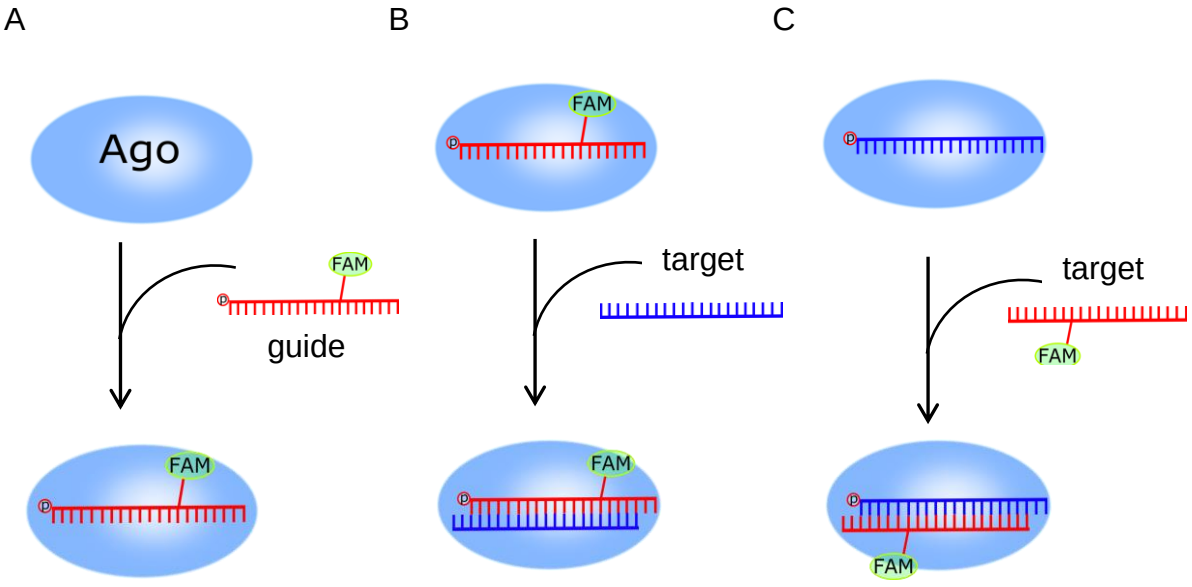


Figure 5.6 Schematic representation of the fluorescence based binding studies. (A) For the measurements of binary complex formation, $as2b^{FAM}$, a 21 nt long guide strand with a fluorescent label at the position 14 was used. The Ago-guide binary complex was also used for analysis of ternary complexes formation. The $as2b^{FAM}$ was used either as a guide (B) or as a target (C) substrate, depending if 5´-phosphorylated or not (guide carries

phosphorylation). In addition, an unlabeled complementary strand was used as target (B) or guide (C) strand. Modified with kind permission of Dr. S. Willkomm (128).

In the present study, transient kinetics of hAgo1 were measured applying the same conditions as previously used for hAgo2 (45). Either the guide or the target substrates were FAM-labeled (Figure 5.6). Firstly, complex formation between recombinant hAgo1 and RNA substrates was analyzed. The same substrates were used as for hAgo2 and the fluorescence change of the FAM labeled 21-mer RNA as2b (Table 5.1) was measured by using a fluorescence spectrometer or a stopped-flow instrument (for method details see section 4.9).

Table 5.1 Sequences of RNA oligonucleotides used in the present study

Name	Sequence (5´ – 3´)
as2b	uag agg uac gug cug agg cTT
as2b^{FAM}	uag agg uac gug c**T***g agg cTT
s2b	gcc uca gca cgu acc ucu aTT
s2b^{FAM}	gcc uca gca cgu acc ucu aT**T***

Boldface letters indicate the position of the fluorescence label: an asterisk indicates the position of C6-linked 5/6-FAM. Uppercase letters represent deoxynucleotides. Unless otherwise indicated, all strands used as guide were 5´-phosphorylated.

For analysis of binary complex formation, 5´-phosphorylated or non-phosphorylated (5´-OH) single stranded guides (as2b^{FAM}) were used. To measure hAgo1/siRNA complex formation, a corresponding 5´-phosphorylated double strand (as2b^{FAM}/s2b) was used.

K_d values of 3.5, 83 and 44 nM (5´-P, 5´-OH and ds-siRNA, respectively) for hAgo1 were determined by performing equilibrium titrations (Figure 5.7A). The values obtained are very close to the values obtained for hAgo2 (7, 106 and 48 nM) (45). For ternary complex formation, the target strand s2b was titrated with a hAgo1/as2b^{FAM} complex. Here, a K_d of 0.5 nM was detected which is higher than the value of 0.2 nM observed for hAgo2 (Figure 5.7B).

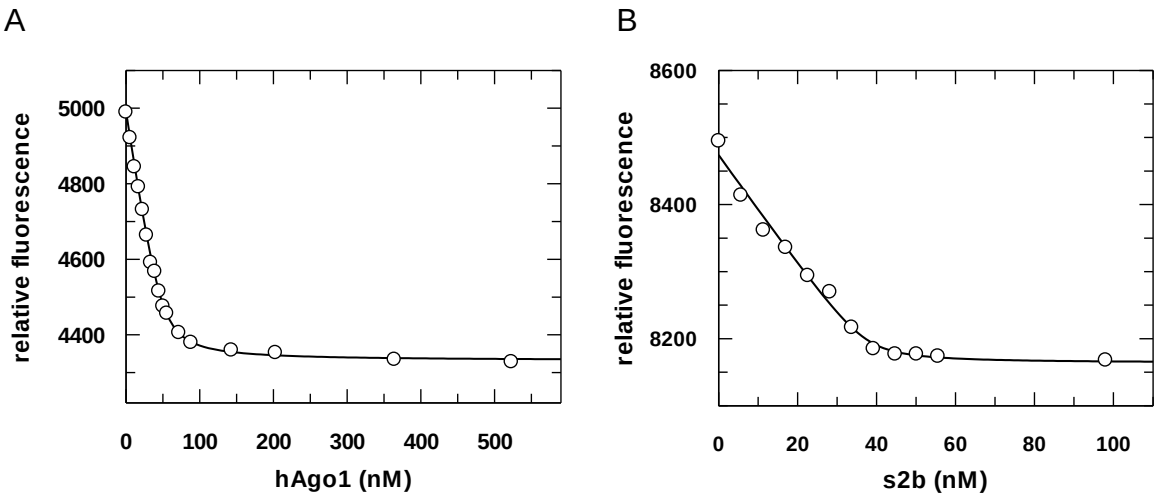


Figure 5.7 Equilibrium titrations to determine affinities of binary and ternary hAgo1/RNA complexes. Exemplary equilibrium titrations results are shown. (A) Guide strand as2b^{FAM} 20 nM was titrated with increasing concentrations of hAgo1. (B) Binary complex of hAgo1/as2b^{FAM} 20 nM was titrated with increasing concentrations of target s2b. The best fit of the experimental data to a quadratic equation is shown with the curves and the values for respective K_{ds} are 3.5 (±0.48) nM (A) and 0.54 (±0.7) nM (B).

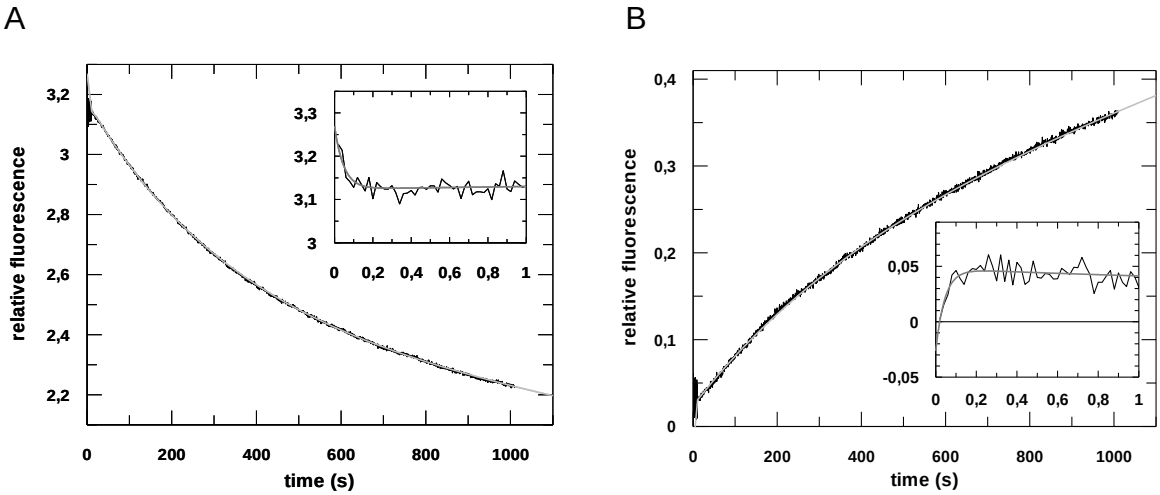


Figure 5.8 Pre-steady-state binding kinetics of association and dissociation of binary hAgo1/as2b^{FAM} complex. Exemplary stopped-flow results of binary complex formation of hAgo1 and as2b^{FAM} are shown: (A) association of hAgo1 (300 nM) and as2b^{FAM} (20 nM); (B) dissociation of hAgo1/as2b^{FAM} complex [pre-incubated hAgo1 (300 nM)/as2b^{FAM} (20 nM) against 2 µM non-labeled as2b as competitor]. To ease differentiation between corresponding rate constants of binary and ternary complex formation subscript "_bin" or "_ter" (e.g. k_{+1_bin} versus k_{+1_ter}, etc.) is added throughout the thesis. Experimental data were fitted to a triple exponential equation yielding three phases. The rate constants for association (A) are the following: $k_{+1\_obs\_bin}$: 20.8 (±1.6) s^{-1}, k_{+2_bin}:

0.17 (±0.009) s^{-1} and k_{+3_bin}: 0.003 (±0.00003) s^{-1}. And those for dissociation (B) are: k_{-1_bin}: 2.6 (±0.4) s^{-1}, k_{-2_bin}: 0.35 (±0.02) s^{-1} and k_{-3_bin}: 0.002 (±0.0001) s^{-1}.

Pre-steady-state binding kinetics were analyzed using a stopped-flow apparatus (see 4.9.2). In a previous study of Prof. Restle's group, three kinetically distinct binding steps for both the binary and the ternary complex formation were defined for hAgo2 (128, 130). As for hAgo2, data from hAgo1 pre-steady-state binding kinetics were fitted best to an equation with three exponential terms (Figure 5.8). These three phases reveal three binding steps representing defined structural transitions. In the case of binary complex formation, pre-steady-state data includes a concentration dependent fast first phase and a slower second and third phase corresponding to guide 5´-end anchoring in the MID domain and 3´-end binding to the PAZ domain, respectively.

The fact that comparable kinetic values were obtained is a strong indication that both hAgo1 and hAgo2 share the same binding mechanism. Similar to the K_d values, the pre-steady-state kinetic values for binary complex formation did not show any difference. Besides similar values for k_{+2_bin} and k_{+3_bin}, the fast first phase is also concentration dependent (Figure 5.9).

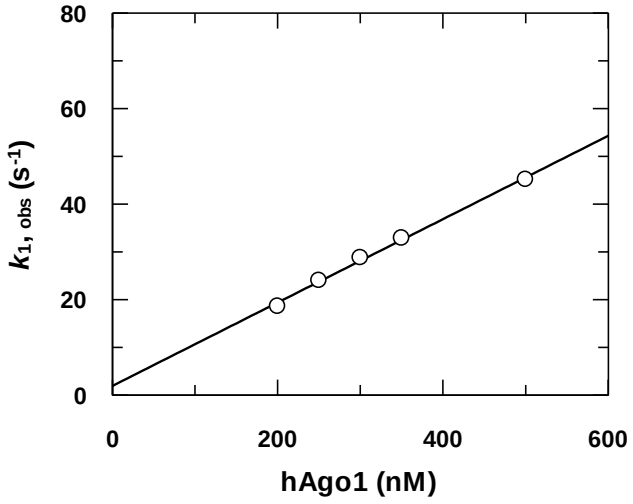


Figure 5.9 Concentration dependence of the first phase of guide RNA binding to hAgo1. A constant concentration of fluorescently labeled guide RNA (as2b^{FAM}, 20 nM) was mixed with increasing amounts of hAgo1. k_{+1_bin} and k_{-1_bin} were determined using the

slope of the linear fit and the intercept of the line with the ordinate and yielded values of 0.87 x 10^8 $M^{-1}s^{-1}$ and 1.9 (±1) s^{-1}, respectively.

Table 5.2 Summary of equilibrium and pre-steady-state binding data for hAgo1

	K_{d_meas}, nM	k_{+1} $M^{-1}s^{-1}$	k_{-1} s^{-1}	k_{+2} s^{-1}	k_{-2} s^{-1}	k_{+3} s^{-1}	k_{-3} s^{-1}	$K_{d\text{-}cal}$, nM
Binary complex								
as2b^{FAM}	3.5	0.9x10^8	2	0.13	0.35	0.004	0.002	29
OH-as2b^{FAM}	83	0.83 x10^8	1.5	0.14	1.79	0.004	0.013	742
as2b^{FAM}/s2b	44	0.62x10^8	6.7	0.17	0.40	0.003	0.002	170
Ternary complex								
as2b^{FAM}-s2b	0.5	2.5x10^8	0.42	0.022	0.44	0.004	0.001	0.82

Numbers to the left represent equilibrium measurements (K_{d_meas}). Numbers to the right were calculated from the corresponding association and dissociation rate constants (K_{d_cal}).

Further, binary complex association and dissociation between hAgo1 and 5´-OH guide RNA (OH-as2b^{FAM}) or siRNA (as2b^{FAM}/s2b) was analysed. Only k_{-2_bin} regarding hAgo1 and siRNA is much slower than for hAgo2 (Table 5.2), overall the kinetics are similar to hAgo2 (45). Ternary complex formation of hAgo1/as2b^{FAM} and s2b was examined as well. The kinetics results are listed in Table 5.2. Compared to hAgo2, dissociation of hAgo1 ternary complexes is much slower, causing a lower affinity of target to hAgo1-guide. The differences in K_d values for binary complexes determined via equilibrium and pre-steady-state measurements most probably arise from a slight protein aggregation, leading to an underestimating of k_{+1_bin} and thus higher K_d values—an incident irrelevant under equilibrium conditions. In case of ternary complex formation, this effect is less pronounced as here most likely aggregated molecules do not contribute to complex formation (compare k_{+1_bin} with k_{+1_ter}).

5.2.2. Different miRNA binding kinetics between hAgo1 and hAgo2

To test whether miRNA has a different binding affinity, the mismatched target s2b-m4 was designed according to the hsa-let-7a target HMGA2 (131, 132). Additionally, considering that guide RNAs can be divided into distinct domains

affecting binding properties and functions (115), miRNA targets s2b-m1, s2b-m2 and s2b-m3 were generated (Table 5.3). Moreover, target s2b with either matches in the 3´-end (seed match, s2b-sm) or matches in the 5´-end (non-seed match, s2b-nsm) were used for further test of more mismatches in both ends. They are all based on s2b but contain mismatches at different positions.

Table 5.3 Sequences of miRNA oligonucleotides

Name	Sequence (5´– 3´)
s2b-m1	gcc uca gac agu acc ucu aTT
s2b-m2	gcc gag gca cgu acc ucu aTT
s2b-m3	cug uca gca cgu acc ucu aTT
s2b-m4	gac gca gac agu acc ucu aTT
s2b-sm	cug cgg ugc gau acc ucu aTT

Uppercase letters represent deoxynucleotides. s2b-sm represent ‘seed match’ s2b.

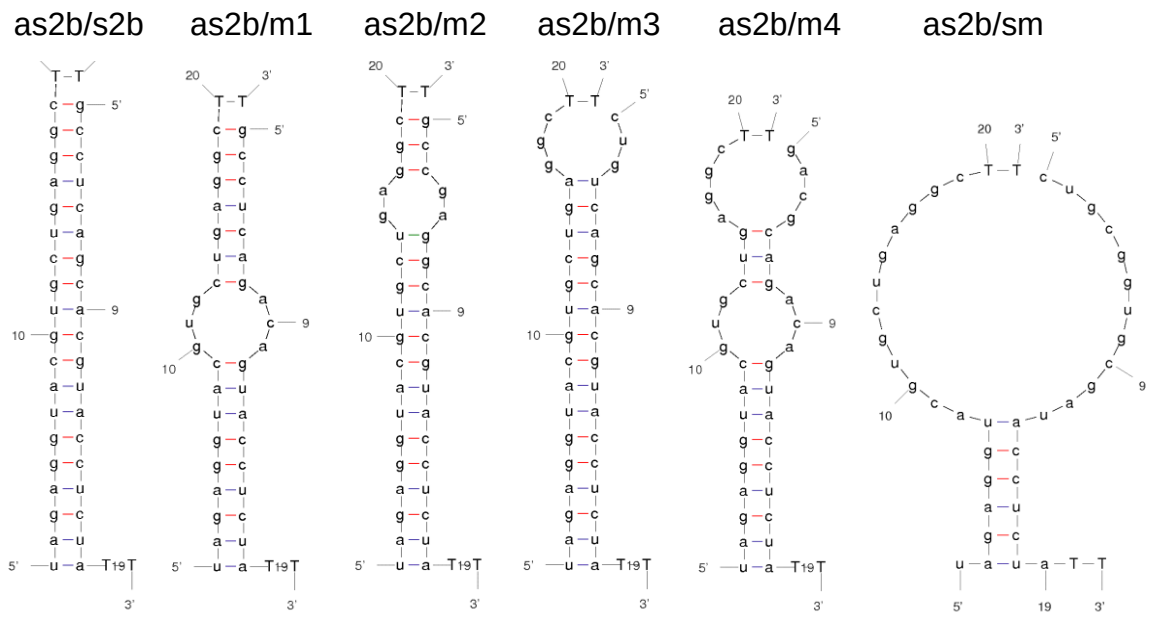


Figure 5.10 Secondary structure predictions of ds miRNA targets. All displayed miRNA double strands are based on the fully complementary siRNA composed of s2b and as2b. These ds miRNAs include mismatches at different positions. Hybridization prediction of double strand duplexes was performed using mfold Web Server. The nucleic acid folding temperature was set to 37°C, other parameters were default.

The mismatched targets listed in the table above were used to generate ds miRNA containing bulges (Figure 5.10). Further these mismatched ds miRNA were used in binary complex formation with hAgo to test the binding properties of Ago and miRNA.

These mismatched targets were analyzed for binding properties applying equilibrium titration and pre-steady-state studies. Interestingly, in case of binary complex formation, both hAgo1 and hAgo2 were bond tighter to a miRNA duplex than to the siRNA duplex. The K_d of hAgo2-miRNA was 35 nM, revealing a slightly tighter binding than 48 nM for hAgo2-siRNA. However, the K_d of hAgo1, it was 7.1 nM for the miRNA versus 44 nM for the siRNA, which suggests that hAgo1 might bind more effective to miRNA. More distinguishing results were found comparing ternary complex formation of hAgo1 and hAgo2 (Figure 5.11).

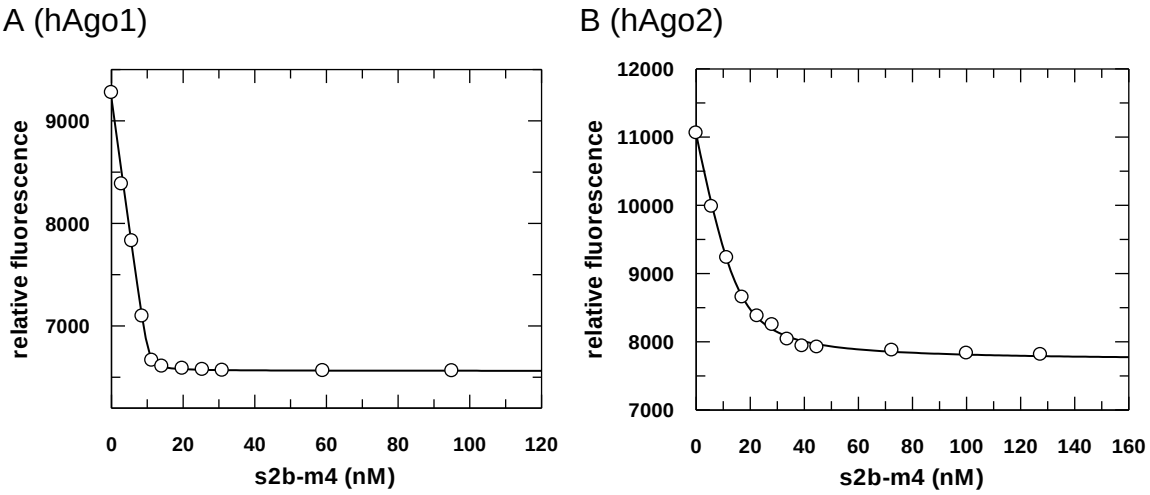


Figure 5.11 Analysis of hAgo1 and hAgo2 miRNA-like ternary complex formation using equilibrium titrations. A binary complex of either hAgo1 (A) or hAgo2 (B) and FAM-labeled guide RNA (as2b^{FAM}) (500 nM hAgo and 20 nM guide RNA) was titrated with increasing concentrations of mismatched target RNA (s2b-m4). The best fit of the experimental data to a quadratic equation is shown by the curves and gives values for respective K_{ds} for hAgo1 and hAgo2 complexes of 0.06 (±0.05) and 2.2 (±0.3) nM, respectively.

Mismatched targets showed a ca. 5-fold tighter binding than fully paired target in hAgo1-complex, with both the second and third phases of association and dissociation being faster or slower, respectively (Figure 5.12 and 5.13).

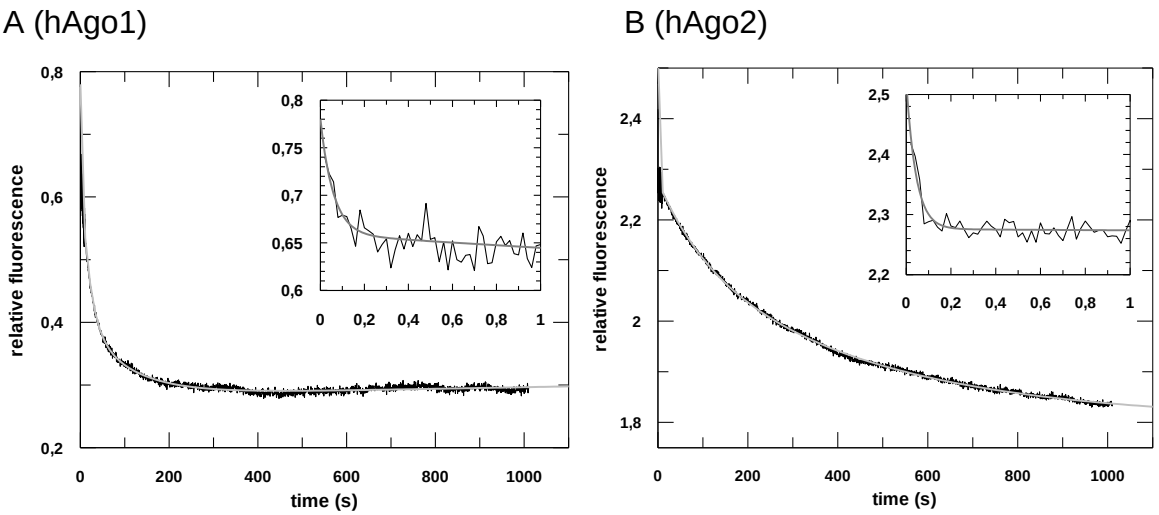


Figure 5.12 Pre-steady-state kinetics of ternary complex formation of hAgo1 and hAgo2 with mismatched target RNA. Exemplary stopped-flow results are shown. Insets show the reactions on a shorter time scale. Here the binding of a target RNA (s2b-m4; 20 nM) to a binary hAgo1 (A) or hAgo2 (B) -guide RNA complex [500 nM Ago and 20 nM guide RNA (as2b^{FAM})] is shown. Experimental data were fitted to a triple exponential equation yielding three phases. For hAgo1 (A): $k_{+1\_obs\_ter}$: 17.4 (±1.1) s^{-1}, k_{+2_ter}: 0.16 (±0.009) s^{-1} and k_{+3_ter}: 0.01 (±0.00008) s^{-1}. For hAgo2 (B): $k_{+1\_obs\_ter}$: 24.6 (±1) s^{-1}, k_{+2_ter}: 0.03 (±0.002) s^{-1} and k_{+3_ter}: 0.005 (±0.0001) s^{-1}.

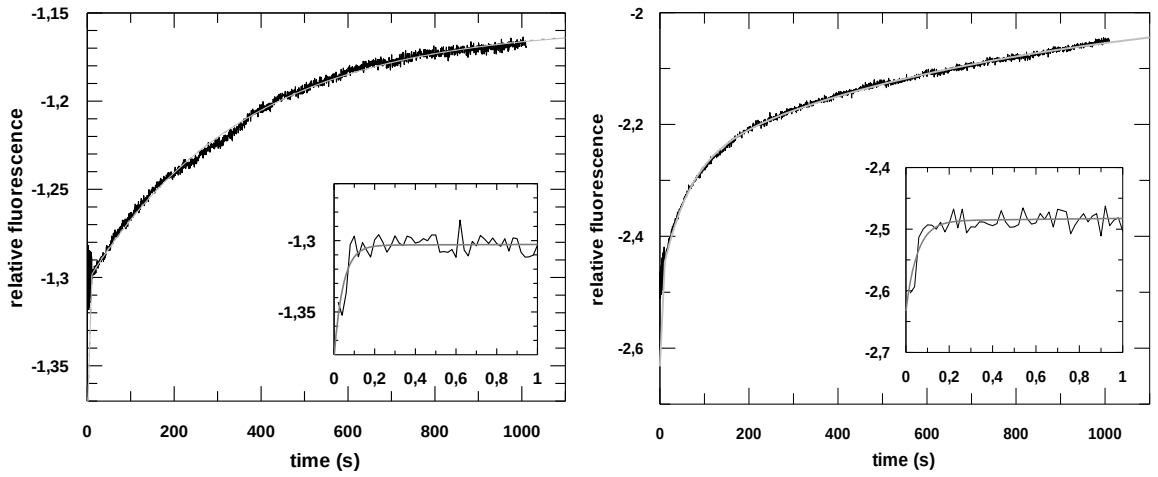


Figure 5.13 Pre-steady-state kinetic analysis of the dissociation of ternary complexes. Guide RNA as2b^{FAM} and mismatched target s2b-m1 was pre-incubated with hAgo1 (A) or hAgo2 (B) and then rapidly mixed with a 33-fold excess of non-labeled competitor guide RNA as2b. The change in fluorescence signal was detected over time to determine dissociation rate constants. The data were fitted to a triple exponential equation

yielding three phases. For hAgo1 (A): k_{-1_ter}: 2.4 (±0.4) s^{-1}, k_{-2_ter}: 0.02 (±0.003) s^{-1} and k_{-3_ter}: 0.0013 (±0.00002) s^{-1}. For hAgo2 (B): k_{-1_ter}: 4 (±0.7) s^{-1}, k_{-2_ter}: 0.02 (±0.002) s^{-1} and k_{-3_ter}: 0.003 (±0.0002) s^{-1}.

Table 5.4 Summary of equilibrium and pre-steady-state data of hAgo binary and ternary complex formation with miRNA. A comparison of miRNA binding properties between hAgo1 and hAgo2 is shown. For binary complex formation, hAgo was mixed with pre-annealed double strand RNA as2b^{FAM}/m4 (represents miRNA duplex). For ternary complex formation, pre-incubated hAgo/as2b^{FAM} was mixed with mismatched targets (m1–m4 and sm).

		K_{d_meas}, nM	k_{+1}, $M^{-1}\cdot s^{-1}$	k_{-1}, s^{-1}	k_{+2}, s^{-1}	k_{-2}, s^{-1}	k_{+3}, s^{-1}	k_{-3}, s^{-1}	K_{d_cal}, nM
Binary complex miRNA (Ago + as2b^{FAM}/m4)									
hAgo1		9.1	$0.48x10^8$	10	0.15	0.35	0.01	0.001	49
hAgo2		35	$0.6x10^8$	3.1	0.26	0.78	0.005	0.004	124
Ternary complex (Ago/as2b^{FAM} + target)									
hAgo1	m1	0.06	$3.4\ x10^8$	2.4	0.053	0.02	0.01	0.0013	0.35
	m2	0.07	$4.2\ x10^8$	0.5	0.1	0.044	0.005	0.0024	0.26
	m3	0.12	$4.5\ x10^8$	0.8	0.032	0.024	0.003	0.0014	0.62
	m4	0.05	$4.8\ x10^8$	1.5	0.16	0.022	0.008	0.006	0.32
	sm	0.1	$3.9\ x10^8$	2	0.14	0.047	0.009	0.002	0.38
hAgo2	m1	1.5	$4.0x10^8$	4	0.03	0.02	0.007	0.003	3
	m2	16.7	$4.2x10^8$	1.1	0.015	0.06	0.002	0.003	15
	m3	1.1	$6.2\ x10^8$	0.9	0.015	0.06	0.005	0.0016	1.8
	m4	2.2	$5.9\ x10^8$	1.3	0.034	0.033	0.005	0.003	1.3
	sm	0.65	$5.7x10^8$	1.1	0.03	0.006	-	-	0.4

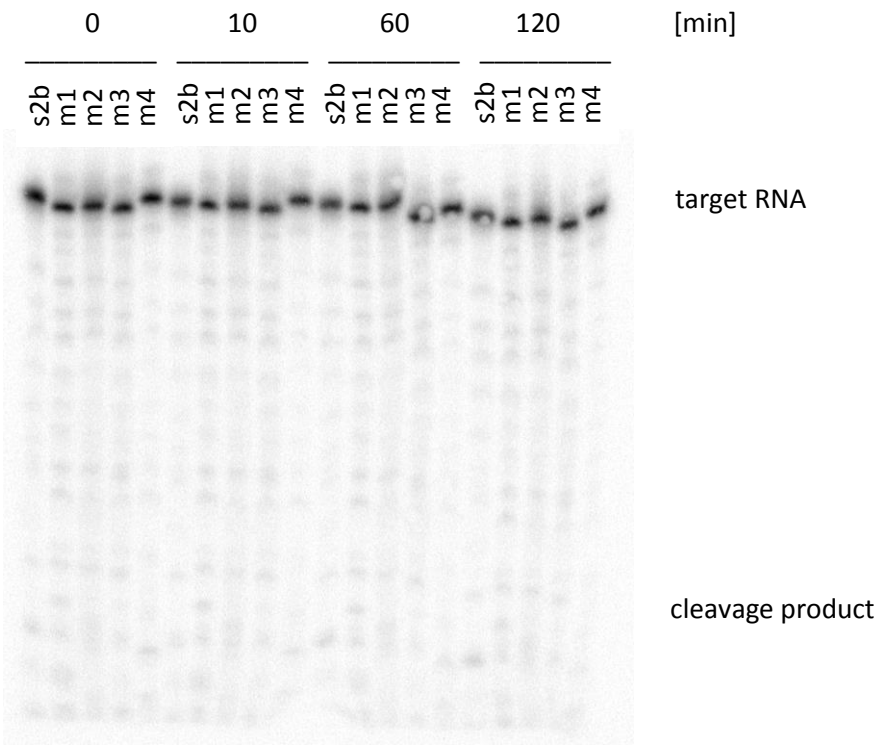
For hAgo2, differences between mi- and siRNA binding could only be observed for ternary complex formation. hAgo2 shows a much lower affinity for targets with mismatches, with the major difference in dissociation. A fully complementary target

fulfills a very slow dissociation reaction with 0.002 s^{-1} for k_{-2} and 0.0002 s^{-1} for k_{-3} (45), which is supposed to promote cleavage of fully complementary target RNAs.

Table 5.2 shows the equilibrium and pre-steady-state binding data for hAgo1 and fully paired target. Here, the binary and ternary complex formation of Ago and mismatched targets were investigated (Table 5.4), instead of the perfect paired target s2b. Targets s2b-m1 to m4 and s2b-sm containing different mismatch positions were used to represent various miRNA targets. The ‘non-seed match’ target s2b-nsm was also tested, but it did not bind tightly to either hAgo1 or hAgo2.

The results indicate that the miRNA targets are loaded faster and tighter to hAgo1, but both miRNA and siRNA can be loaded to both hAgo1 and hAgo2. The differences observed are more significant for ternary than for binary complex formation. In other words, hAgo1 loads miRNA targets faster to ternary complexes than to binary complexes. Next, potential cleavage activities of hAgo1 and hAgo2 were compared, using fully complementary and mismatched targets (Figure 5.14 and 5.15).

A



B

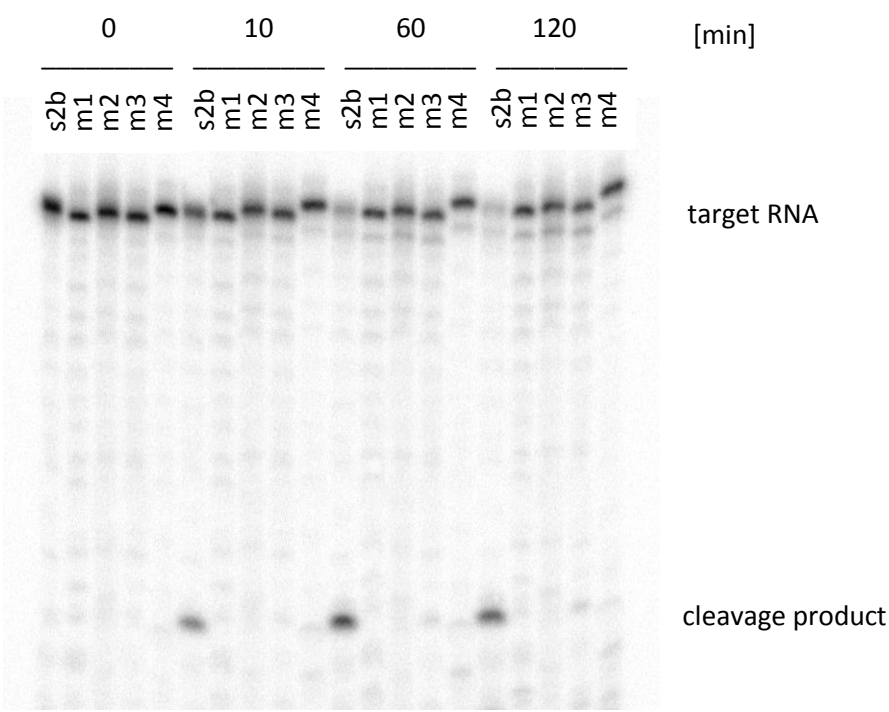
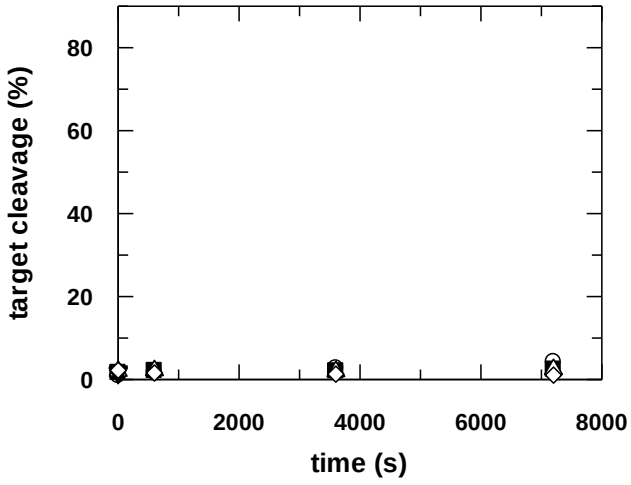


Figure 5.14 Autoradiography of hAgo-mediated cleavage reactions. First, 3.7 µM hAgo1 (A) or hAgo2 (B) and 100 nM ss guide RNA (as2b) were preincubated in cleavage buffer for 10 min. Reaction was started by adding 2.5 nM of one 5´-^{32}P-labeled target RNAs (s2b, s2b-m1, s2b-m2, s2b-m3 or s2b-m4). Reactions were analyzed using 20% (w/v) denaturing PAGE and visualized via autoradiography. Samples were collected at four different time points 0, 10, 60 and 120 min.

A (hAgo1)



B (hAgo2)

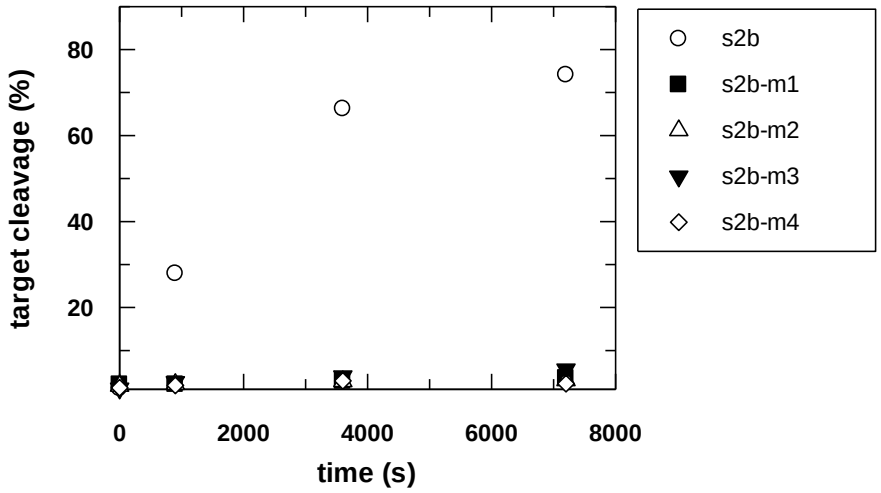


Figure 5.15 cleavage assays with hAgo1 and hAgo2 proteins and different si- or mi-RNA substrates. Cleavage assays were conducted using hAgo1 (A) or hAgo2 (B), details described in Figure 5.14. The experimental data of s2b cleaved by hAgo2 was fitted to a single exponential equation, yielding a rate constant of 0.0005 (±0.00004) s^{-1}.

5.2.3. The mismatch position affects binding of miRNA to Ago

Next, the effect of several mismatch positions within a miRNA target on hAgo ternary complex formation was investigated. All miRNA duplexes that contain mismatches showed increased binding affinity to hAgo1, whereas mismatches in central or supplementary region were more effective than mismatches in the tail region (the position of mismatches see Figure 5.10). While mismatches enhance miRNA binding to hAgo1, the opposite is true for hAgo2. All mismatches tested, caused a slight reduction of K_d, whereas mismatches in the supplementary region caused a reduction up to a 10-fold lower K_d.

It seems likely that hAgo proteins show a preference for binding of different miRNA. This may be caused by the position of mismatches within the duplex which influences the binding properties differently. The data indicate a preference of hAgo1 to bind mismatched targets, whereas a seed match might be sufficient for proper binding since the second and third phase of hAgo1-as2b/s2b-sm ternary complex formation (k_{+2_ter} and k_{+3_ter} in Figure 5.16) are similar to a fully paired target. This indicates hAgo1 binds a miRNA target in a base-pair hybridization

independent manner as long as the seed region is paired to the target RNA. For hAgo2, the situation is different. s2b-sm is loaded to hAgo2 with only two phases (Table 5.2), the third phase is missing which is proposed to represent the 3´-end of guide RNA release from the PAZ domain.

A (hAgo1) B (hAgo2)

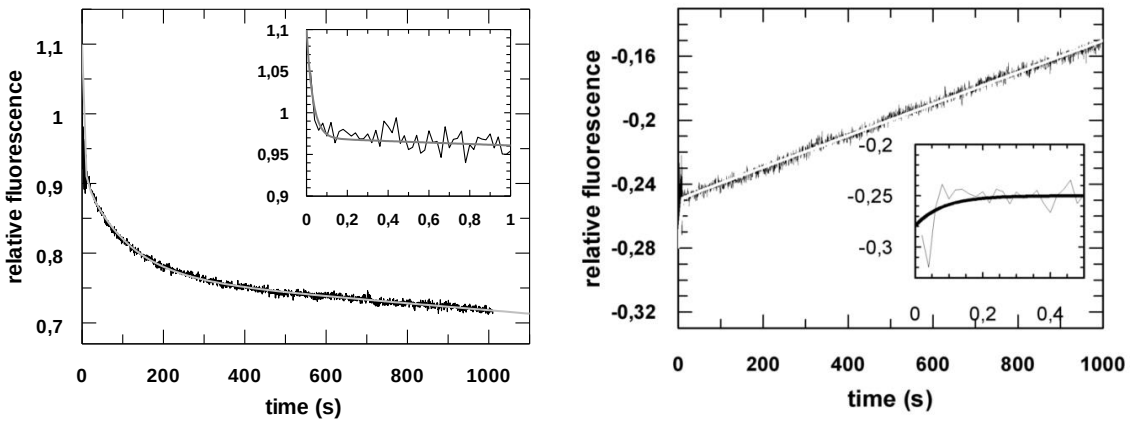


Figure 5.16 Pre-steady-state kinetics of ternary complex formation using a seed matched target. Results of stopped-flow measurements of the assembly of hAgo/as2b^{FAM} and s2b-sm ternary complex are shown. s2b-sm has a mismatched 3´-region. In case of hAgo1 (A), the graph shows three phases with the following rate constants: $k_{+1\_obs\_ter}$: 31.9 (±4.1) s^{-1}, k_{+2_ter}: 0.14 (±0.006) s^{-1} and k_{+3_ter}: 0.009 (±0.0002) s^{-1}, which is very close to what was observed for fully paired target. For hAgo2 (B), the experimental data were fitted best to a double exponential equation. Here, the third phase is missing, which indicates the seed matched target does not trigger the release of guide 3´-end from PAZ domain (45).

5.3. PAZ domain is important for miRNA binding

5.3.1. Domain swapping

Compared to hAgo2, hAgo1 shows higher miRNA binding affinity. To investigate which domain of hAgo1 is responsible for the differences observed in mismatched target binding, domain-swapping experiments were conducted substituting each functional domain of hAgo2 with the corresponding domain of hAgo1. The MID domain which binds to the 5´-end of guide RNA was not changed, because both Ago1 and Ago2 most likely recognize the seed region in a comparable manner using the MID domain. The chimera Ago2^{Ago1_PIWI} which is hAgo2 containing the

PIWI domain of hAgo1, did not show tighter mismatched target binding, but a loss in cleavage activity (Figure 5.15 and 5.16). To test the influence of the N terminal domain of hAgo1, N and PAZ domains between hAgo1 and hAgo2 were exchanged. By swapping the N and PAZ domains of hAgo1 to hAgo2, it was possible to convert hAgo2 into a hAgo1-like protein with a K_d for miRNA binding similar to Wt hAgo1. To further exemplify this point, the N domain or PAZ domain from hAgo1 to hAgo2 was swapped separately, but Ago2^{Ago1_PAZ} did not fold properly causing N-terminal degradation products (Figure 5.4). As described above (section 5.1.3), the PAZ domains were rearranged (split in two parts, PAZ1 and PAZ2) resulting in expressed chimeras with full length 124 kDa. Then all chimeras were analyzed for cleavage functions.

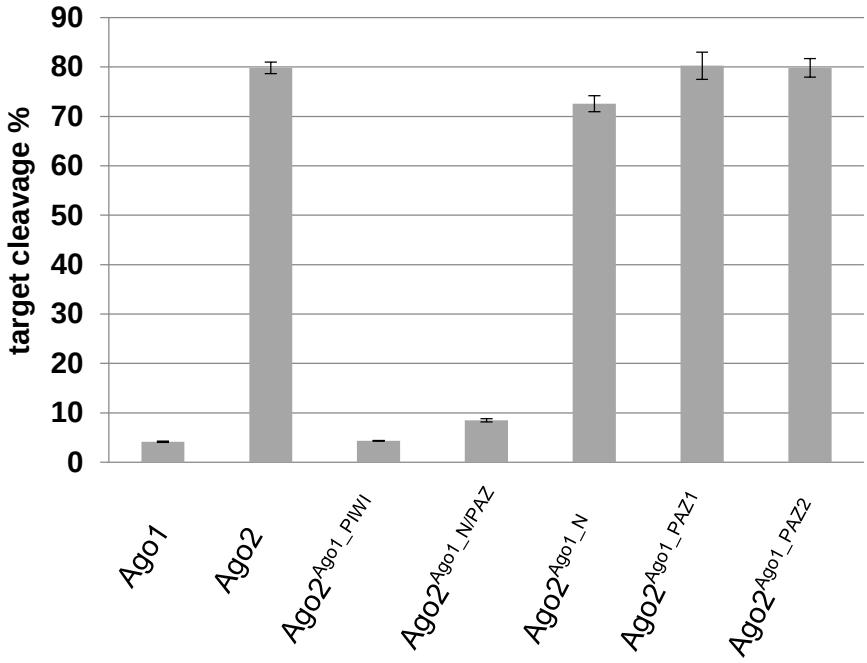
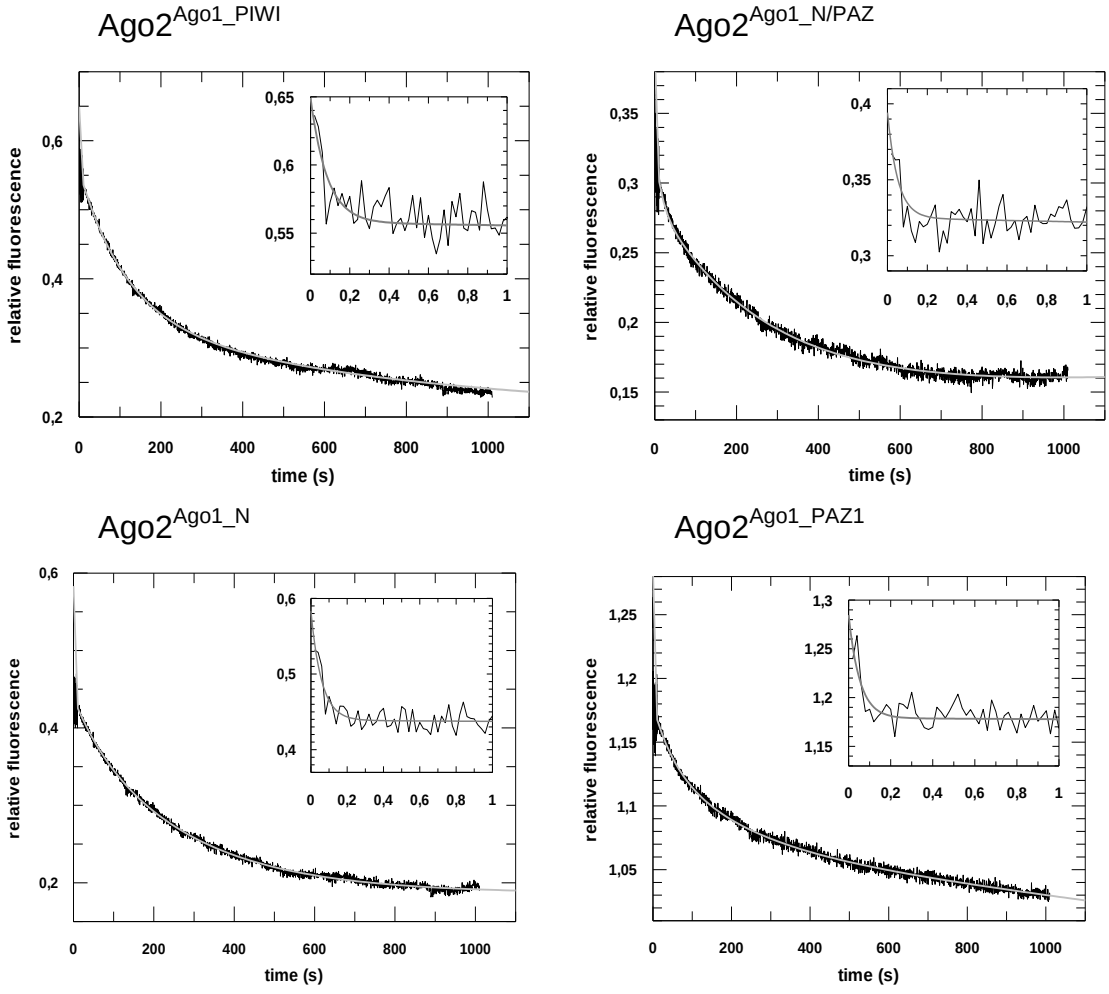


Figure 5.17 Effect of hAgo1 domains on target-RNA cleavage. Cleavage assays for WT hAgo1, hAgo2 and chimeras of hAgo2 are shown each containing single or double domains of hAgo1. For the assay 3.7 µM Ago protein, 100 nM guide RNA (as2b) and 2.5 nM target RNA s2b was used. Reactions were analyzed with 20% PAGE.

Slicer activities of wild type and chimeras were tested by a cleavage assay (Figure 5.17). First of all, none of the targets with mismatches was cleaved by any chimera. The standard target s2b was cleaved by hAgo2, Ago2^{Ago1_N}, Ago2^{Ago1_PAZ1} and Ago2^{Ago1_PAZ2}, but not by hAgo1, Ago2^{Ago1_PIWI} or Ago2$^{Ago1_N/PAZ}$. In case of

Ago2^{Ago1_PIWI}, the PIWI domain of hAgo1 lacks amino acids in the “catalytic site” which are critical for slicing activity. The inactivity of Ago2$^{Ago1_N/PAZ}$ might be due to a difference in the nucleic acid binding channel which is formed by the N and the PAZ domain.

The miRNA binding properties of Ago chimeras were analyzed by equilibrium and pre-steady-state studies. The major differences between chimeras were observed for Ago miRNA ternary complex formation (Figure 5.18)**,** while overall the kinetics between Wt hAgo (Table 5.4) and chimeras (Table 5.5) were comparable. Surprisingly, Ago2^{Ago1_PAZ1} showed kinetic values very close to Wt hAgo1. On the other hand, Ago2^{Ago1_PAZ2} more closely resembled Wt hAgo2. Considering the minor difference of binding affinity compared to Ago2^{Ago1_N}, PAZ1 is the essential region for binding of mismatched target.



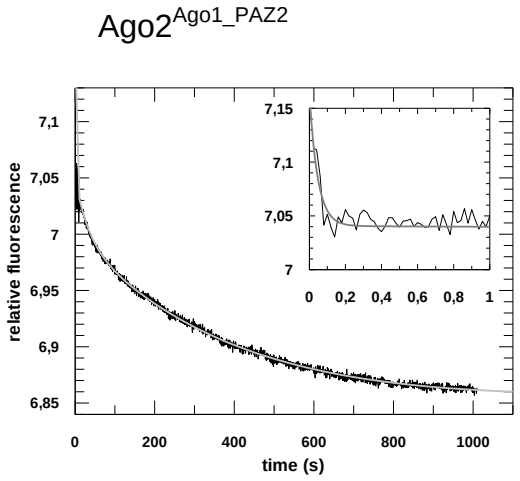


Figure 5.18 Analysis of pre-steady-state kinetics of chimeras during ternary complex formation (as2b^{FAM}/m4). Stopped flow binding data of a mismatched target RNA (s2b-m4, 20 nM) to an Ago/as2b complex [500 nM Ago chimeras and 20 nM guide RNA (as2b^{FAM})]. Experimental data were fitted to a triple exponential equation yielding three phases with the rate constants given in the Table 5.5.

Table 5.5 Summary of equilibrium and pre-steady-state data of hAgo2 chimeras. For simplicity**,** merely s2b-m4 was used as an exemplary mismatched target; for pre-steady-state diagrams see Figure 5.18. Data concerning the dissociation process was not determined since no significant difference was expected according to the data obtained from the association experiments.

	K_d (nM)	k_{+1} (M^{-1} s^{-1})	k_{+2} (s^{-1})	k_{+3} (s^{-1})
Binary complex miRNA (Ago + as2b^{FAM}/m4)				
Ago2^{Ago1_PIWI}	33.5	1.36 x10^8	0.1	0.005
Ago2$^{Ago1_N/PAZ}$	10.5	1.25 x10^8	0.11	0.006
Ago2^{Ago1_N}	21	0.8 x10^8	0.07	0.003
Ago2^{Ago1_PAZ1}	15	1.3 x10^8	0.134	0.07
Ago2^{Ago1_PAZ2}	31	0.6 x10^8	0.06	0.004
Ternary complex (Ago/as2b^{FAM} + m4)				
Ago2^{Ago1_PIWI}	1.9	4.8 x10^8	0.01	0.004
Ago2$^{Ago1_N/PAZ}$	0.1	4 x10^8	0.06	0.014
Ago2^{Ago1_N}	0.85	3.7 x10^8	0.025	0.004
Ago2^{Ago1_PAZ1}	0.17	4.3x10^8	0.04	0.01
Ago2^{Ago1_PAZ2}	1	3.5x10^8	0.016	0.003

5.3.2. hAgo1 is evolutionarily more conserved than hAgo2

hAgo1 and hAgo2 proteins display a primary sequence identical of 84%, but seem to have evolved independently because different Ago proteins exist in all organisms. In this thesis amino acid sequence alignments of Ago1 and 2 among different species were performed. It is remarkable that both Ago1 and 2 are highly conserved in mammals. For example, they share 99 % identity from human to mouse or rat. However, Ago1 has much less amino acid changes than Ago2 (Table 5.6). Moreover, there is not only a higher degree of identity among mammals or vertebrates, hAgo1 shows even 73% identity with flies (Table 5.6). In contrast, Ago2 shows a lower degree of identity in all respective comparisons; the identity with flies is only 33 %. So it appears that hAgo1 is more conserved than Ago2 during evolution, and one can speculate that hAgo1 preserved more functions in lower organisms. On the other hand, hAgo2 might have obtained some new functions in more advanced organisms.

Table 5.6 Sequence alignments of Ago1 and 2 among different species. The degrees of identity of Ago1 and Ago2 between humans and other species are shown in percentage. Identified amino acids numbers are given in brackets. All sequences are from the NCBI Protein database. The alignments were done with Align Sequences Protein BLAST.

Homo sapiens	*Mus musculus*	*Rattus norvegicus*	*Danio rerio*	*Drosophila melanogaster*	*Arabidopsis thaliana*
Ago1	100% (857/857)	99% (856/857)	96%	73%	43%
Ago2	99% (853/860)	99% (852/860)	92%	33%	32%

5.3.3. PAZ domain of hAgo1 is more flexible than in hAgo2

Ago is a well-known flexible protein, and nucleic acid binding triggers rearrangement of the domains (42). Thus, in this thesis the structural flexibility between hAgo2 binary (4W5N) and ternary complexes (4W5O) were compared using B-factors analysis, which describes the displacement of atoms from their

mean position in a crystal structure and thus diminished scattered X-ray intensity. Generally, according to the B-factors, the ternary complex is more stable than the binary complex. Especially the C-terminal part of hAgo2 is stabilized by binding to a short target. Notably, hAgo2 shows a more flexible MID domain in the binary complex, but a more flexible PAZ domain in the ternary complex (Figure 5.19). This indicates that seed region pairing of the target RNA stabilizes the MID domain and activates the PAZ domain. It suggests a stepwise propagation for target recognition from MID to PAZ although the details have not been understood yet.

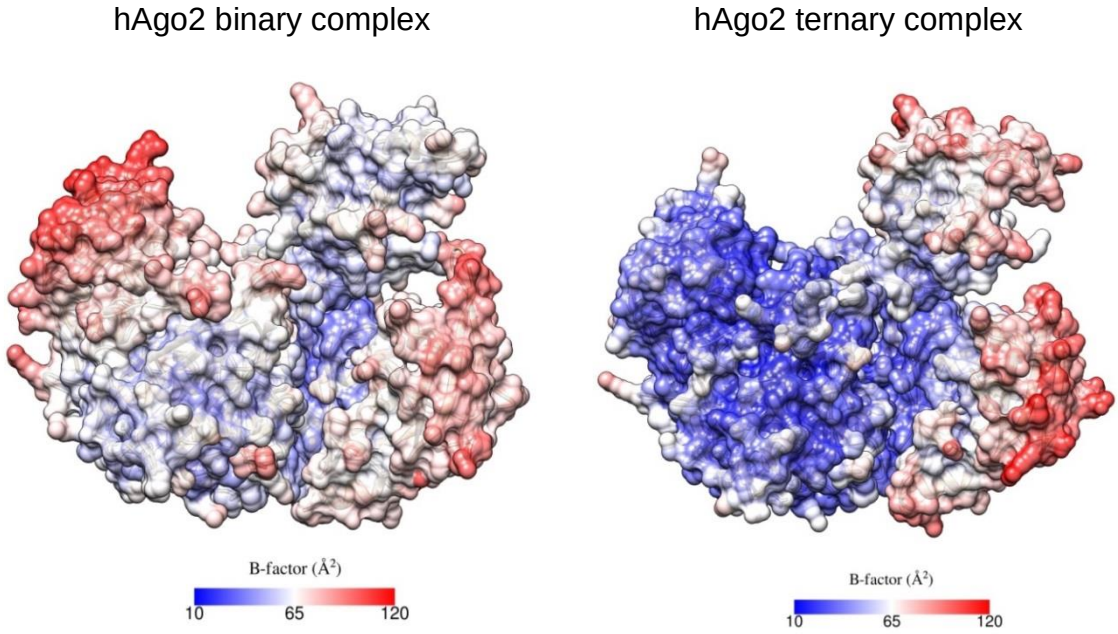


Figure 5.19 Normalized B factor analysis between binary and ternary complexes of hAgo2. The PDB files 4W5N and 4W5O were used for binary and ternary complex representation of hAgo2. The colors from blue to white and to red indicate B factors from small to large. The structures are shown in surface representation.

Additionally flexibility comparisons between hAgo1 (4KRF) and hAgo2 (4W5N) by B-factors, reveal hAgo1 was generally more stable than hAgo2, but with a very flexible spot in the PAZ domain (Figure 5.20). From the sequence alignment it is evident hAgo1 and hAgo2 differ by a few amino acids in the PAZ domain (Figure 5.21). This region is covered by the chimera $Ago2^{Ago1\_PAZ1}$. From the structure of the protein, those differences are in a random coil located at the first half of PAZ domain (Figure 5.20). Because of the major differences observed between hAgo2

and Ago2^{Ago1_PAZ1}, this indicates that the flexible coil might play an important role in miRNA binding.

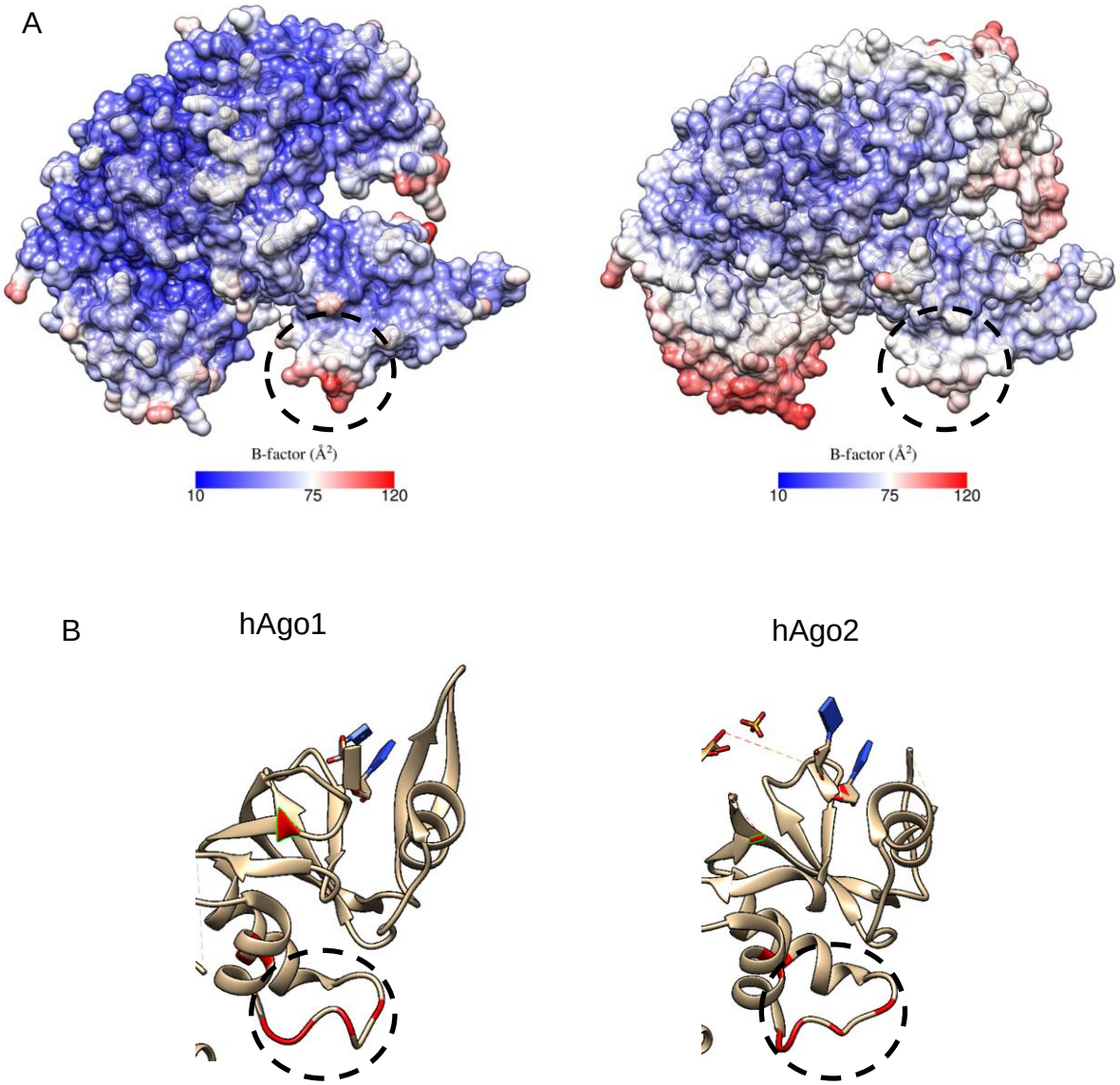


Figure 5.20 Structure and B-factor difference of PAZ domain between hAgo1 and hAgo2. The PDB files 4KRF and 4W5N were used for hAgo1 and hAgo2, respectively. Circles indicate the difference in the PAZ domain. (A) Normalized analysis of binary complexes of hAgo1 and hAgo2 are shown. The structures are shown as surfaces and the colors from blue to white and to red indicate B factors from small to large. (B) Structure differences in PAZ domain (ribbon representation) are shown. Differences in amino acid positions are labeled in red.

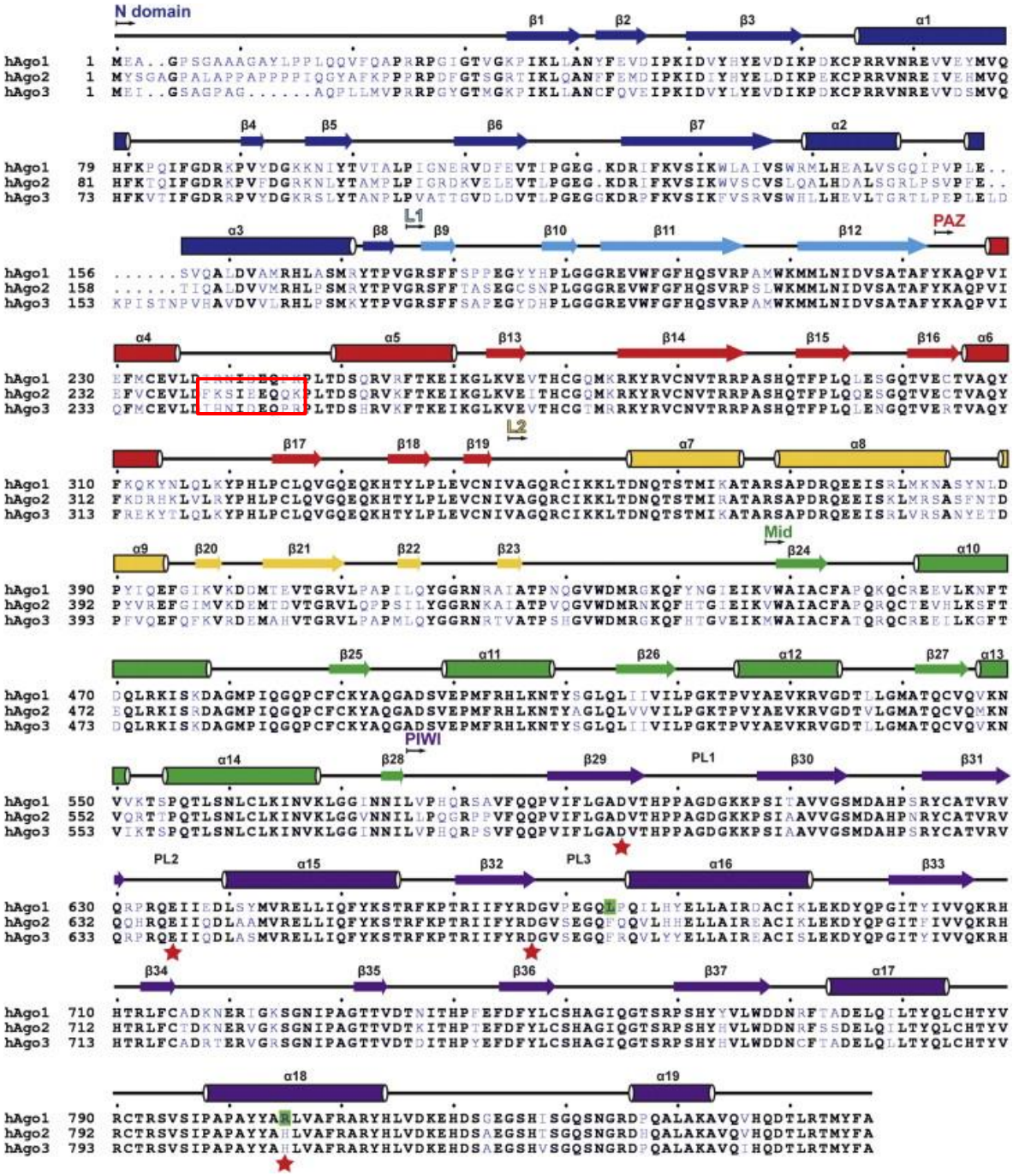


Figure 5.21 Sequence alignments of hAgo1, hAgo2, and hAgo3. Secondary structure elements extracted from the hAgo1 structure are shown above. Conserved amino acids are in black and non-conserved amino acids are shaded blue. The location of the PL1, PL2 and PL3 loops are labeled. Residues that make up the catalytic tetrad are indicated with a red asterisk below. Amino acids differences between hAgo2 and Ago2^{Ago1_PAZ1} are labeled by a red box. Modified from Faehnle *et al.* (55), with kind permission of Elsevier.

5.4. CRISPR-Cas9 system

As described in the introduction, the CRISPR system works as adaptive bacterial immunity against viruses and plasmids, using RNA to guide the silencing of invading DNA. Cas9 from the CRISPR/Cas type II system is becoming a powerful tool in genome engineering because it is the sole protein responsible for the RNA-guided DNA interference. In this thesis, the used techniques to obtain kinetic data of Ago were applied to study the mechanism of CRISPR-Cas system.

5.4.1. Expression and purification of Cas9

Streptococcus pyogenes Cas9 (SpyCas9) was expressed from a pET-based T7 promoter-containing plasmid (pMJ806, from Addgene, www.addgene.org) in the *E. coli* strain BL21 Gold DE3. The fusion protein construct contains an N-terminal His6-tag, followed by a maltose-binding protein (MBP) polypeptide, a tobacco etch virus (TEV) protease cleavage site, and the SpyCas9 sequence spanning residues 1–1368 (Figure 5.22). dCas9 (pMJ841, from Addgene) is constructed the same way expect there are two amino acids mutated, Aspartate 10 to Alanine (D10A) and Histidine 840 to Alanine (H840A), which inactivated the enzyme catalytically. Both plasmids (pMJ806 and pMJ841) were transformed into *E. coli* BL21 (see section 4.5.4), and proteins were expressed as described in section 4.6.1 “protein expression”.

The purification protocol generally follows previously published procedures with minor modifications (87, 133) and includes two chromatography steps: immobilized metal ion affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). Briefly, cell lysate was bound in batch to Ni-NTA agarose and the protein was eluted in elution buffer containing 200 mM imidazole.

The His6-MBP affinity tag was removed by cleavage with TEV protease, then the cleaved Cas9 protein was separated from the fusion tag by purification via a Superdex 200 16/60 column. Eluted protein was analyzed by SDS-PAGE. dCas9 was purified following the same procedure as for the wild-type Cas9 protein. For method details see section 4.6.2 “protein purification”.

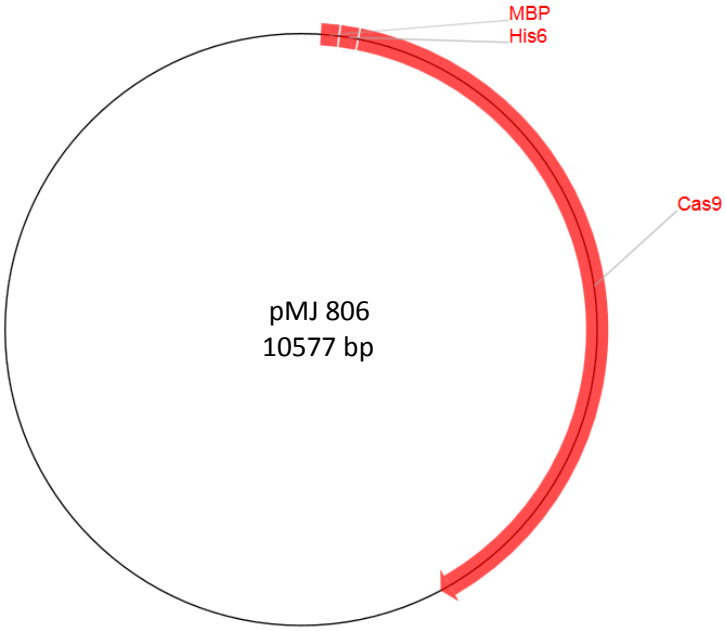


Figure 5.22 Plasmid map of pMJ806

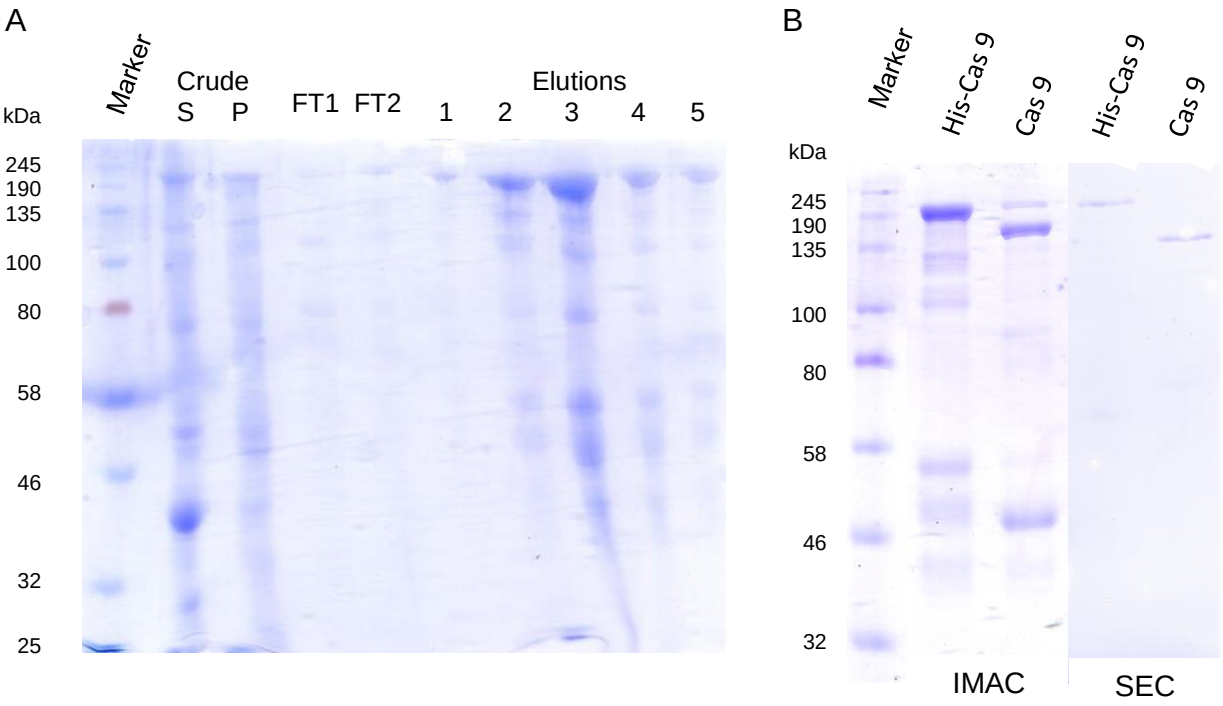


Figure 5.23 SDS-PAGE analysis of recombinant Cas9 purification. (A) First purification step by IMAC. Samples in the 8% SDS-PAGE from left to right are: protein ladder, crude protein supernatant (S) and pellet (P), flow through 1 (wash buffer; FT1), flow through 2 (wash buffer + 10 mM imidazole; FT2), and elutions (5 ml each; from 1 to 5). (B) His-tag cleaved by TEV protease followed by SEC to improve the purity. Samples in the 8% SDS-PAGE from left to right are: protein ladder, elution from IMAC, elution

treated by TEV protease, elution from SEC (with His-tag) and elution from SEC (His-tag cleaved).

As Figure 5.23 shows, His-Cas9 protein which has a molecular weight of 202 kDa carrying a His-tag was purified by the IMAC. Cas9 is reduced to 159 kDa when the His-tag was removed.

5.4.2. Single guide RNA preparation

Originally Cas9 is guided by two RNAs, that is why it is called dual-RNA guided DNA endonuclease. CRISPR RNA (crRNA) guide is composed of a 5´-terminal 20-nt guide sequence, followed by an invariant repeat-derived sequence at the 3´-end; trans-activating crRNA (tracrRNA) sequence remains identical for all guide crRNAs, so partially base-paired to the 3´-end of crRNAs (87). Alternatively, Cas9 can be guided by a chimeric single-guide RNA (sgRNA) that combine the essential parts of the crRNA and tracrRNA molecules in a single oligonucleotide chain (Figure 5.24).

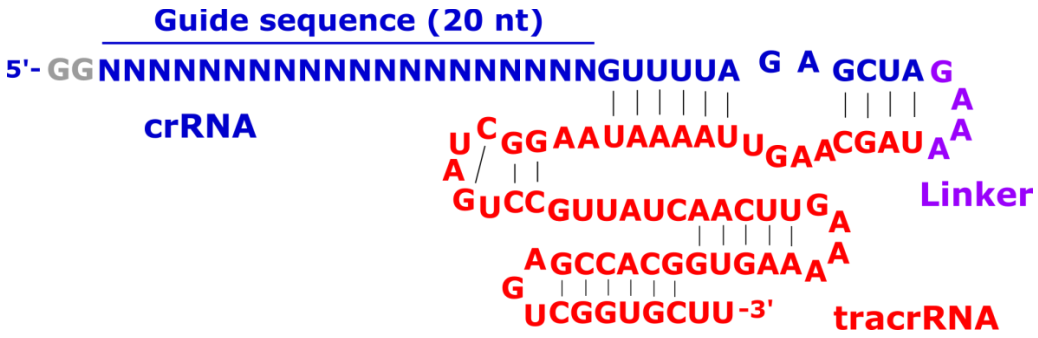


Figure 5.24 Schematic representation of chimeric sgRNA. sgRNA was designed to combine both sequences from crRNA and tracrRNA. The crRNA-derived (colored in blue) and tracrRNA-derived (colored in red) sequences are connected by a 5´-GAAA-3´ tetraloop linker (colored in purple).

In vitro transcription was performed to prepare sgRNA. The sequence was designed according to crystal structures (110), which is shown in Figure 5.25.

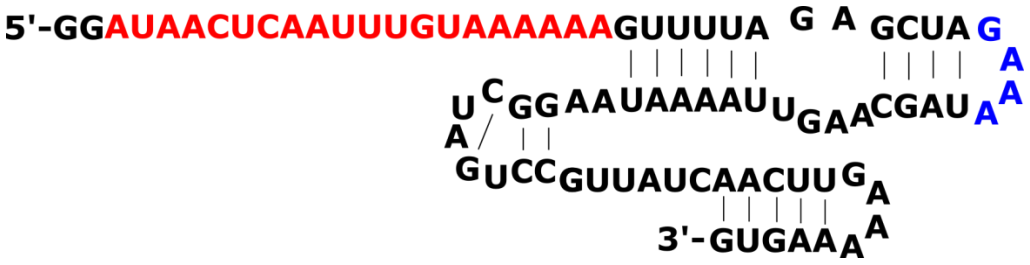


Figure 5.25 sgRNA used in this study. Annealed 83 nt RNA is shown. The 20 nt guide region is colored in red and the linker is in blue.

For *in vitro* transcription, a custom DNA template was used. A few different concentrations of NTPs and Mg^{2+} were tested to get the highest transcription yield, the used reaction condition is giving in Table 5.7. To get sufficient RNA, a 5 ml scale reaction for *in vitro* transcription was performed. After the reaction the product was mixed with loading buffer, loaded on a sequencing gel, the band of interest was excised and RNA was recovered by electroelution using a Biotrap device. The final RNA concentration was determined using spectrophotometry with an UV wavelength of 260 nm. The amount of pure RNA was ca. 2.2 mg.

Table 5.7 *In vitro* transcription reaction

	Final concentration
Transcription buffer	1x
Hybridized template*	1 µM
$MgCl_2$	30 mM
NTPs	5 mM each
DTT	5 mM
T7 RNA polymerase	0.1 mg ml^{-1}

*Hybridized template: first DNA template and T7 promoter oligonucleotide with a molar ratio of 1:1.5 were denatured at 90°C for 5 min and second annealed while letting the mixture slowly cool down to room temperature.

5.4.3. Cleavage and gel shift assay using Cas9

After the preparation of Cas9 and sgRNA, the endonuclease function of Cas9 was tested by cleavage assays. FAM-labeled 31 nt long oligonucleotide duplexes were used in cleavage assays. The FAM-label is positioned at the 3´-end of the target strand. The complementary unlabeled non-target strand contains the PAM 5´-TGG-3´ sequence (see 3.4.1). Target and nontarget strands were annealed by mixing equimolar amounts to generate the oligonucleotide duplex (see 4.7.5). Using the FAM-labeled substrate in the cleavage assay of Cas9 leads to a cleaved product with a length of 17 nt.

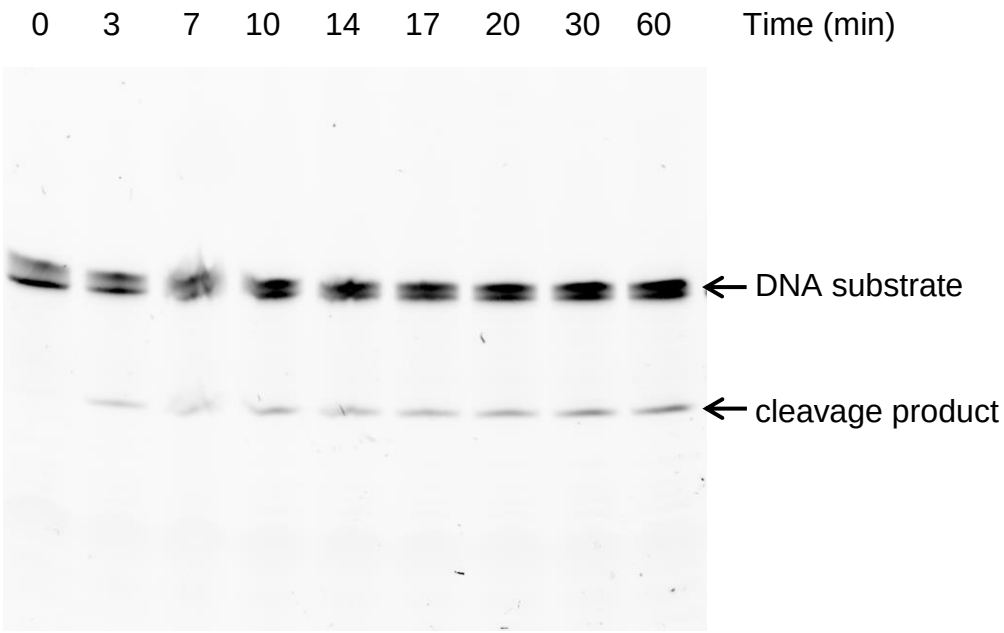


Figure 5.26 Cleavage assay of Cas9 using a double-stranded DNA target. Cleavage assays were conducted using 100 nM Cas9-sgRNA and 10 µM DNA duplex substrate. The cleavage reaction was stopped at indicated time points and analyzed by electrophoresis on a 16% denaturing (7 M urea) polyacrylamide gel. FAM caused fluorescence was detected using a FLA9500 laser scanner.

To determine the cleavage rate of Cas9, cleavage assays were performed applying single turnover and multiple turnover conditions, in which different ratios of Cas9-sgRNA and DNA target were used. The results showed Cas9 is a multiple

turnover endonuclease. The rate of single turnover is 0.14-0.17 s^{-1} and 0.005-0.01 s^{-1} for multiple turnover (Figure 5.27).

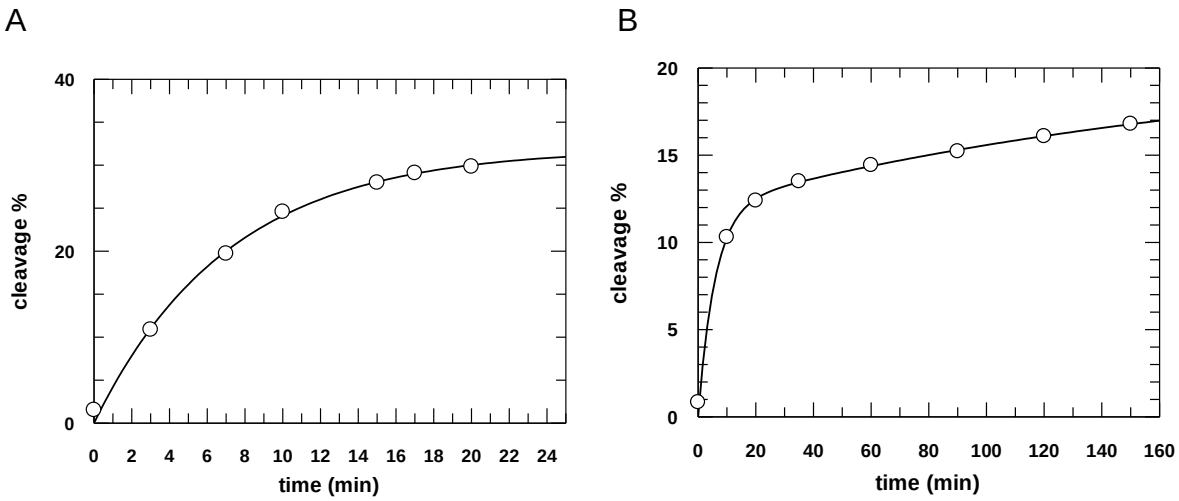
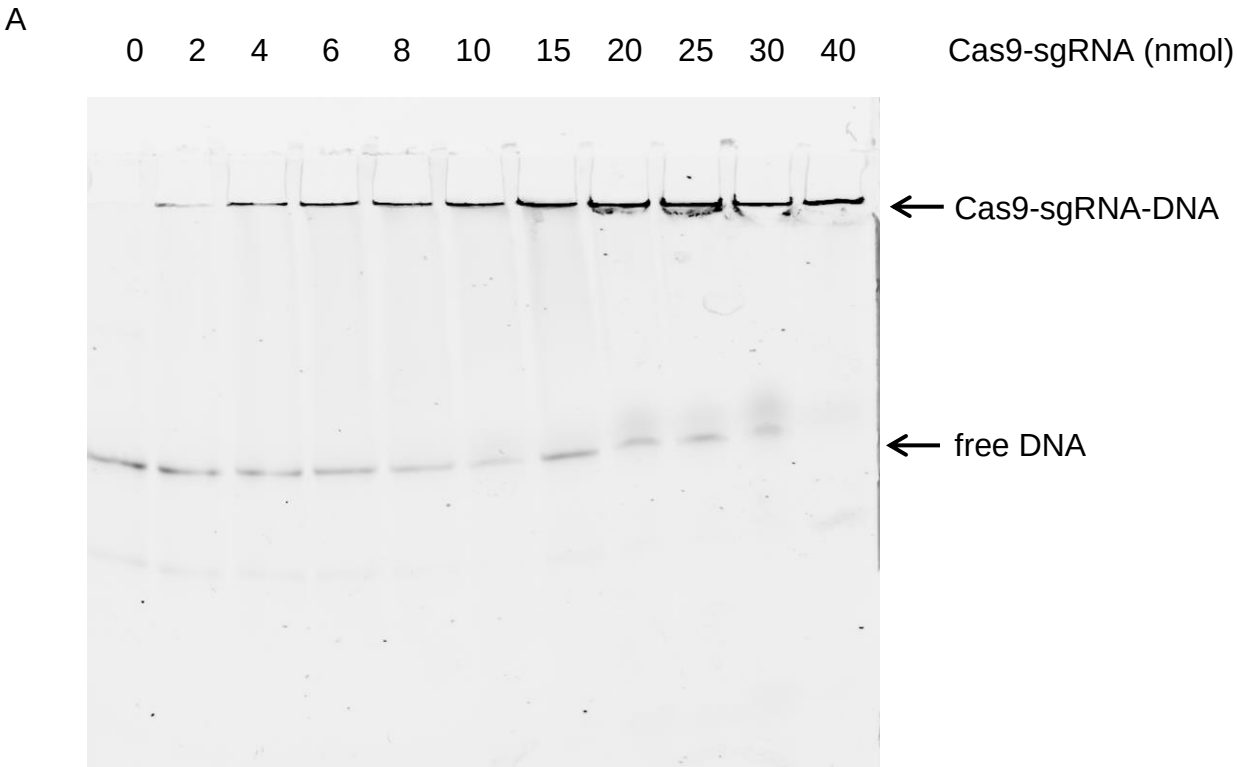


Figure 5.27 Cas9-catalyzed target-DNA cleavage reactions. 25 nM Cas9-sgRNA (A) or 1 nM Cas9-sgRNA (B) was pre-incubated for 10 min at 37°C in cleavage buffer. Reaction was started by adding 5 nM FAM-labeled target DNA duplex. Samples were taken at different time points and analyzed by denaturing PAGE. Fitting of the experimental data to a single exponential equation (A) yielded a rate constant of 0.14 (± 0.01) s^{-1} or to a double exponential equation (B) yielded rate constants of 0.17 (± 0.03) and 0.006 (± 0.002) s^{-1}.



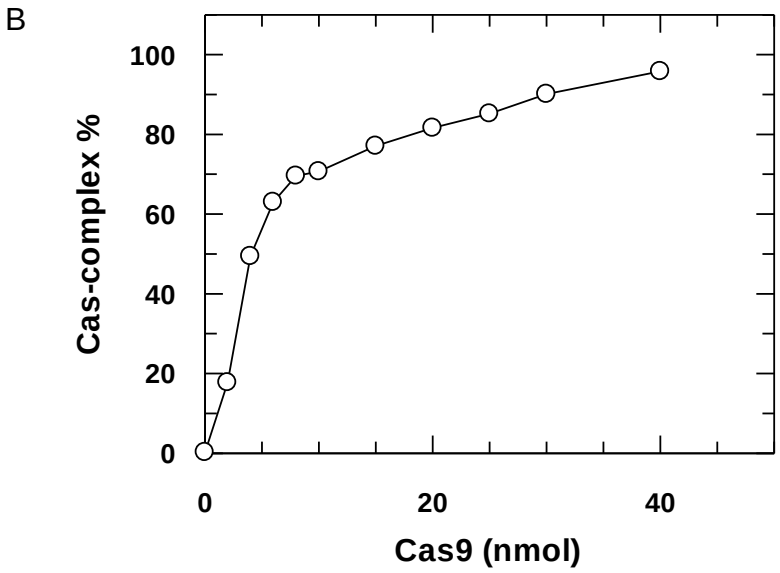


Figure 5.28 Gel shift assay of Cas9-sgRNA with dsDNA substrate. Gel shift assays were conducted using 10 nmol FAM labeled DNA duplex and the amount of Cas9-sgRNA was increased from 0 to 40 nmol. (A) Binding was analyzed by electrophoresis on a 16% native polyacrylamide gel and detected by a FLA9500 laser scanner. (B) Quantification of the shifted bands was done using Image Quant 5.2.

The gel shift, or electrophoretic mobility shift assay was used to detect Cas9-sgRNA/dsDNA complexes. The Cas9-sgRNA/dsDNA complexes were analyzed using native PAGE and the gel was scanned using the FLA9500 laser scanner. By increasing the amount of Cas9-sgRNA, the amount of free DNA substrates is decreased. The final percentage of Cas9-sgRNA/DNA complex reached up to 95% (Figure 5.28).

5.5. Mechanistic model of Cas9-sgRNA/nucleic acid interaction

After setting up the CRISPR-Cas9 system, it was tried to reveal the molecular mechanism of nucleic acid binding and cleavage. So it was focused on studies of Cas9-sgRNA dsDNA binding. Steady-state and pre-steady-state characterization revealed more molecular details about target DNA recognition and cleavage by Cas9.

5.5.1. Binding kinetics of Cas9

FAM-labeled DNA duplexes were used as the target substrates (Figure 5.29). The fluorescence signal change was recorded by using a fluorescence spectrometer or a stopped-flow instrument. Two dsDNAs with a FAM-label at different positions were used in the measurements.

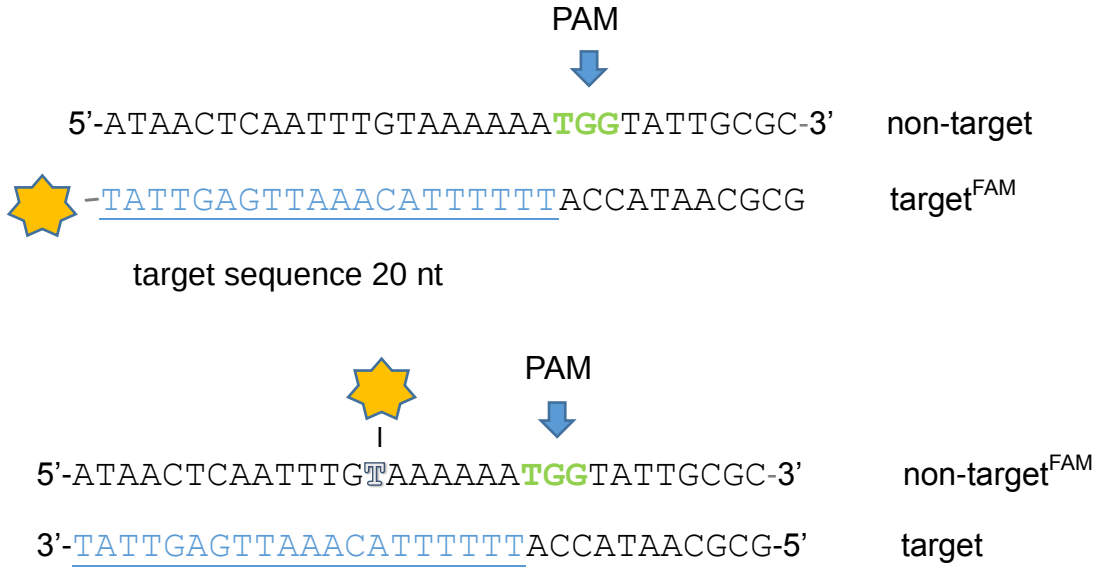


Figure 5.29 Schematic representation of FAM-labeled DNA duplexes. Boldface letters indicate the position of the PAM and the guide RNA matching sequence in the target DNA is underlined. The FAM label is located at the 3´-end of target^{FAM} or at position 14 of the non-target^{FAM} (indicated by an asterisk).

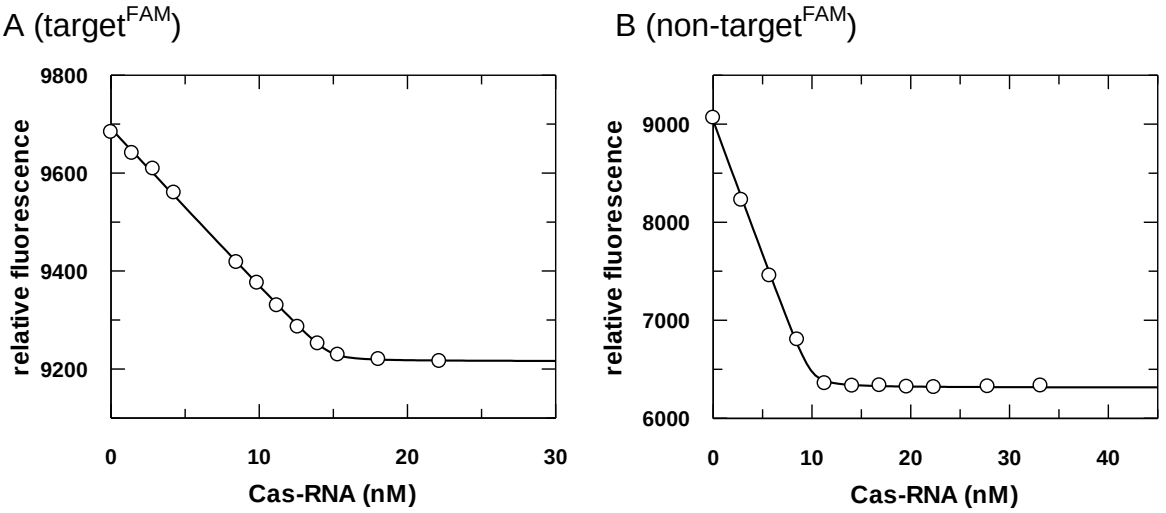


Figure 5.30 Equilibrium titrations of DNA duplexes and Cas9-sgRNA. DNA duplexes of target^{FAM} (A) or non-target^{FAM} (B) were titrated with increasing concentrations of Cas9-

sgRNA complex. The best fit of the experimental data to a quadratic equation is shown by the curves and give values for respective K_{ds} of Cas/nucleic acid complexes of 0.03 (A) and 0.04 (B) nM.

Equilibrium titrations (Figure 5.30) revealed a very tight binding between Cas9-sgRNA and target DNA. The binding affinities of target^{FAM} and non-target^{FAM} are very similar and both in the range of 20-50 pM.

Pre-steady-state binding kinetics showed a three or four-step binding pathway (Figure 5.31). In the presence of Mg^{2+} there are three phases for Cas9-sgRNA binding to target^{FAM} and four phases to non-target^{FAM}. The second phase of Cas9-sgRNA non-target^{FAM} binding most likely represents an additional unrelated phase compared to target^{FAM}, because the other three phases have similar rate constants with the three from target^{FAM}.

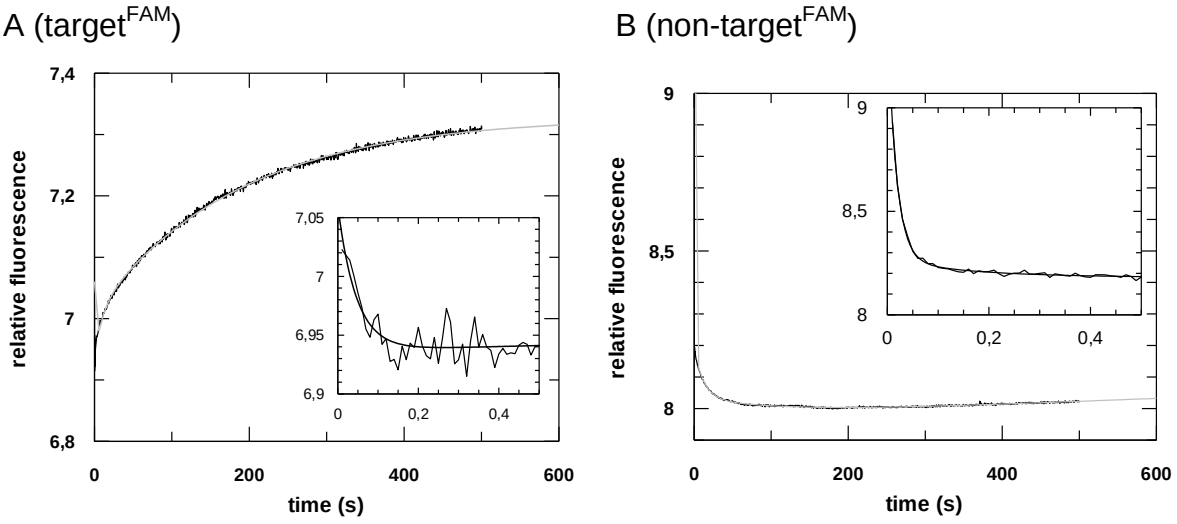


Figure 5.31 Pre-steady-state kinetic analysis of DNA duplex Cas9-sgRNA association. Exemplary stopped-flow results are shown. Insets show the reactions on a shorter time scale. Binding of Cas9-sgRNA complexes (100 nM Cas9 and 100 nM sgRNA) to 20 nM target^{FAM} (A) or non-target^{FAM} (B) DNA duplex is shown. The binding buffer contained 2 mM Mg^{2+}. For (A), the data were fitted to a triple exponential equation yielding three phases with the following rates: k_{+1_obs}: 21.7 (±1.1) s^{-1}, k_{+2}: 0.1 (±0.004) s^{-1} and k_{+3}: 0.006 (±0.00004) s^{-1}. Data of non-target^{FAM} (B) were fitted to a quadruple exponential equation yielding four phases: k_{+1_obs}: 58 (±0.7) s^{-1}, k_{+2}: 5.3 (±0.5) s^{-1}, k_{+3}: 0.1 (±0.002) s^{-1} and k_{+4}: 0.009 (±0.0006) s^{-1}.

The fast first phase was dependent on the protein concentration while the others were not. To relate the kinetically defined states to particular binding steps or corresponding structural transitions, a comparison of them was made to the rates from the cleavage assay. The analysis shows the second phase of Cas9-sgRNA dsDNA (target^{FAM}) complex formation (or third phase of Cas9-sgRNA dsDNA complex formation using non-target^{FAM}) is similar to the cleavage kinetic of single turnover and the third phase of Cas9-sgRNA dsDNA (target^{FAM}) complex formation (or fourth phase of Cas9-sgRNA dsDNA complex formation using non-target^{FAM}) is similar to the kinetic of multiple turnover.

Notably, the second phase of Cas9-sgRNA and non-target^{FAM} complex formation was missing using target^{FAM}. This could represent an interaction of PAM and Cas9. The PAM interacting domain is located in the core of the protein. So the position of FAM may explain the missing phase in Cas9-sgRNA target^{FAM} complex formation. Since FAM is positioned at the 3´-end on target^{FAM} it is far from the protein center compared to FAM on non-target^{FAM}. Thus, the PAM interaction may be too limited to cause any change in a fluorophor that far away.

5.5.2. The effects of Mg^{2+} on binding and cleavage

To further investigate the binding kinetics, the cleavage reaction was analyzed without Mg^{2+} (Figure 5.32). Because Cas9 is a Mg^{2+}-dependent endonuclease, the target DNA can not be cleaved in the absence of Mg^{2+}.

Compared to the Cas9-sgRNA dsDNA complex formation in the presence of Mg^{2+}, the last phase with the slowest rate is missing, which could represent the “cleavage state”. But the third phase remained in the absence of Mg^{2+}, which might represent a transition independent of Mg^{2+}. Additionally, the second phase of Cas9-sgRNA non-target^{FAM} complex formation showed up using target^{FAM}, which was considered could not being detectable in the presence of Mg^{2+} (Figure 5.31) because of the FAM position (see discussion section 6.5). It may represent a rearrangement of the protein after binding.

A (target^{FAM}) B (non-target^{FAM})

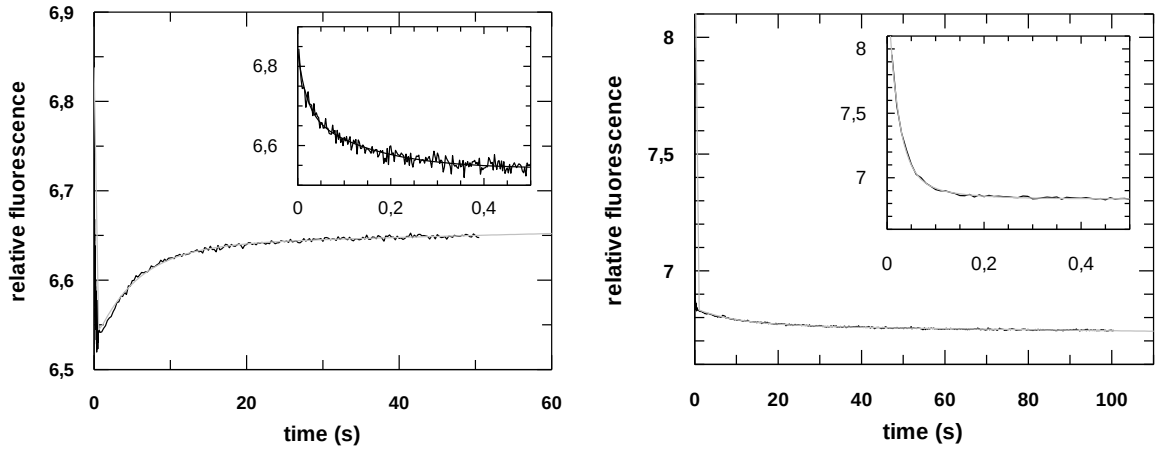


Figure 5.32 Pre-steady-state kinetic analysis of Cas9-sgRNA binding to dsDNA without Mg^{2+}. The same measurements were performed as described in Figure 5.31 but with a buffer containing 2 mM EDTA instead of 2 mM Mg^{2+}. The data were fitted to a triple exponential equation with the following rates for target^{FAM} (A): k_{+1_obs}: 53.3 (±6.4) s^{-1}, k_{+2}: 3.4 (±0.5) s^{-1} and k_{+3}: 0.18 (±0.015) s^{-1}; and (B) k_{+1_obs}: 47.6 (±0.9) s^{-1}, k_{+2}: 6.5 (±0.7) s^{-1} and k_{+3}: 0.1 (±0.003) s^{-1} for non-target^{FAM}.

Cas9 is a Mg^{2+}-dependent endonuclease with 3'-5' exonuclease activity (87). Therefore, it was attempted to investigate the pre-steady-state kinetics of the cleavage reaction triggered by the addition of $MgCl_2$. Figure 5.33 shows the fluorescence change of pre-assembled Cas9-sgRNA/dsDNA complex rapidly mixed with Mg^{2+}.

A (target^{FAM}) B (non-target^{FAM})

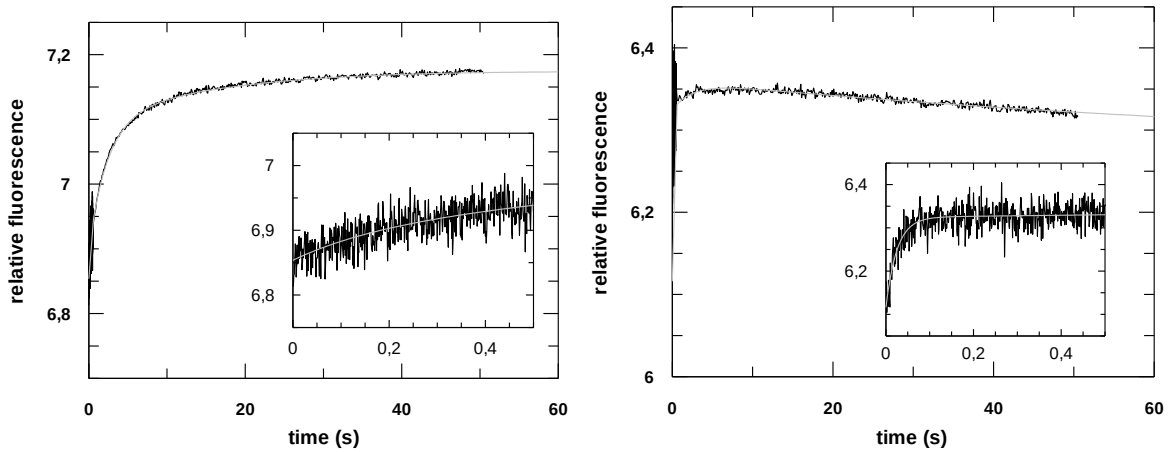


Figure 5.33 Pre-steady-state kinetic analysis of Cas9 cleavage reaction triggered by Mg^{2+}. Cas9-sgRNA (50 nM) was bound to DNA duplex (20 nM) of target ^{FAM} (A) or non-

target ^{FAM} (B) in buffer without Mg^{2+}, was rapidly mixed with buffer containing 4 mM Mg^{2+}. The data were fitted to a triple exponential equation yielding three phases, with the following rates for target^{FAM}: (A) k_{+1_obs}: 5.5 (±1) s^{-1}, k_{+2}: 0.4 (±0.04) s^{-1} and k_{+3}: 0.05 (±0.02) s^{-1} and (B) k_{+1_obs}: 39.8 (±1.7) s^{-1}, k_{+2}: 0.4 (±0.09) s^{-1} and k_{+3}: 0.011 (±0.003) s^{-1} for non-target ^{FAM}.

The first phase showed large fluctuation regarding amplitude and rate constants. This arises most likely from mixing of the fluorophor with Mg^{2+} and is not related to any conformational transition of protein/nucleic acid complex. But the second and third phases had very stable value and the rate constants are slightly faster but comparable to the kinetics observed in the cleavage assays (Figure 5.27). The second phase most likely represents a conformational change preceding cleavage (slicing associated step) and the third phase corresponds to cleavage product release. These results are in support of Cas9 is a multiple turnover endonuclease (87), and indicate the cleavage is Mg^{2+}-dependent but nucleic acid binding is not.

5.5.3. Influence of PAM on Cas9 nucleic acid binding

To test whether PAM recognition is a requirement of Cas9-sgRNA dsDNA binding and cleaving, a DNA duplex with sequence changes (5'-TGG-3' to 5'-TTA-3') in the PAM region was used as a target (Figure 5.34). In cleavage assays, no cleavage product could be detected. In pre-steady-state analysis no signal change beside the fast first phase could be observed (Figure 5.35).

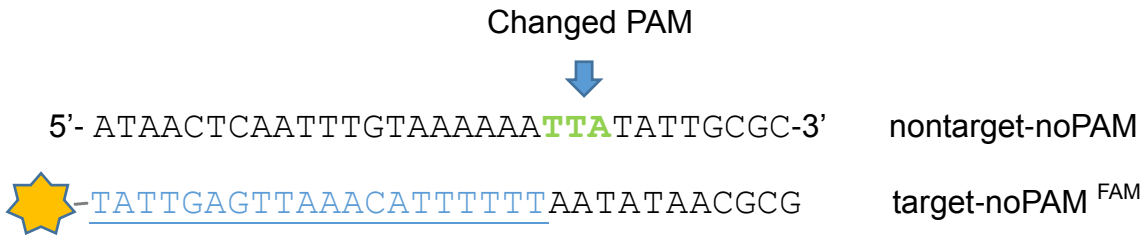


Figure 5.34 Schematic representation of modified PAM target duplex. Boldface green letters indicate the position of modified PAM sequence and the guide RNA matching sequence in target DNA is underlined. The FAM-label is located at the 3´-end of target-noPAM^{FAM} (indicated by an asterisk).

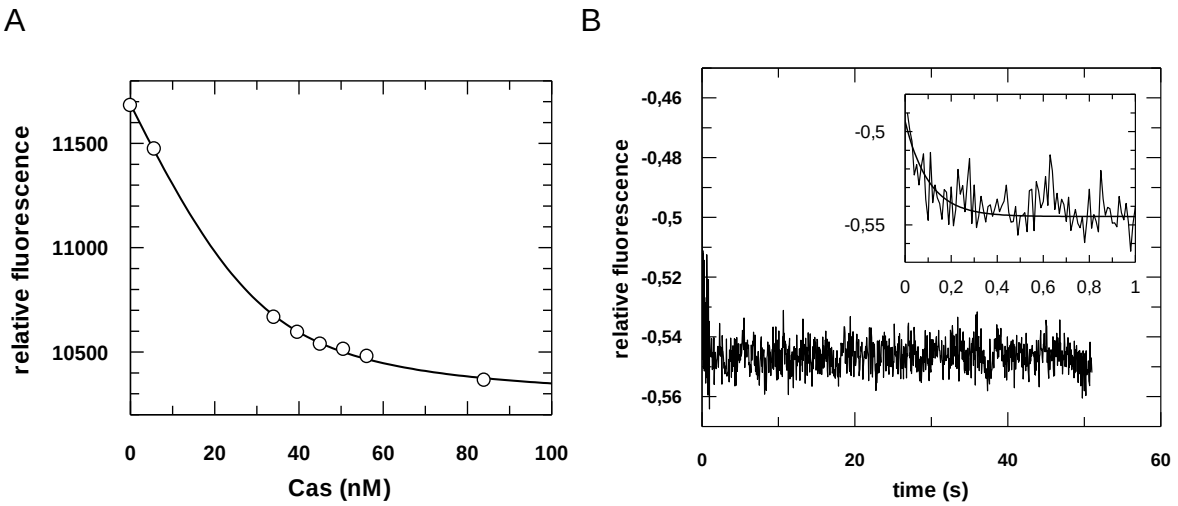


Figure 5.35 Equilibrium and pre-steady-state analysis of binding of Cas9-sgRNA to the modified PAM duplex. (A) dsDNA with modified PAM was titrated with increasing concentrations of Cas9-sgRNA complex. The best fit of the experimental data to a quadratic equation is shown by the curve, and give a value for respective K_d of 5.4 (±0.5) nM. (B) Kinetic data of complex formation of dsDNA with modified PAM (20 nM) using Cas9-sgRNA (100 nM) (obtained by stopped-flow measurements). Experimental data were fitted to a single exponential equation yielding a rate constant of k_{+1_obs}: 21.5 (±2) s^{-1}.

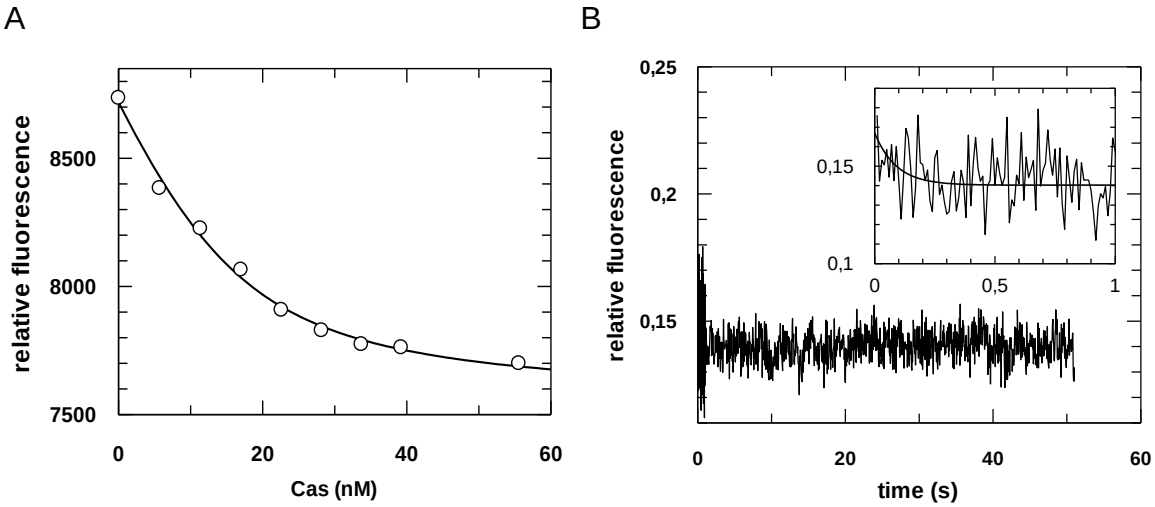


Figure 5.36 Equilibrium and pre-steady-state analysis of Cas9 DNA interaction in the absence sgRNA. (A) dsDNA target was titrated with increasing concentrations of Cas9. The best fit of the experimental data to a quadratic equation is shown by the curve, and give a value for respective K_d of 5.6 (±1.2) nM. (B) Stopped-flow results of dsDNA (20 nM) and Cas9 protein association (100 nM). Experimental data were fitted to a single exponential equation yielding a rate constant of k_{+1_obs}: 13.2 (±2.5) s^{-1}.

Interestingly, the K_d for the modified PAM target is 5.4 nM as determined by equilibrium titrations, which is more than 100-times lower as compared to the results obtained using a target with unmodified PAM sequence. Moreover, binding between Cas9 protein only and DNA duplex in the absence of sgRNA was analyzed as a negative control (Figure 5.36). This set up revealed comparable data to the experiment using the modified PAM target. The relatively tight binding (K_d of 5.6 nM) shows that Cas9 has a high affinity for DNA oligonucleotides, this tight binding could help searching for potential targets. On the other hand, usage of unmodified PAM sequence enhanced the Cas9-DNA binding affinity dramatically, resulting in an exact positioning of the nucleic acid for further interaction.

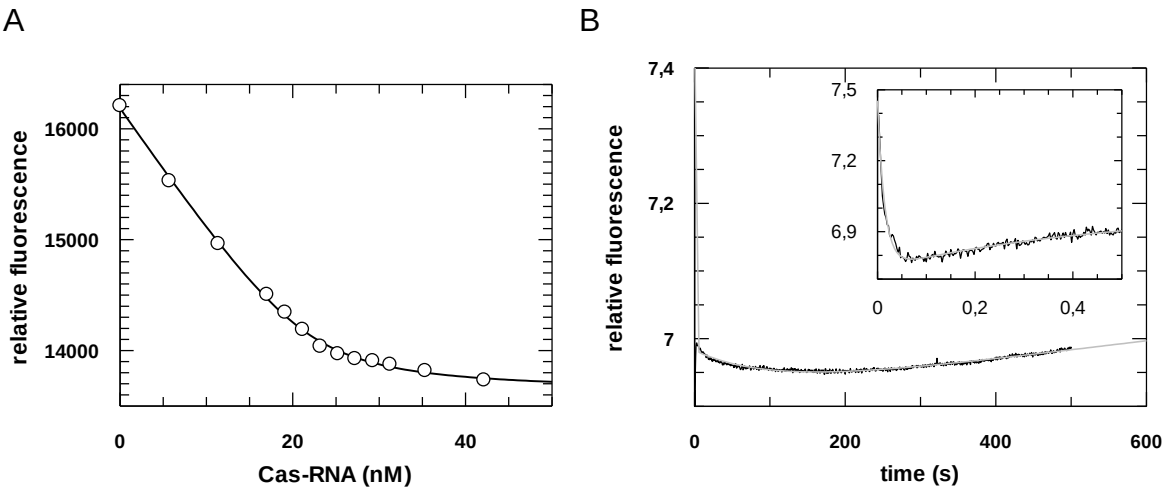


Figure 5.37 Equilibrium and pre-steady-state analysis of Cas9-sgRNA and ssDNA interaction. (A) ssDNA non-target^{FAM} was titrated with increasing concentrations of Cas9-sgRNA complex. The best fit of the experimental data to a quadratic equation is shown by the curve with a value for the respective K_d of 1.3 (±0.3) nM. (B) Stopped-flow results of Cas9-sgRNA complex (100 nM) ssDNA target non-target^{FAM} (20 nM) association. Experimental data were fitted to a triple exponential equation yielding three phases with the following rates: k_{+1_obs}: 67.2 (±0.9) s^{-1}, k_{+2}: 2.2 (±0.03) s^{-1} and k_{+3}: 0.01 (±0.0007) s^{-1}.

To further investigate the importance of the PAM sequence, an analysis of the binding affinity was performed using PAM only target to Cas9-sgRNA. Here, a non-target^{FAM} ssDNA was used, containing the PAM sequence but without a target strand pairing with sgRNA. The observed affinity (1.3 nM, Figure 5.37A) was much lower than for the duplex (0.04 nM, see Figure 5.30), but higher than for a dsDNA

without PAM sequence (> 5 nM; see Figure 5.35A). Pre-steady-state kinetic analyzes revealed the presence of three phases (Figure 5.37B). Compared to the four phases of Cas9-sgRNA dsDNA complex formation (Figure 5.31B), one phase was not detected. Most likely the missing phase refers to the third phase of Cas9-sgRNA dsDNA complex formation, which is supposed to represent unwinding of dsDNA. Here, the remaining three phases of Cas9-sgRNA ssDNA complex formation are suggested to represent (1) collision complex formation, (2) structural transition triggered by PAM recognition and (3) target DNA release, respectively.

5.5.4. dCas9 missed one transition upon substrate binding

To further demonstrate that the slow phase observed during transient binding analysis indeed represents a structural transition triggered by cleavage (compare section 5.5.2), dCas9 was used as a reference. There are three phases observed from pre-steady-state kinetic analysis (Figure 5.38A), which are very similar to Cas9-sgRNA dsDNA binding without Mg^{2+}. Compared to the results obtained with active Cas9 (Figure 5.38B), the fourth phase is missing, which is the slow phase equivalent to multiple turnover cleavage kinetic. This indicates the slowest phase represents the release of cleavage products, which represents the rate limit step of the cleavage reaction under multiple turnover conditions.

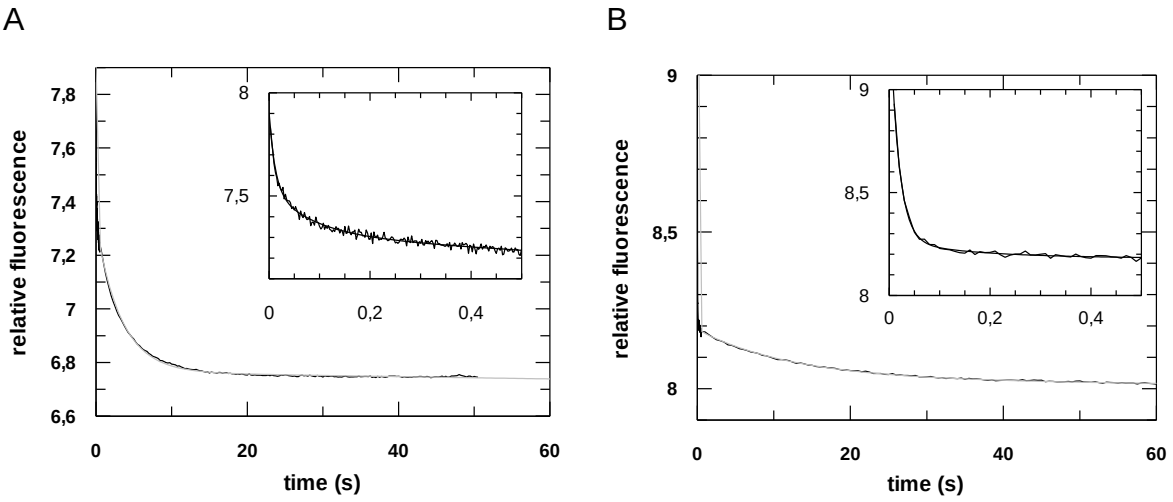


Figure 5.38 Pre-steady-state kinetic analysis of dCas9 (or Cas9)-RNA binding to dsDNA. The stopped flow data show binding of dCas9-sgRNA (A) or Cas9-sgRNA (B) complex (100 nM) to 20 nM target DNA duplex using non-target^{FAM}. The binding buffer contained 2 mM Mg^{2+}. For dCas9, experimental data were fitted to a triple exponential

equation yielding three phases with the following rates: k_{+1_obs}: 84.5 (±6) s^{-1}, k_{+2}: 3.6 (±0.3) s^{-1} and k_{+3}: 0.3 (±0.005) s^{-1}. For Cas9, the data were fitted to a quadruple exponential equation yielding four phases with the following rates: k_{+1_obs}: 56.9 (±1.9) s^{-1}, k_{+2}: 3.5 (±0.4) s^{-1}, k_{+3}: 0.11 (±0.004) s^{-1} and k_{+4}: 0.014 (±0.0005) s^{-1}.

The third phase in this experiment (Figure 5.38A) definitely does not represent the cleavage reaction because dCas9 is catalytic inactive. Moreover, the third phase here (k_{+3} of 0.3 s^{-1}) is slightly faster than the corresponding phase in Cas9-sgRNA dsDNA complex formation (k_{+3} of 0.11 s^{-1}) or the single turnover rate. It suggests this third phase might be a transition preceding cleavage, for example target duplex unwinding or RNA-DNA heteroduplex formation. This indicates there is a step preceding cleavage, which is detected during dCas9-sgRNA dsDNA substrate association (k_{+3} in Figure 5.38A).

Next, the dissociations of dCas9-sgRNA/dsDNA or Cas9-sgRNA/dsDNA complexes were analyzed by using 33-fold excess of non-labeled DNA as competitor (Figure 5.39). Both graphs showed opposite signal change compared to association experiments.

A (dCas9-sgRNA/dsDNA) B (Cas9-sgRNA/dsDNA)

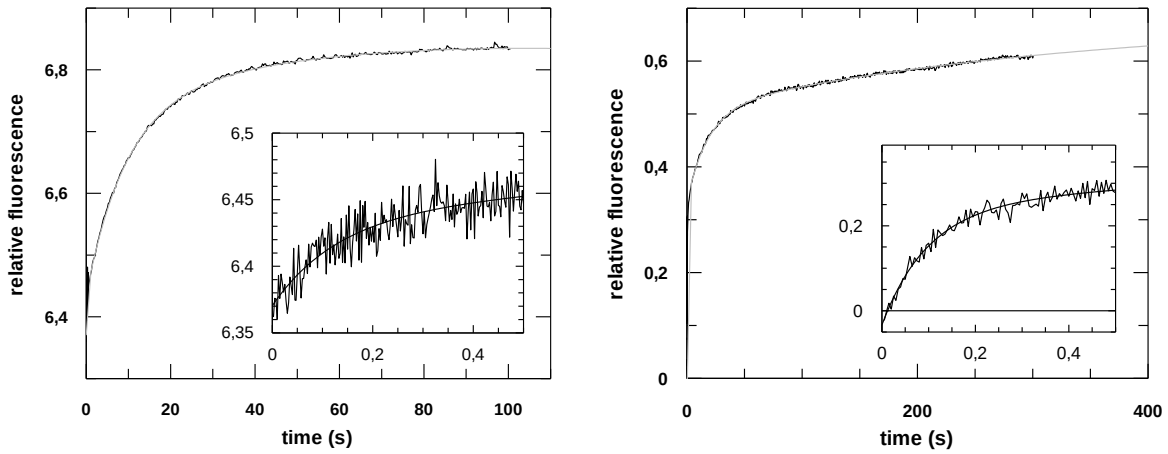


Figure 5.39 Pre-steady-state kinetic analysis of the dissociation of dCas9 (or Cas9)-RNA/dsDNA complexes. 20 nM DNA duplex using non-target^{FAM} was pre-incubated with dCas9-sgRNA (A) or Cas9-sgRNA (B) complex and then rapidly mixed with a 33-fold excess of non-labeled competitor dsDNA. The change in fluorescence signal was detected over time to determine dissociation rate constants. For dCas9, experimental data

were fitted to a triple exponential equation yielding three phases with the following rates: k_{-1}: 7.6 (±0.6) s^{-1}, k_{-2}: 0.1 (±0.004) s^{-1} and k_{-3}: 0.012 (±0.002) s^{-1}. For Cas9, the data were fitted to a quadruple exponential equation yielding four phases with the following rates: k_{-1}: 10.3 (±0.4) s^{-1}, k_{-2}: 1.1 (±0.1) s^{-1}, k_{-3}: 0.05 (±0.002) s^{-1} and k_{-4}: 0.003 (±0.0005) s^{-1}.

Sum up, to investigate the molecular mechanism of CRISPR-Cas9 system, steady-state and pre-steady-state studies of Cas9-sgRNA DNA binding were performed. The overall kinetic data is shown in Table 5.8.

Table 5.8 Summary of equilibrium and pre-steady-state data of Cas9-sgRNA/DNA complex formation. A comparison of Cas9-sgRNA DNA binding under different conditions. Here, non-target^{FAM} used as ssDNA is, and pre-annealed target/non-target^{FAM} used as dsDNA. Unless otherwise indicated, the binding buffer contains 2 mM Mg^{2+}.

	Cas9-sgRNA dsDNA	Cas9-sgRNA dsDNA without Mg^{2+}	Cas9-sgRNA ssDNA	Cas9-sgRNA dsDNA modified PAM	dCas9-sgRNA dsDNA
K_d (nM)	0.04	0.05	1.3	5.4	0.07
k_{+1_obs} (s^{-1})	58	47.6	67.2	21.5	84.5
k_{+2} (s^{-1})	5.3	6.5	2.2	-	3.6
k_{+3} (s^{-1})	0.1	0.1	0.01	-	0.3
k_{+4} (s^{-1})	0.009	-	-	-	-
k_{-1} (s^{-1})	10.3	12	14	n.d.	7.6
k_{-2} (s^{-1})	1.1	0.5	0.96	n.d.	0.1
k_{-3} (s^{-1})	0.05	0.01	0.11	n.d.	0.012
k_{-4} (s^{-1})	0.003	-	-	n.d.	-

5.5.5. Influence of the seed sequence on Cas9 nucleic acid affinity

There is a 20 nt long sequence within the target DNA which hybridises with the sgRNA whereas not the entire stretch does contribute to recognition and binding (49). To evaluate the effects of partial complementarity on binding activity, target strands of different lengths in the duplex region were analyzed (Figure 5.40). The pairing “seed” region was considered to consist of 8-12 nt (49), and seeds with 8, 12 or 16 nt length were tested (Figure 5.41). Full length target (seed20) was analyzed as standard target in section 5.5.1.

5’-ATAACTCAATTTGTAAAAAATGGTATTGCGC-3’ non-target FAM
3’- TATTGAGTTAAACATTTTTTACCATAACGCG-5’ target (seed20)

5’-ATAACTCAATTTGTAAAAAATGGTATTGCGC-3’ non-target FAM
3’- GAGTTAAACATTTTTTACCATAACGCG-5’ target (seed16)

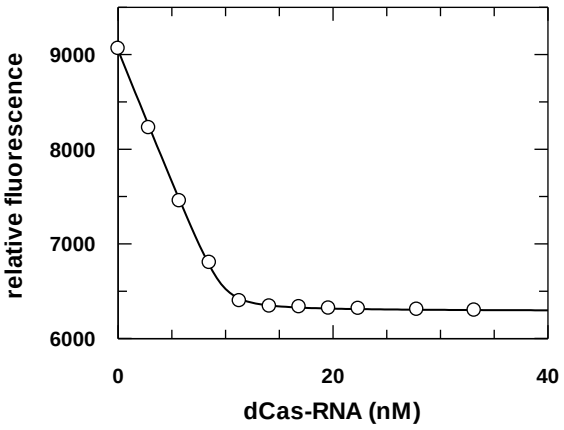
5’-ATAACTCAATTTGTAAAAAATGGTATTGCGC-3’ non-target FAM
3’- TAAACATTTTTTACCATAACGCG-5’ target (seed12)

5’-ATAACTCAATTTGTAAAAAATGGTATTGCGC-3’ non-target FAM
3’- CATTTTTTACCATAACGCG-5’ target (seed8)

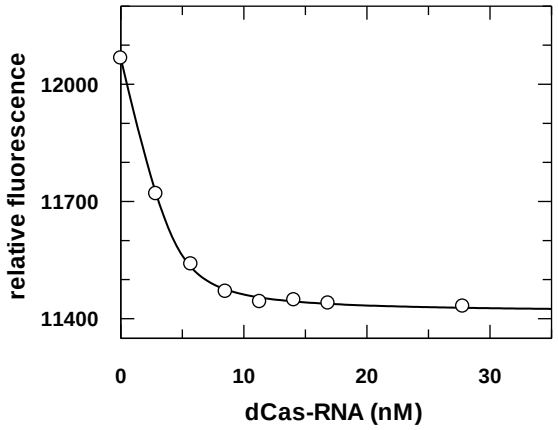
Figure 5.40 Schematic representation of targets with different seed lengths. Boldface letters indicate the position of the PAM. The guide RNA matching sequence in target DNA is underlined. The position of the FAM-label is marked with an outlined T.

Equilibrium titrations showed the affinities were decreased by reducing base pairing within the seed region. There is a positive correlation between affinity and length of the seed; a full length target has dozens-fold tighter binding affinity than a target with an 8 nt long seed region.

A (dsDNA-seed16)



B (dsDNA-seed12)



C (dsDNA-seed8)

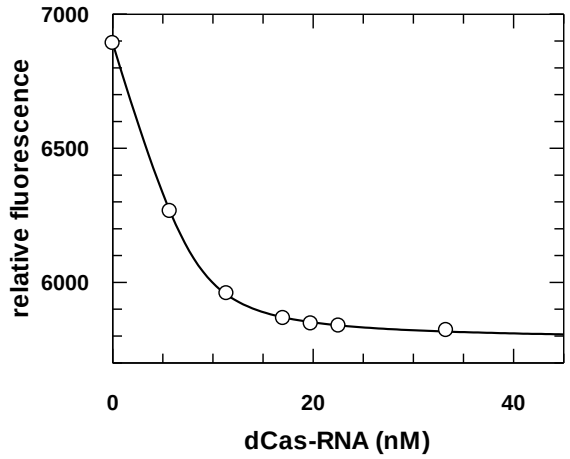
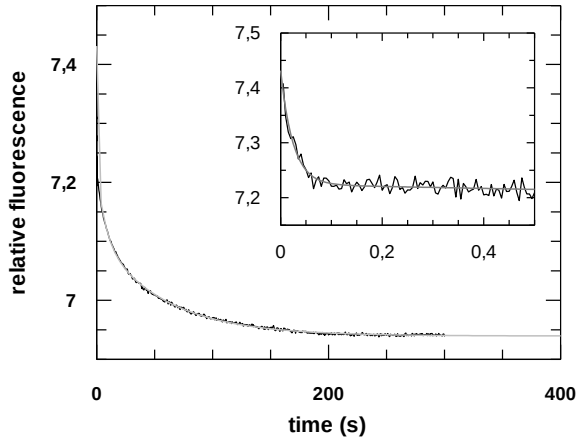
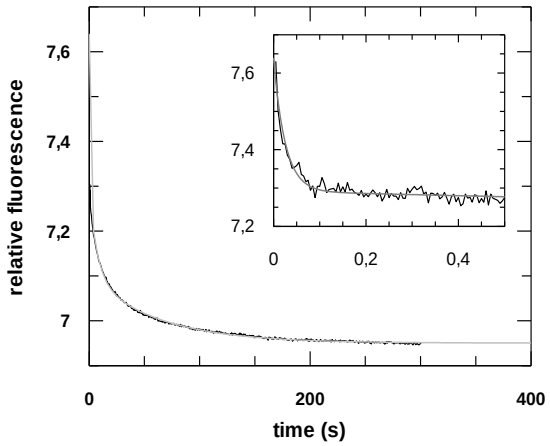


Figure 5.41 Equilibrium titrations of different DNA duplexes with dCas-sgRNA. DNA duplexes with partially complementary targets were titrated with increasing concentrations of dCas9-sgRNA complex. The best fits of the experimental data to a quadratic equation are shown by the curves and reveal values for respective K_{ds} from seed16 to seed8 of 0.1 (±0.02), 0.45 (±0.09) and 1.2 (±0.09) nM, respectively.

A (dsDNA-seed16)



B (dsDNA-seed12)



C (dsDNA-seed8)

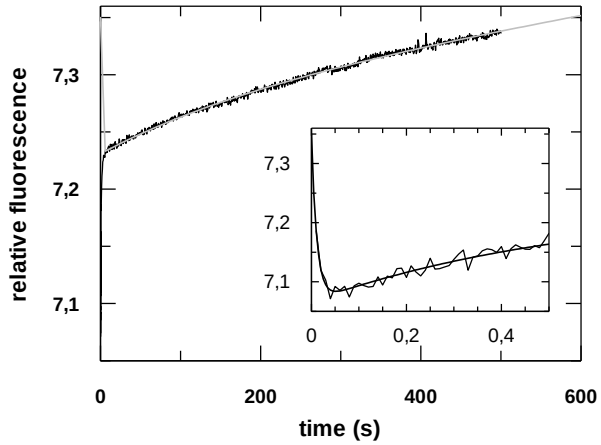


Figure 5.42 Pre-steady-state kinetic analysis of the association of different target duplexes with dCas9-sgRNA. Exemplary stopped-flow results of binding of dCas9-sgRNA complex (100 nM) to a partial complementarity target (20 nM) are shown. To prevent fluorescent changes arising from cleavage, binding experiments were performed with dCas9. Experimental data were fitted to a triple exponential equation yielding three phases with rates given in Table 5.9.

Pre-steady-state studies were performed to reveal the details of short target binding (Figure 5.42, Table 5.8). Since cleavage might interfere with the dissociation reaction, inactive dCas9 was used for the dissociation measurements here to exclude interference with substrate cleavage. The K_d was calculated from association and dissociation rate constants and fits well with the K_d from equilibrium measurements.

Table 5.9 Summary of equilibrium and pre-steady-state binding data for interaction of targets with different seeds with dCas9-sgRNA. Numbers to the left represent equilibrium measurements (K_{d_meas}). Numbers to the right were calculated from the corresponding association and dissociation rate constants (K_{d_cal}).

	$K_{d\text{-}meas}$ (nM)	k_{+1}(M^{-1} s^{-1})	k_{-1} (s^{-1})	k_{+2} (s^{-1})	k_{-2}(s^{-1})	k_{+3}(s^{-1})	k_{-3}(s^{-1})	$K_{d\text{-}cal}$
seed8	1.2	6.3×10^{8}	5.5	1.8	0.49	0.02	0.04	4.7
seed12	0.45	5.8×10^{8}	4.7	2.1	0.41	0.18	0.06	0.52
seed16	0.1	6.1×10^{8}	5.8	2.0	0.50	0.37	0.018	0.11
seed20	0.04	5.5×10^{8}	7.6	3.6	0.1	0.3	0.012	0.02

Compared to full length target, the biggest difference is observed for short target with 8 nt seed sequence. Here, the signal went in the opposite direction and the third phase k_{+3} was more than 10-fold slower compared to full length. Above, two binding steps were suggested including the second phase k_{+2} representing PAM interaction and the third phase k_{+3} representing RNA-DNA heteroduplex formation of RNA guide and DNA target (Figure 5.38A). The results obtained with the 8 nt seed sequence indicate that RNA-DNA heteroduplex formation might be obstructed because of too less base pairing in the seed region.

6. Discussion

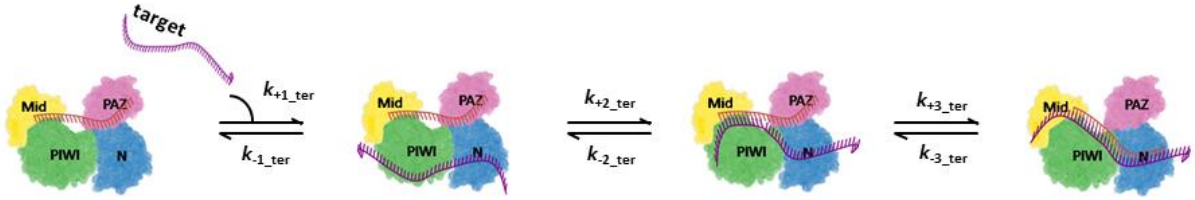
In this PhD study, the interactions of non-coding small nucleic acids and associated proteins were investigated in two distinct systems. The first one describes hAgo protein-mediated RNAi pathways. Based on a mechanistic model for the hAgo2-catalyzed RNA slicing reaction (128), the binding properties of hAgo1 and hAgo2 were analyzed in detail to determine potential differences between these two proteins. Relating transient kinetic data to conformational changes of hAgo nucleic acid complexes, it could be shown that hAgo1 has higher binding affinities for mismatched targets than hAgo2. This is a strong indication that hAgo1 is involved in the miRNA pathway.

Recently, CRISPR was discovered as prokaryotic adaptive immune system, including the associated enzyme Cas9 which is an RNA-guided endonuclease that uses guide RNA strands to target invading DNA (87). Numerous studies were reported in just a couple of years and Cas9 became a promising tool for genome engineering (70, 71, 100). The second project in this thesis is about CRISPR-Cas9. Here, the technique already used for kinetic studies of Ago was applied and adapted for the investigation of the CRISPR/Cas system. By relating the obtained experimental data to published structural and biochemical information, it was possible to establish a minimal mechanistic model for Cas9-sgRNA DNA target binding and cleavage, representing the first detailed model of such kind.

hAgo and SpyCas9 are nucleic acid assisted enzymes which are involved in gene regulation. Target recognition is accomplished by base pairing, and regulatory functions are fulfilled by binding to and cleaving of nucleic acids. Here, transient kinetic analyses were related to corresponding mechanistic steps in two proposed models. Furthermore, different targets as well as different conditions were tested, which provided in-depth insights into complex interactions of protein and corresponding target molecules. These findings allowed us to uncover new details regarding hAgo as well as Cas9 associated regulation.

6.1 A binding model for hAgo nucleic acid interaction

According to a previous study of the hAgo2-catalyzed RNA slicing reaction (45), three kinetically distinct binding steps were defined relating structural transitions to both binary and ternary complex formation. In the case of the ternary complex (Scheme 6.1), the model includes a concentration dependent fast first phase and a slower second and third phase corresponding to base pairing within the seed region and release of the 3´-end from the PAZ domain, respectively (45). In the present work an analogous binding mechanism was discovered for hAgo1 with similar kinetic values. Thus, hAgo1 shows similar binding properties than hAgo2 but lacks the cleavage function.



Scheme 6.1 Proposed kinetic model of ternary complex formation of hAgo2 and siRNA. The four domains of Ago (N, PAZ, MID and PIWI) are individually colored. The relative positions of protein and RNA substrates during binding are indicated. Firstly, hAgo2 forms a collision complex with the incoming target strand, followed by Watson-Crick base pairing of the target RNA with the bound guide RNA, starting from the prearranged seed-region. Further extension beyond nucleotide ~12 (calculated from the 5´-end of the guide) is restricted because of the 3´-end of the guide is bound to the PAZ domain. Accordingly, the release of 3´-end of the guide from the PAZ domain is required to complete guide and target hybridization. With kind permission of Dr. A. Deerberg (130).

The observed transitions for hAgo2 during target recognition are interpreted structurally in scheme 6.1. Two steps of target strand interaction (seed base pairing and 3´-end release) are also verified by a single-molecule FRET study of Argonaute from *Methanocaldococcus jannaschii* (MjAgo) (44). The interaction between Argonaute and DNA was studied using fluorescent dyes that were covalently attached as probes at different positions of the protein and the DNA guide. The FRET results demonstrated that the 3´-end of the guide strand is released from the PAZ domain upon target strand loading. This is in agreement with the pre-steady-state kinetic data of this thesis, which proposes that there is a

comparable binding mechanism for both prokaryotic and eukaryotic Ago-mediated silencing.

Studying hAgo1, kinetic data of ternary complex formation indicated three binding steps which are in agreement with the findings of hAgo2 (45). Furthermore, miRNA targets were tested for binding to hAgo1 or hAgo2, showing different binding affinities and kinetics (see section 5.2.2). Notably, both hAgo1 and hAgo2 can bind to mismatched targets. Compared to fully paired siRNA, miRNA binding was even tighter for hAgo1. In both binary and ternary complex formation, all mismatched targets tested showed stable binding to hAgo2 (Table 5.4), including target s2b-sm which only match in the seed region. These results suggest that a seed-match is sufficient for stable binding of a target to both hAgo1 and hAgo2. However, hAgo1 binds tighter to miRNA than hAgo2.

Recently, single-molecule studies provided detailed insights into Argonaute-nucleic acid interactions with a highlight on miRNA targeting (134, 135). Jo *et al.* (135) used the single-molecule fluorescence technique to characterize the interaction of hAgo2 with a small guide RNA. The study showed that target binding by Argonaute protein can result in four distinct reaction pathways: transient binding, stable binding, Argonaute unloading or target cleavage. Notably target cleavage required extensive sequence complementarity but stable binding to hAgo2 is efficiently established with the seed match only. This is supported by the study of Chandradoss *et al.* (136), which suggests that the seed region is the minimal motif required for stable target binding (nt position 2–8).

Furthermore, single-molecule imaging results published by Salomon *et al.* (134) showed Ago proteins reshape the fundamental properties of nucleic acid hybridization. In other words, Argonautes distinguish among substrates based on sequence complementarity between guide and target. For example, mouse Ago2 binds tighter to miRNA than to siRNA targets, even though siRNA forms more base pairs with the target. These findings reveal potential binding properties of Argonaute which are beyond the rules for nucleic acid hybridization and suggest

miRNA targets could interact with Argonaute in a more sophisticated manner. Similar to the results of Jo *et al.* (135), Salomon *et al.* also found that seed pairing is required for rapid target finding and stable binding.

These recent publications (134–136) strongly support the seed-match rule: Argonaute identifies miRNA targets by seed pairing, and this provides stable binding to target sites. Notably, more than just base pairing between nucleic acids, seed-match enables Argonautes to further regulate target binding and dissociation by reshaping the rules for nucleic acid hybridization (134). These new findings argue in support of tight binding of hAgo2 to miRNA targets. However, since present studies mainly focus on hAgo2, the role of hAgo1 in the miRNA pathway is still unclear. The kinetic data showed that miRNA binding properties of hAgo are beyond the rules for nucleic acid hybridization, which provides a potential explanation for miRNA target search and selection.

6.2 Argonaute and miRNA pathway

As described in section 6.1, seed-matching was reported to be sufficient for stable binding of hAgo2 and the target (135, 136). It was proposed that mammalian Ago2 compared to fly Ago2, might be more optimized for miRNA interaction than for siRNA (115). In the present study, it was found that hAgo1 and hAgo2 are both binding to miRNA. In the case of binary complexes, both hAgo1 and 2 bind to a siRNA duplex equally tight, while hAgo2 shows a 3-fold lower affinity for miRNA duplexes. Thus, hAgo1 appears to be more specialized for binding miRNA than hAgo2. This especially holds true for ternary complexes where hAgo2 shows significantly slower dissociation from complementary target RNAs which in turn favours target cleavage. It appears that both proteins are involved in the miRNA pathway, but Ago1 may be better suited for miRNA targets, because it binds faster and tighter to mismatched duplexes.

As introduced in section 2.1.2, many studies indicated that sRNAs are sorted into different Agos. This especially holds true for flies and worms (13). Many factors such as 1) Dicer that processes the precursor, 2) the structure of the small RNA

duplex, 3) the terminal nucleotides, 4) thermodynamic properties and 5) Ago proteins could influence sorting (13, 137). Additionally, several reports indicate the importance of a bias for U at the first position (48, 138–140). The sRNA sorting mechanism was mostly studied using the model system *Drosophila melanogaster*. Unlike in mammals, there are two distinct Dicer proteins processing small RNA duplexes in flies. Generally, miRNAs which are processed by Dicer 1 enter Ago1 complexes. siRNAs processed by Dicer 2 prefer Ago2 complexes. Besides in flies, sorting was also reported in *Caenorhabditis elegans* and *Arabidopsis thaliana* (23, 141). Worm miRNAs and siRNAs are partitioned among distinct Argonaute subfamily proteins. miRNAs are sorted into ALG-1 (Argonaute-like gene 1) or ALG-2, whereas siRNAs preferentially load to RDE-1 (RNAi defective 1) (141, 142). This also holds true for plant Argonautes: miRNAs processed by DCL1 (Dicer-like 1) are loaded to Ago1; heterochromatic sRNAs (hcRNAs) prefer DCL3 and Ago4 and siRNAs load to Ago2. The major question is how small RNAs are sorted and what is the function of Argonaute in sRNA sorting. In flies small RNAs are actively sorted into Ago-containing complexes according to their intrinsic structures. For example, a central mismatch favors sRNA loading into Ago1 (12). In plants, the preferential loading by distinct Ago proteins is related to specific 5´-nucleotides of the small RNAs. *Arabidopsis thaliana* Ago1 (AtAgo1) shows a strong bias towards a terminal U. AtAgo2 and AtAgo4 select sequences that begin with an A and AtAgo5 mainly binds to RNAs starting with a C (22, 23, 143).

So in some less complex organisms, there is a sorting mechanism ensuring that siRNAs and miRNAs are loaded to functionally distinct Ago-RISCs based on duplex structure and 5´-nucleotide. Generally, Ago1 prefers binding of miRNAs, whereas Ago2 associates with siRNAs. However, mammals seem to lack a strict system for such small RNA sorting, because the mammalian Ago clade proteins show no preference for specialized structural properties like for the 5´-nucleotide (52, 144). It seems this might have gotten lost in mammals during evolution. The sRNA duplexes with perfectly matched or bulged stems are loaded to either Ago1 or Ago2 without distinguishability. Therefore, it is outstanding unresolved how and why the sorting may have changed during the evolutionary journey.

Wee *et al.* (115) published information concerning functional domains of sRNA guides as determined by the interaction with Argonaute (Figure 6.1A). It was one of the references used for the design of the mismatch containing targets (Table 5.3, Figure 5.10). The observed kinetic data using different targets (Table 5.4) are in accordance with these proposed divided domains of sRNAs and might be an explanation for the sorting of miRNA. Moreover, according to their model (Figure 6.1B), fly Ago2-RISC forms in a cleavage competent conformation and cleaves the target RNA completely. Mouse Ago2-RISC on the other hand has a slow catalytic rate often allowing the target to escape before being sliced. Besides, siRNA from fly Ago2 dissociates much slower than from mouse Ago2 and mouse Ago2 binds much tighter to mismatched targets than fly Ago2 (115). All these results indicate that mammalian Ago2 evolved towards miRNA regulation even though it is the only enzymatically active Ago. From the basic sequence alignment of Ago1 and 2, it appears Ago2 is evolutionarily more active than Ago1. The comparison of human and mouse Ago sequences shows both Ago1 proteins display exactly the same sequence while there are seven amino acid exchanges for Ago2; there is a 73% similarity among Ago1 between human and fly, but only 33% for Ago2. So the apparent missing of a sorting mechanism is most probably due to the evolution of Ago2, which involves miRNA regulation, taking over the function from Ago1 protein.

However, there are four Argonaute proteins in humans. Except hAgo4, which is not expressed in most human cell lines and is the most divergent of the four proteins, the remaining hAgo 1-3 all show high expression levels (59). So it is believed that there are particular functions for the distinct Agos in humans although the details remain unclear. Based on the kinetic data and the sequence conservation of Ago1 among mammals, it was proposed that hAgo1 preferentially binds miRNA. It might be more efficient than hAgo2 because of the faster association of ternary complexes and it might be more specific because of a different tolerance to mismatch positions (compared to Ago2, Table 5.4). This strongly indicates that the selection mechanism in mammals is working differently compared to less complex organisms since hAgo2 is involved in miRNA regulation. But hAgo1 still might be the main player in the miRNA pathway; it is of major

importance because of some specific features as discussed. Moreover, a recently discovered Ago1-specific sRNA with novel functionality suggests the sorting mechanism is functional in mammals but maybe more sophisticated (27).

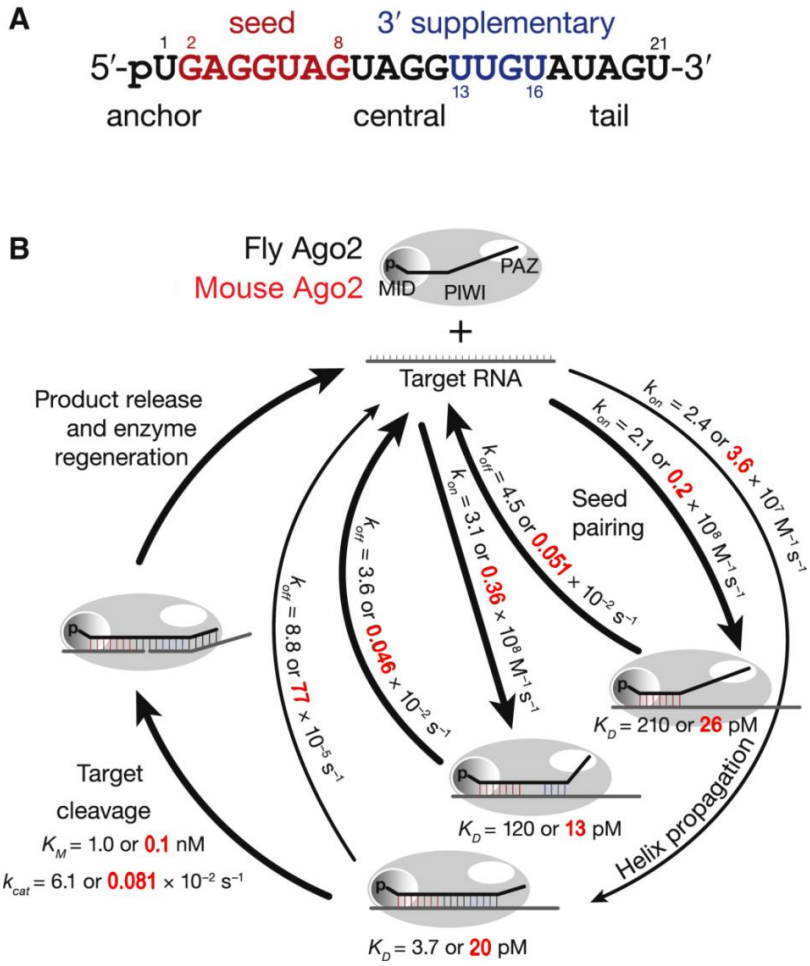


Figure 6.1 Model for RISC function. (A) Argonaute defines distinct functional domains within the RNA guide. A small RNA guide can be divided into anchor (g1), seed (g2–g8), central (g9–g12), 3′-supplementary (g13–g17), and tail (g18–g21) region. (B) A model for RISC binding and cleavage of target RNA. The numbers for fly Ago2 are in black and the numbers for mouse Ago2 are in red. K_D and k_{off} were measured by an equilibrium competition assay (adding competitor RNA to pre-incubated Ago2-RISC). Experimentally determined K_D and k_{off} were used to calculate k_{on} (= k_{off} /K_D). K_M = (k_{off} + k_{cat})/k_{on}. When k_{cat} << k_{off}, K_M~K_D. Modified from Wee *et al.* (115), with kind permission of Elsevier.

6.3 Structural comparison between hAgo 1 and 2

Ago proteins are found in all domains of life and share remarkable structural conservation. The overall nucleic acid binding mechanism is the same regardless of DNA or RNA substrates (145). However, recently interesting differences have been discovered as more biochemical and structural information is obtained from different Agos. For example, some Ago proteins strictly select their guide substrates by the 5´-nucleotide (146) whereas others have the same cleavage efficiencies regardless of the 5´-nucleotides (65). For the development of therapeutics, there have been studies reported which indicate that chemical modifications of therapeutically administered siRNAs or miRNAs can improve serum stability and enhance efficiency of their delivery (147, 148). So the important question here was, how structurally highly homologous proteins differentiate between various substrates and act in different pathways.

hAgo1 and hAgo2 share 84% amino acid sequence identity and adopt quite similar overall folding topologies, whereas there are no obvious structural differences that could explain differential activity between hAgo1 and hAgo2 from a structural point of view (55). Recent publications focus on activating enzymatically inactive hAgo1 and hAgo3 to become slicers. This could be accomplished by two groups (54, 61) using either domain swapping or DNA family shuffling. Besides the difference within the catalytic center, it became obvious that mutations distal to the catalytic center are of major relevance for the cleavage activity. These results revealed a role for noncatalytic structural features for the enzymatic function of human Ago proteins. This is a strong indication that the interaction between Ago and RNA or other accessory proteins, for example TRBP [human immunodeficiency virus (HIV) transactivating response RNA (TAR) binding protein] or PACT (protein activator of the interferon induced protein kinase), could activate a rearrangement of different Ago domains. A crystal structure of a hAgo2/guide/target ternary complex was published recently (43). This structure serves as structural basis for Ago targeting of miRNA. The structures with or without target RNA suggest a stepwise mechanism of miRNA recognition in which the initial target pairing with guide nucleotides 2 to 5 promotes conformational changes that expose more nucleotides for further target recognition. So far, only seed region pairing has been resolved structurally using X-ray crystallography, while the non-seed region, including

mismatches, remains unclear. However, the B-factor analysis showed that seed pairing stabilizes the MID domain and activates the PAZ domain. Moreover, it was reported some years ago (149), that the conformational flexibility and rearrangement of the PAZ domain may be critical for regulation of miRNA binding. The high flexibility of the N-terminus in the ternary complex shows that further target pairing might involve the N-terminal part of Ago and likely triggers further rearrangements.

The domain swapping experiments mentioned above, demonstrated that the N-terminus of hAgo1 is supportive of high affinity miRNA binding, especially the PAZ1 region. The function of the PAZ domain is still poorly understood and mostly known for guide 3´-end binding. Recent research (150) suggest an additional role for the Ago PAZ domain in mammalian miRNA pathways, which shows the PAZ domain is required for Ago loading of slicing-independent RISC. Similar results were found in *Drosophila* (DmAgo) (58). Here, the N-terminal lobe of Argonaute regulates slicer activity, and the interactions with guide RNA determine whether the target mRNA enter a slicer-dependent or independent pathway. Notably, lacking of PAZ domain blocked the silent Ago (hAgo1, 3, 4) to form mature RISC (150), which suggest that the PAZ domain is required for sRNA duplex unwinding or ejection. Since the PAZ domain might be more important for hAgo1 than hAgo2, it was expected to find some differences in the PAZ domain by structural comparison between hAgo1 and hAgo2. The static structure however is very similar and the most obvious difference was found in B-factors. A notable flexible random coil was found in the PAZ domain of hAgo1 which is different from hAgo2, and is exactly the major sequence difference between hAgo2-PAZ1 and hAgo2.

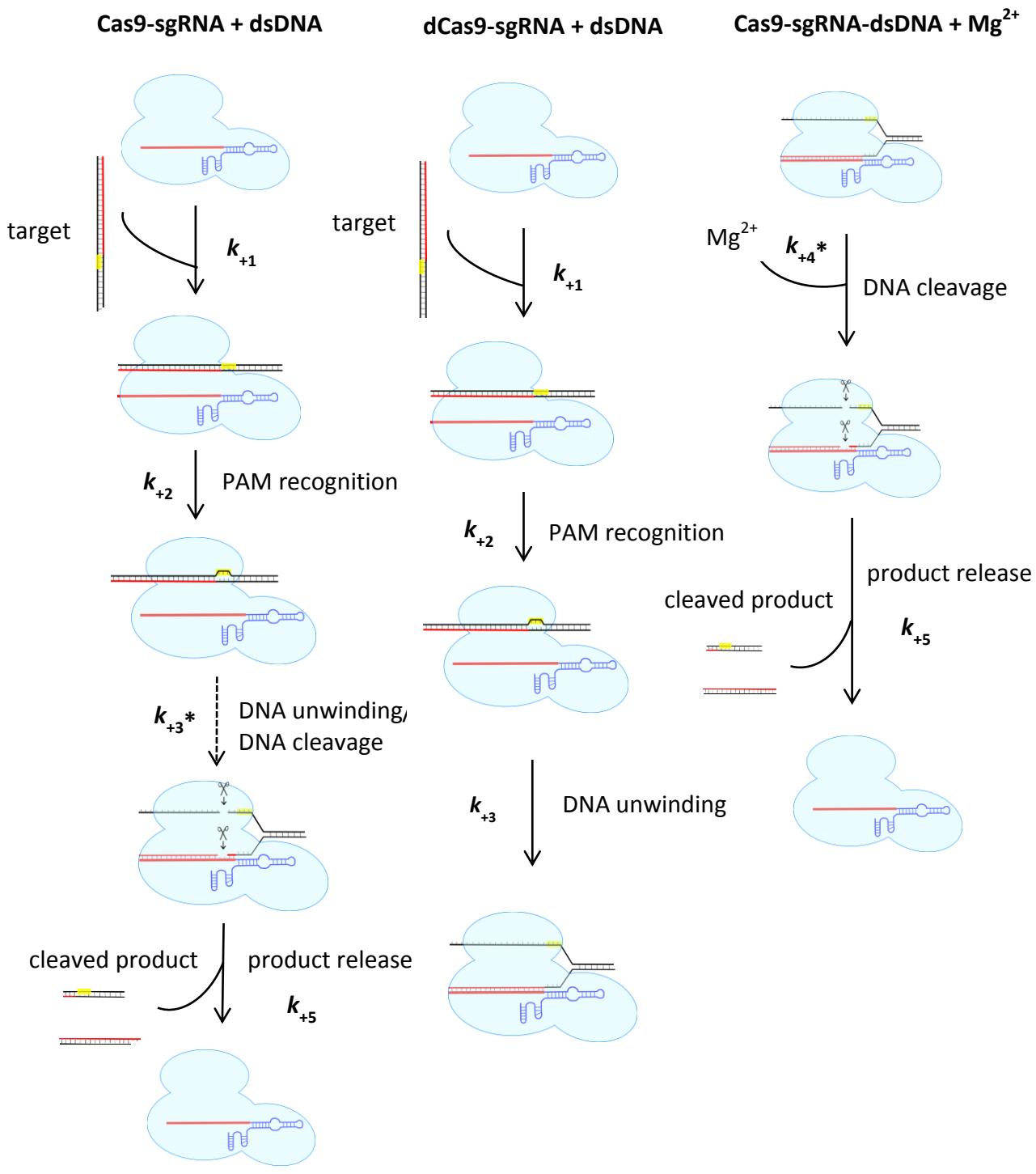
Overall, both hAgo1 and hAgo2 show strong interaction with either siRNA or miRNA, and the binding kinetic differences are less significant for binary than ternary complex formation. It was assumed that they are both adapted to deal with siRNA as well as miRNA, but sorting takes place during ternary complex formation Ago2 shows slower dissociation from fully paired target RNAs whereas hAgo1 shows faster association with mismatched targets. This suggests an adapted

sRNA sorting system in humans, which is more sophisticated and complicated than in the less complex organisms.

6.4 A kinetic model of Cas9 nucleic acid interaction

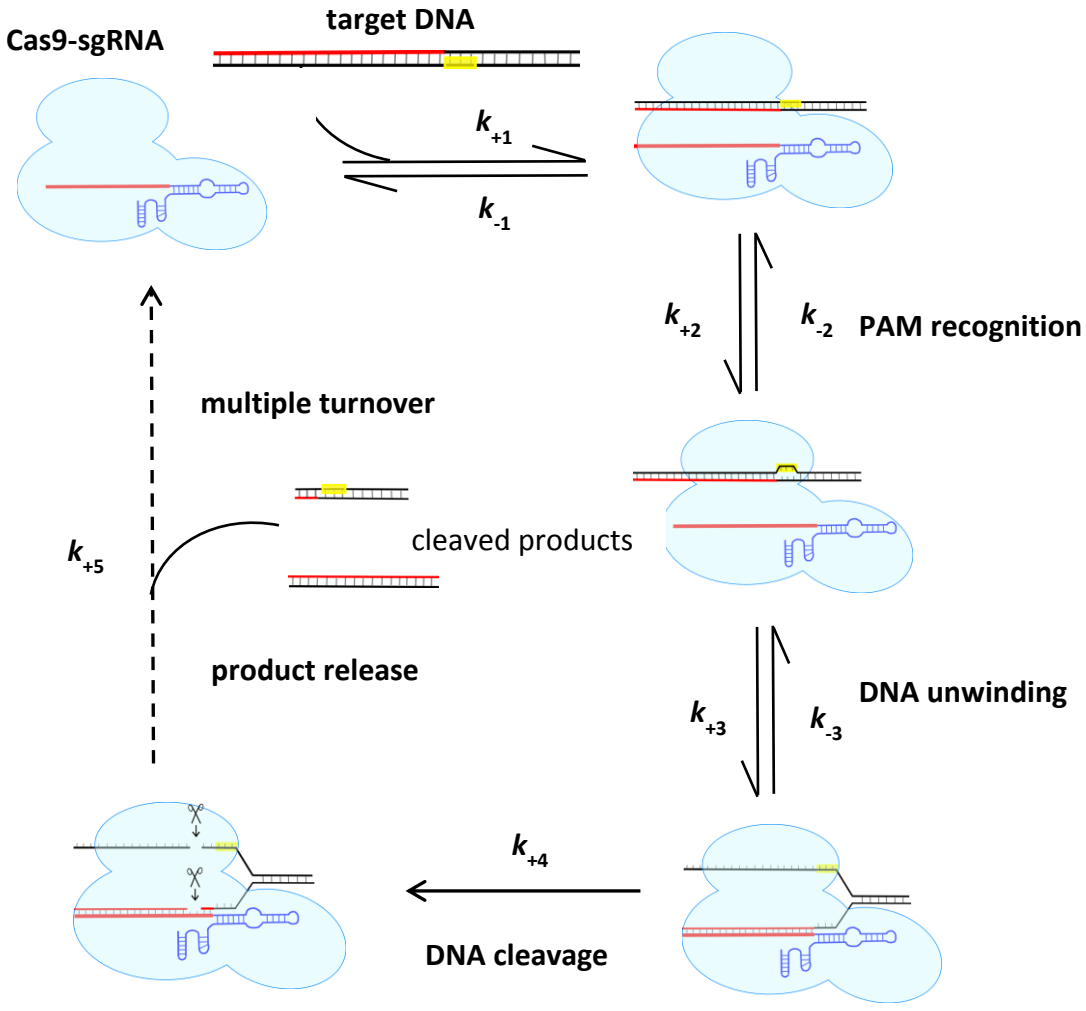
The CRISPR-Cas system is one of the breakthrough technologies in recent years, which fruitfully stimulates related research. In the present study, the kinetics of Cas9 nucleic acid interactions were measured. With the help of structural and biochemical studies by others (110, 112), it was possible to relate the kinetic data to conformational changes along the pathway. Here, three minimal mechanistic models for Cas9 were proposed implementing the different pre-steady-state kinetic measurements (Scheme 6.2).

The Cas9 nucleic acid interaction starts with PAM (protospacer adjacent motif) dependent searching. Once PAM is recognized by a binary complex of Cas9-sgRNA, dsDNA target unwinding and RNA-DNA heteroduplex formation is triggered. Following target base pairing, the dsDNA is cleaved and the product is released. The aim of the CRISPR-Cas9 project in the present study was to establish a comprehensive minimal mechanistic model for Cas9-sgRNA/DNA binding and cleavage. Based on the kinetic data of Cas9 nucleic acid interaction measured applying different conditions, it was possible to propose a Cas9-sgRNA dsDNA binding model (Scheme 6.3).



Scheme 6.2 Analysis of Cas9-nucleic acid interactions by stopped flow measurements applying different experimental conditions. Relating the pre-steady-state data to the binding mechanism five steps k_{+1} to k_{+5} in the order of occurrence were proposed (irrelevant of phase order in individual measurement), consistent with Scheme 6.3. Cas9-sgRNA+dsDNA was related to the stopped flow experiment of Cas9-sgRNA dsDNA complex formation (Figure 5.31B). The pre-steady-state kinetic analysis suggested four phases, with the following transitions: diffusion-controlled formation of a

collision complex (k_{+1}), PAM recognition (k_{+2}), DNA unwinding and cleavage (k_{+3}*) and product release (k_{+5}). dCas9-sgRNA+dsDNA was related to the stopped flow experiment of dCas9-sgRNA dsDNA complex formation (see Figure 5.38A). Pre-steady-state kinetic analysis suggested three phases, with the following transitions: diffusion-controlled formation of a collision complex (k_{+1}), PAM recognition (k_{+2}) and DNA unwinding (k_{+3}). Furthermore, Cas9-sgRNA-dsDNA+Mg^{2+} was related to the stopped flow experiment of Cas9-sgRNA-dsDNA cleavage reaction triggered by Mg^{2+} (Figure 5.33B). Pre-steady-state kinetic analysis suggested two phases with the following transitions: a DNA cleavage associated step (k_{+4}*), which does not represent the actual chemical hydrolysis reaction, and product release (k_{+5}). k_{+3}* differ from k_{+3} because k_{+3}* as it is proposed to include two steps: DNA unwinding and a conformational change due to cleavage. However they could not be differentiated because of the overlay of two phases.



Scheme 6.3 Proposed kinetic model for Cas9-sgRNA dsDNA targeting and cleavage. Cas9-sgRNA binds DNA target by a PAM-dependent search process. After PAM recognition (k_{+2}) the target is loaded to Cas9. Starting with PAM melting, the ds DNA is

sequentially unwound (k_{+3}) followed by RNA-DNA heteroduplex formation. When the two strands are separated and positioned within the corresponding domains, slicing occurs (k_{+4}). The DNA is cleaved, but the cleaved products initially remain bound to Cas9-sgRNA and are subsequently released at a slow rate (k_{+5}). The first fast phase (k_{+1}) depends on the concentration of binding partners, and represents a diffusion-controlled formation of a collision complex. It is proposed that minimally there are five structurally defined steps for the entire reaction. These are collision complex formation (k_{+1}/k_{-1}), PAM recognition (k_{+2}/k_{-2}), DNA unwinding including RNA-DNA hybridization (k_{+3}/k_{-3}), DNA cleavage (k_{+4}*) and product release (k_{+5}). The corresponding rate constants are: k_{+1}: 3.2 × 10^8 $M^{-1}{\cdot}s^{-1}$, k_{-1}: 7.6 s^{-1}, k_{+2}: 3.6 s^{-1}, k_{-2}: 0.1 s^{-1}, k_{+3}: 0.3 s^{-1}, k_{-3}: 0.012 s^{-1}, k_{+4}: (n.d.), and k_{+5}: 0.006 s^{-1}. k_1–k_3 were determined by stopped flow measurements and k_5 by DNA cleavage assay. k_{+4}, the actual chemical hydrolysis step was too fast to be determined accurately.

6.5 From kinetic data to mechanism

In the proposed model, five steps for Cas9-sgRNA dsDNA interaction were defined (Scheme 6.3). Basically, these distinct steps can be divided into target binding and target cleavage. The first three rate constants (k_1–k_3) represent target binding. k_3 can only be observed when using dCas9 (or Cas9 in Mg^{2+} free buffer) and the remaining two steps (k_{+4} and k_{+5}) represent Mg^{2+} dependent target cleavage and product release. However, these five steps could not be observed applying a single stopped flow measurement.

Four phases were observed during binding of non-target^{FAM} dsDNA to Cas9-sgRNA using cleavage conditions with the first fast phase that depends on the concentration of binding partners (Scheme 6.2, Cas9-sgRNA+dsDNA). The two slow phases k_{+3}* and k_{+5} are considered to be associated with slicing and release of cleavage products, because they are very close to the kinetics observed during substrate cleavage under either single or multiple turnover conditions. Comparing these values to binding in the absence of Mg^{2+}, it turned out that the slowest phase k_{+5} is missing, which strongly suggests this phase might be associated with cleavage. However, k_{+2} and k_{+3} were unaffected and close to the corresponding rates in the presence of Mg^{2+}, indicating these steps are Mg^{2+} independent. This is further supported by pre-steady-state analysis of dCas9-sgRNA dsDNA complex formation. The binding kinetics of dCas9 were almost indistinguishable from those

of Cas9 in experiments in the absence of Mg^{2+}. Notably, the third phase (k_{+3}=0.3 s^{-1}) of dCas9-sgRNA dsDNA complex formation (or Cas9-sgRNA dsDNA in the absence of Mg^{2+}) is slightly faster than the third phase (k_{+3}*=0.11 s^{-1}) of Cas9-sgRNA dsDNA complex formation applying cleavage condition. Thus, the phases may represent different steps of reaction. The third phase of dCas9-sgRNA dsDNA complex formation is definitely not related to cleavage because dCas9 is catalytic inactive. For Cas9, the third phase is supposed to represent cleavage, since the rate constant shows the same value as for substrate cleavage under single turnover conditions. Moreover, cleavage was measured in stopped flow experiments. For this a pre-assembled ternary Cas9-sgRNA-dsDNA complex was rapidly mixed with Mg^{2+} to trigger the cleavage reaction. Applying such conditions two phases were observed which are supposed to represent a conformational change associated with cleavage (Scheme 6.2, Cas9-sgRNA-dsDNA+Mg^{2+}). Here, the corresponding rate k_{+4}*=0.4 s^{-1} is faster than the single turnover cleavage rate. As a consequence, it is proposed that the third phase of dCas9-sgRNA+dsDNA complex formation rather represents a step prior to cleavage, most likely RNA-DNA heteroduplex formation (and DNA unwinding), which represents the rate-limiting step of single turnover cleavage. Therefore, the observed third phase of Cas9-sgRNA+dsDNA complex formation might actually include two steps: DNA unwinding and slicing (Scheme 6.2).

According to a study of Sternberg *et al.* (112), the mechanism of Cas9 searching for targets is associated with PAM sequences. Their single-molecule study visualized transient interactions of Cas9 with PAM-rich sequences and showed Cas9-sgRNA finds targets by three-dimensional diffusion, which helps to locate a particular target within the vast amount of DNA in the cell (112). The equilibrium and pre-steady-state studies point in the same direction: if no PAM is present, no binding occurs. The modified PAM target used in this study had a very low binding affinity for Cas9-sgRNA. Stopped flow experiments showed only one phase of Cas9-sgRNA/modified PAM DNA complex formation. This was reasonable as merely collision complex formation could be observed. Thus, PAM recognition is the initial step, followed by R-loop formation. Recent single DNA molecule supercoiling experiments enabled direct observation of R-loop formation in real

time (113, 151). These results indicated R-loops propagate directionally starting from the PAM and expanding toward the distal protospacer end. Moreover, it could be shown that R-loop progression can be monitored close to base-pair resolution and target recognition occurs exclusively through a directional R-loop-zipping process (113). A recent structural study further revealed the interaction between the PAM motif and Cas9 residues in the carboxy-terminal domain (CTD) of the protein (110). The GG dinucleotide of the non-target strand is read out via major-groove interactions with conserved arginine residues from the CTD. This suggests a mechanism for PAM-dependent target DNA melting and RNA-DNA hybrid formation. A few months ago, two distinct DNA binding modes were reported by using single-molecule FRET (152). These results revealed that target recognition initiates from the PAM and promotes with high fidelity via the seed sequence. Although this single-molecule study was performed in the type I CRISPR-Cas systems (using a large ribonucleoprotein surveillance complex called Cascade), it suggests that the stepwise target recognition may also hold true for the type II CRISPR-Cas9 system. So the mechanism of target binding is based on the current understanding. Cas9 finds targets by searching for PAM, followed by PAM melting (R-loop formation), and RNA-DNA hybrid formation (or dsDNA unwinding) mediated by complementary seed sequences. These findings are the basis of a model proposed by Sternberg *et al.* (Figure 6.3) (112). For PAM dependent target searching, the process starts with random three-dimensional collisions of Cas9-sgRNA and DNA. Cas9-sgRNA rapidly dissociates from non-PAM DNA which greatly enhances efficiency. Once bound to a PAM sequence, Cas9 samples the adjacent DNA for guide RNA complementarity. When the target matches a sequence, an RNA-DNA heteroduplex forms, and R-loop expansion propagates via sequential unwinding. If it is the correct target dsDNA will be cleaved and the cleaved products remain bound to Cas9-sgRNA. It was observed that the cleavage products remain tightly bound to Cas9-sgRNA in both single-tethered “DNA curtains” and biochemical gel-shift assays (112). The cleavage assay and stopped flow experiments also showed a slow rate of 0.006 s^{-1}. Cleavage product release is suggested to be the rate-limiting step, which overall limits multiple turnover cleavage processes.

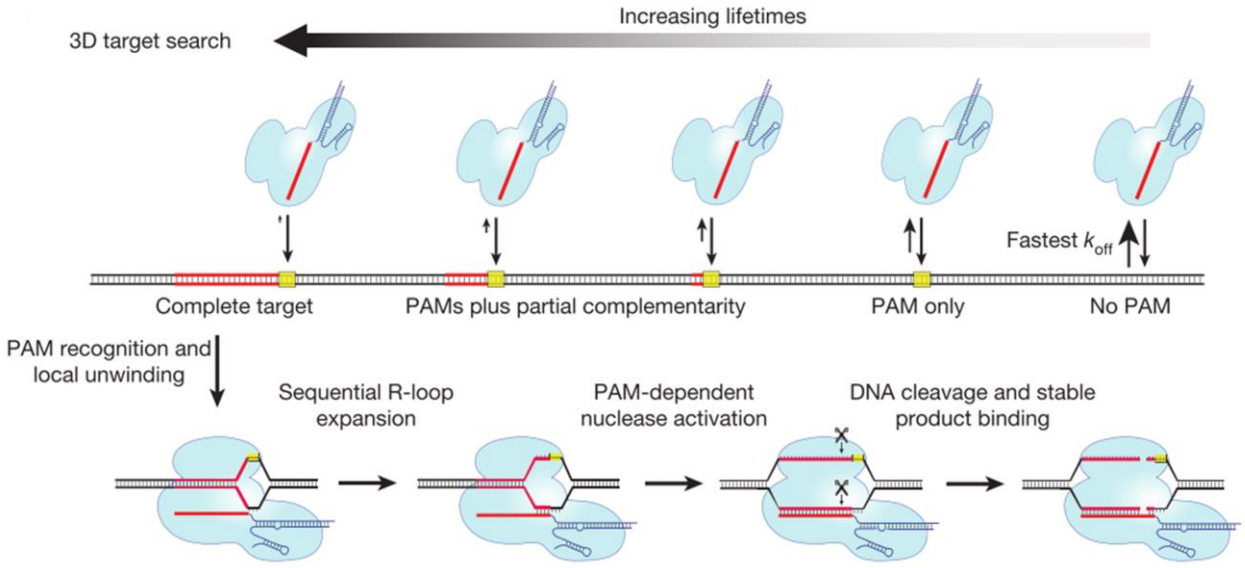


Figure 6.3 Model for target search, recognition and cleavage by Cas9-sgRNA. Details are explained in the text see above. Taken from Sternberg et al. (112), with kind permission of Nature Publishing Group.

6.6 The “seed” mechanism

Recent single-molecule studies of Argonaute provided molecular details for the process of target finding, binding, cleavage, and release (5). The seed (nt position 2–8) is the minimal motif required and promotes stable association of Ago to miRNA targets (136). This emphasizes the importance of the base pairing in seed region for the accuracy of target selection. Together with crystal structure analysis of hAgo2 bound to a miRNA guide and short RNA target sequences (43), this suggests seed pairing is the key mechanism for target binding used by Argonaute proteins during RNA interference in eukaryotes. In this study, a similar seed mechanism for CRISPR-Cas9 was identified and characterized, apparently indicating a possible convergent evolution of Ago and Cas9.

Here, binding affinities of Cas9-sgRNA to DNA strands comprising different seed length from 8nt to 20nt were measured. The determined affinity was strongly seed-dependent with a dramatically reduced affinity when the amount of base pairing in the seed region drops below a certain threshold. This is in accordance with results

from Sternberg *et al.* (112). Higher complementarity results in slower k_{off} rates and longer lifetime of the targets (Figure 6.3). A detailed pre-steady-state kinetic analysis revealed, targets with a 12 or 16 nt seed length showed similar binding properties compared to the 20 nt full-length target, but a target of 8 nt seed length did not. Moreover, the 8 nt seed substrate did show the opposite fluorescence signal change upon interaction with Cas9-sgRNA compared to the other three targets with k_{+3} being 15-fold slower. Together, this indicates that a seed of 8 nts might not be sufficient for proper binding to Cas9-sgRNA complex. According to a recent study (116), a 10-nucleotide RNA "seed" sequence is required for initial DNA interaction which is preordered in an A-form conformation.

Observations from crystal structures of Cas9 bound to sgRNA revealed the Cas9-sgRNA complex is preorganized for target DNA recognition (116). Thus, structural rearrangements might occur before target DNA duplex binding. Single-particle electron microscopy reconstructions showed that the two structural lobes which form the nucleic acid binding cleft perform a guide RNA-induced reorientation to form a central channel in which DNA substrates can be bound (108). sgRNA binding triggers a rearrangement of the PAM recognition region. Moreover, structural observations of the PAM-dependent target DNA recognition was reported by another crystallography study (110). It was revealed that target DNA recognition is based on PAM-dependent searching followed by GG dinucleotide interactions with conserved arginine residues from the Cas9 CTD (110). Comparison of CTD before and after Cas9-sgRNA binding to DNA revealed the CTD domain becomes structured to accommodate the PAM duplex. This is followed by target DNA melting and RNA-DNA hybrid formation. These conformational changes might cause the observed signal change in the fluorescence experiments, when using DNA targets with a FAM label at different positions (Figure 5.31). The 3´-end labeled DNA target^{FAM} showed a smaller signal change and one phase less than the non-target^{FAM} with FAM in the 7th position (calculated from the PAM, or 14th position from 5'-end). Comparing the binding data received with these two different substrates, it turned out that k_{+2} from non-target^{FAM} could not be observed when using target^{FAM}. It is believed the difference arises due to the position of FAM of target^{FAM} being too far away from the protein

center, especially the CTD, so interactions with CTD might not be detected using target^{FAM}. For the same reason, the amplitude of the signal change is smaller than for non-target^{FAM}, which has a FAM-label that is closer to the core of Cas9 protein.

After PAM-dependent DNA target recognition, the PAM region melts followed by RNA-DNA hybrid formation. Since Cas9 has no energy dependent helicase activity, the mechanism of local DNA unwinding remains unknown. A sequential, stepwise unwinding mechanism was proposed by Sternberg *et al.* (112). According to that formation of the RNA-DNA heteroduplex starts at the PAM sequence position and proceeds sequentially towards the distal 5´-terminus of the target sequence, which is consistent with a Brownian ratchet (153). Because the dsDNA unwinding might rely on this thermal energy, a signal change in the opposite direction was observed for the 8 nt seed target with an extremely slow k_{+3} rate constant. A reasonable explanation might be a too short seed region provides too less energy for unwinding, which in turn affects binding and cleaving. This is in agreement with the requirement for a minimal 10 nt long seed sequence.

6.7 Outlook

In the present study, basic mechanistic models for both hAgo1 RNA binding and Cas9-sgRNA DNA targeting and cleavage were established. The detailed kinetic data from steady-state and pre-steady-state analyses in this study provide information about protein nucleic acid complex formation. For Argonaute, I could show that hAgo1 does bind mismatched targets differently compared to hAgo2, highlighting preferential association of hAgo1 with miRNA. This suggest different Argonautes having different functions, and supports the claim of sorting of short RNA is relevant in humans. From a biotechnological point of view, usage of these data could lead to the development of improved miRNA based drugs for biological and clinical applications, especially potential markers for diagnosis (154). For example, the technique here could be used as a platform for testing miRNA targets *in vitro*. Affinity and binding kinetic data could help to predict the therapeutically properties of RNA based drugs.

For Cas9 the model presented here is in support of previous models based on structural and biochemical data (110, 112). It provides a clearer understanding of this new remarkable tool. Applications of Cas9 were explosively expanding in the past two years, but the clinical applications are limited by the potential high amount of off-target effects. Therefore, improving target specificity and accurate predicting potential off-targets remain a significant goal (155). The model system in this thesis is suitable to measure not only the affinity of a given target but also to determine the binding kinetics of each individual reaction step. This will be useful in evaluating binding properties of DNA targets to reduce off-target effects. The final goal is to generate higher fidelity Cas enzymes which can be safely used in gene or cell therapy.

7. References

1. Jonas, S., and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433

2. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*. **391**, 806–811

3. Randall, G., Panis, M., Cooper, J. D., Tellinghuisen, T. L., Sukhodolets, K. E., Pfeffer, S., Landthaler, M., Landgraf, P., Kan, S., Lindenbach, B. D., Chien, M., Weir, D. B., Russo, J. J., Ju, J., Brownstein, M. J., Sheridan, R., Sander, C., Zavolan, M., Tuschl, T., and Rice, C. M. (2007) Cellular cofactors affecting hepatitis C virus infection and replication. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12884–12889

4. Ghildiyal, M., and Zamore, P. D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108

5. Herzog, V. A., and Ameres, S. L. (2015) Approaching the Golden Fleece a Molecule at a Time: Biophysical Insights into Argonaute-Instructed Nucleic Acid Interactions. *Mol. Cell*. **59**, 4–7

6. Jinek, M., and Doudna, J. a (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature*. **457**, 405–412

7. Kobayashi, H., and Tomari, Y. (2015) RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochim. Biophys. Acta - Gene Regul. Mech.* 10.1016/j.bbagrm.2015.08.007

8. Richard W. and Sontheimer, E. J. C. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*. **136**, 642–655

9. Ha, M., and Kim, V. N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524

10. Fabian, M. R., and Sonenberg, N. (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol.*

Biol. **19**, 586–593

11. Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P. D. (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA*. **16**, 43–56

12. Tomari, Y., Du, T., and Zamore, P. D. (2007) Sorting of Drosophila Small Silencing RNAs. *Cell*. **130**, 299–308

13. Czech, B., and Hannon, G. J. (2011) Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.* **12**, 19–31

14. Förstemann, K., Horwich, M. D., Wee, L., Tomari, Y., and Zamore, P. D. (2007) Drosophila microRNAs Are Sorted into Functionally Distinct Argonaute Complexes after Production by Dicer-1. *Cell*. **130**, 287–297

15. Rand, T. a, Ginalski, K., Grishin, N. V, and Wang, X. (2004) Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14385–14389

16. Rand, T. a., Petersen, S., Du, F., and Wang, X. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*. **123**, 621–629

17. Kim, K., Lee, Y. S., and Carthew, R. W. (2007) Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *RNA*. **13**, 22–29

18. Miyoshi, K., Miyoshi, K., Tsukumo, H., Tsukumo, H., Nagami, T., Nagami, T., Siomi, H., Siomi, H., Siomi, M. C., and Siomi, M. C. (2005) Slicer function of. *Genes Dev.* **2**, 2837–2848

19. Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M. C. (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666

20. Leuschner, P. J. F., Ameres, S. L., Kueng, S., and Martinez, J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320

21. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., and Zamore, P. D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*. **123**, 607–620

22. Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008) The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins. *Plant Cell Physiol.* **49**, 493–500

23. Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J., and Qi, Y. (2008) Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. *Cell*. **133**, 116–127

24. Wang, D., Zhang, Z., O'Loughlin, E., Lee, T., Houel, S., O'Carroll, D., Tarakhovsky, A., Ahn, N. G., and Yi, R. (2012) Quantitative functions of argonaute proteins in mammalian development. *Genes Dev.* **26**, 693–704

25. Burroughs, A. M., Ando, Y., de Hoon, M. J. L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C. O. (2011) Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol.* **8**, 158–177

26. Dueck, A., Ziegler, C., Eichner, A., Berezikov, E., and Meister, G. (2012) MicroRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res.* **40**, 9850–9862

27. Yamakawa, N., Okuyama, K., Ogata, J., Kanai, A., Helwak, A., Takamatsu, M., Imadome, K. I., Takakura, K., Chanda, B., Kurosaki, N., Yamamoto, H., Ando, K., Matsui, H., Inaba, T., and Kotani, A. (2014) Novel functional small RNAs are selectively loaded onto mammalian Ago1. *Nucleic Acids Res.* **42**, 5289–5301

28. Juvvuna, P. K., Khandelia, P., Lee, L. M., and Makeyev, E. V. (2012) Argonaute identity defines the length of mature mammalian microRNAs. *Nucleic Acids Res.* **40**, 6808–6820

29. Su, H., Trombly, M. I., Chen, J., and Wang, X. (2009) Essential and overlapping functions for mammalian argonautes in microRNA silencing. *Genes Dev.* **23**, 304–317

30. Meister, G. (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* **14**, 447–59

31. Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Banning,

C. (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.* **17**, 170–180

32. Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V, Patel, D. J., and van der Oost, J. (2014) The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753

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

34. Wei, K. F., Wu, L. J., Chen, J., Chen, Y. F., and Xie, D. X. (2012) Structural evolution and functional diversification analyses of argonaute protein. *J. Cell. Biochem.* **113**, 2576–2585

35. Song, J., Smith, S. K., and Hannon, G. J. (2004) Crystal Structure of Argonaute Slicer Activity. **305**, 1434–1441

36. Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., and Patel, D. J. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell*. **19**, 405–419

37. Parker, J. S., Roe, S. M., and Barford, D. (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737

38. Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D. J. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*. **456**, 921–926

39. Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D. J. (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature*. **456**, 209–213

40. Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T., and Patel, D. J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*. **461**, 754–761

41. Elkayam, E., Kuhn, C. D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., and Joshua-Tor, L. (2012) The structure of human argonaute-2 in complex with miR-20a. *Cell*. **150**, 100–110

42. Schirle, N. T., and MacRae, I. J. (2012) The Crystal Structure of Human Argonaute2. *Science (80-.).* **336**, 1037–1040

43. Schirle, N. T., Sheu-Gruttadauria, J., and MacRae, I. J. (2014) Structural basis for microRNA targeting. *Science (80-.).* **346**, 608–613

44. Zander, A., Holzmeister, P., Klose, D., Tinnefeld, P., and Grohmann, D. (2014) Single-molecule FRET supports the two-state model of Argonaute action. *RNA Biol.* **11**, 45–56

45. Deerberg, A., Willkomm, S., and Restle, T. (2013) Minimal mechanistic model of siRNA-dependent target RNA slicing by recombinant human Argonaute 2 protein. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17850–5

46. Ma, J.-B., Yuan, Y.-R., Meister, G., Pei, Y., Tuschl, T., and Patel, D. J. (2005) Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature*. **434**, 666–670

47. Parker, J. S., Roe, S. M., and Barford, D. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*. **434**, 663–666

48. Frank, F., Sonenberg, N., and Nagar, B. (2010) Structural basis for 5[prime]-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*. **465**, 818–822

49. Künne, T., Swarts, D. C., and Brouns, S. J. J. (2014) Planting the seed: Target recognition of short guide RNAs. *Trends Microbiol.* **22**, 74–83

50. Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D. C., van der Oost, J., Patel, D. J., and Wang, Y. (2014) Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci.* **111**, 652–657

51. Nakanishi, K., Weinberg, D. E., Bartel, D. P., and Patel, D. J. (2012) Structure of yeast Argonaute with guide RNA. *Nature*. **486**, 368–374

52. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*. **15**, 185–197

53. Liu, J., Carmell, M. a, Rivas, F. V, Marsden, C. G., Thomson, J. M., Song, J.-J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*. **305**, 1437–1441

54. Hauptmann, J., Dueck, A., Harlander, S., Pfaff, J., Merkl, R., and Meister, G. (2013) Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nat. Struct. Mol. Biol.* **20**, 814–7

55. Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J., and Joshua-Tor, L. (2013) The Making of a Slicer: Activation of Human Argonaute-1. *Cell Rep.* **3**, 1901–1909

56. Kwak, P. B., and Tomari, Y. (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* **19**, 145–151

57. Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* **11**, 576–577

58. Hur, J. K., Zinchenko, M. K., Djuranovic, S., and Green, R. (2013) Regulation of Argonaute slicer activity by guide RNA 3′ end interactions with the N-terminal lobe. *J. Biol. Chem.* **288**, 7829–7840

59. Petri, S., Dueck, A., Lehmann, G., Putz, N., Rüdel, S., Kremmer, E., and Meister, G. (2011) Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. *RNA*. **17**, 737–749

60. Nakanishi, K., Ascano, M., Gogakos, T., Ishibe-Murakami, S., Serganov, A. a., Briskin, D., Morozov, P., Tuschl, T., and Patel, D. J. (2013) Eukaryote-Specific Insertion Elements Control Human ARGONAUTE Slicer Activity. *Cell Rep.* **3**, 1893–1900

61. Schürmann, N., Trabuco, L. G., Bender, C., Russell, R. B., and Grimm, D. (2013) Molecular dissection of human Argonaute proteins by DNA shuffling. *Nat. Struct. Mol. Biol.* **20**, 818–26

62. Nakanishi, K., Ascano, M., Gogakos, T., Ishibe-Murakami, S., Serganov, A. a., Briskin, D., Morozov, P., Tuschl, T., and Patel, D. J. (2013) Eukaryote-Specific Insertion Elements Control Human ARGONAUTE Slicer Activity. *Cell Rep.* **3**, 1893–1900

63. Shabalina, S. A., and Koonin, E. V. (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587

64. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D., and Aravin, A. (2013) Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. *Mol. Cell*. **51**, 594–605

65. Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J. J., and van der Oost, J. (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*. **507**, 258–61

66. Swarts, D. C., Koehorst, J. J., Westra, E. R., Schaap, P. J., and van der Oost, J. (2015) Effects of Argonaute on Gene Expression in Thermus thermophilus. *PLoS One*. **10**, e0124880

67. Wiedenheft, B., Sternberg, S. H., and Doudna, J. a. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. **482**, 331–338

68. Westra, E. R., Buckling, A., and Fineran, P. C. (2014) CRISPR-Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.* **12**, 317–26

69. Marraffini, L. a. (2015) CRISPR-Cas immunity in prokaryotes. *Nature*. **526**, 55–61

70. Hsu, P. D., Lander, E. S., and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. **157**, 1262–1278

71. Mali, P., Esvelt, K. M., and Church, G. M. (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods*. **10**, 957–963

72. Doudna, J. A., and Charpentier, E. (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346**, 1258096–1258096

73. Jiang, W., and Marraffini, L. A. (2015) CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annu. Rev. Microbiol.* **69**, 209–228

74. Sánchez-Rivera, F. J., and Jacks, T. (2015) Applications of the CRISPR-Cas9 system in cancer biology. *Nat. Rev. Cancer*. **15**, 387–395

75. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, a (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433

76. Bolotin, A., Quinquis, B., Sorokin, A., and Dusko Ehrlich, S. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. **151**, 2551–2561

77. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182

78. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. a, and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**, 1709–12

79. Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V, and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. **321**, 960–964

80. Marraffini, L. a, and Sontheimer, E. J. (2008) CRISPR Interference Limits Horizontal Targeting DNA. *Science (80-.).* **322**, 1843–1845

81. Andersson, A. F., and Banfield, J. F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*. **320**, 1047–1050

82. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009) RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*. **139**, 945–956

83. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*. **329**, 1355–1358

84. Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L., and White, M. F. (2013)

Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell*. **52**, 124–134

85. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**, 9275–9282

86. Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012) PNAS Plus: Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109**, E2579–E2586

87. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. a., and Charpentier, E. (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–821

88. Barrangou, R., and Marraffini, L. a. (2014) CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell*. **54**, 234–244

89. Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. a, Gregory, P. D., Pabo, C. O., and Rebar, E. J. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–85

90. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-Guided Human Genome Engineering via Cas9_Sup. *Science (80-.).* **339**, 823–6

91. Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., and Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*. **153**, 910–918

92. Makarova, K. S., Aravind, L., Grishin, N. V, Rogozin, I. B., and Koonin, E. V (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **30**, 482–496

93. Makarova, K. S., Grishin, N. V, Shabalina, S. a, Wolf, Y. I., and Koonin, E. V (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional

analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct*. **1**, 7

94. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., and Joung, J. K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–84

95. Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. a., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013) Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*. **154**, 1380–1389

96. Gilbert, L. a., Larson, M. H., Morsut, L., Liu, Z., Brar, G. a., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. a., Lim, W. a., Weissman, J. S., and Qi, L. S. (2013) XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. **154**, 442–451

97. Chen, B., Gilbert, L. a., Cimini, B. a., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*. **155**, 1479–1491

98. Chen, S., Sanjana, N. E., Zhang, F., Sharp, P. a, Chen, S., Sanjana, N. E., Zheng, K., Shalem, O., Lee, K., Shi, X., and Scott, D. a (2015) Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis Resource Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell*. **160**, 1–15

99. Shi, J., Wang, E., Milazzo, J. P., Wang, Z., Kinney, J. B., and Vakoc, C. R. (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 1–10

100. Sternberg, S. H., and Doudna, J. A. (2015) Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell*. **58**, 568–574

101. Fu, Y., Foden, J. a, Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., and Sander, J. D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–6

102. Hsu, P. D., Scott, D. a, Weinstein, J. a, Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. a, Bao, G., and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9

nucleases. *Nat. Biotechnol.* **31**, 827–32

103. Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., and Church, G. M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–8

104. Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. a, and Liu, D. R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–43

105. Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. a, Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., and Joung, J. K. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–76

106. Guilinger, J. P., Thompson, D. B., and Liu, D. R. (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–82

107. Koch, L. (2015) Technology: Multi-tasking CRISPR RNA scaffolds. *Nat. Rev. Genet.* **16**, 70–70

108. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. a (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. **343**, 1247997

109. Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. **156**, 935–949

110. Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. **513**, 569–573

111. Jiang, F., and Doudna, J. a (2015) The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **30**, 100–111

112. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. a (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.

Nature. **507**, 62–7

113. Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., and Seidel, R. (2015) Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep.* **10**, 1534–1543

114. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003) Prediction of Mammalian MicroRNA Targets. *Cell*. **115**, 787–798

115. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., and Zamore, P. D. (2012) Argonaute divides Its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*. **151**, 1055–1067

116. Jiang, F., Zhou, K., Gressel, S., and Doudna, J. A. (2015) A Cas9 – guide RNA complex preorganized for target DNA recognition. **348**, 1477–1482

117. Chorn, G., Zhao, L., Sachs, A. B., Flanagan, W. M., and Lim, L. P. (2010) Persistence of seed-based activity following segmentation of a microRNA guide strand. *RNA*. **16**, 2336–2340

118. Grimson, A., Farh, K. K. H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007) MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol. Cell*. **27**, 91–105

119. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. a (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–9

120. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. a, Eckert, M. R., Vogel, J., and Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. **471**, 602–607

121. Papenfort, K., and Vogel, J. (2009) Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res. Microbiol.* **160**, 278–287

122. Friedman, R. C., Farh, K. K. H., Burge, C. B., and Bartel, D. P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105

123. Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M., and Vogel, J. (2010) Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20435–20440

124. Balbontín, R., Fiorini, F., Figueroa-Bossi, N., Casadesús, J., and Bossi, L. (2010) Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in salmonella enterica. *Mol. Microbiol.* **78**, 380–394

125. Semenova, E., Jore, M. M., Datsenko, K. a, Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J. J., and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10098–10103

126. Datsenko, K. a., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., and Semenova, E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945

127. Maier, L.-K., Lange, S. J., Stoll, B., Haas, K. a, Fischer, S., Fischer, E., Duchardt-Ferner, E., Wöhnert, J., Backofen, R., and Marchfelder, A. (2013) Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B. *RNA Biol.* **10**, 865–74

128. Willkomm, S. (2015) *Funktionelle Charakterisierung von Komponenten des humanen und des Methanocaldococcus jannaschii RNA-induced silencing Komplexes*. Ph.D. thesis, Universität zu Lübeck

129. Parthasarathy, S., and Murthy, M. R. (1997) Analysis of temperature factor distribution in high-resolution protein structures. *Protein Sci.* **6**, 2561–2567

130. Deerberg, A. (2011) *Biochemische Charakterisierung der siRNA-vermittelten Erkennung und Spaltung von target RNA durch humanes Argonaute2 Protein*. Ph.D. thesis, Universität zu Lübeck

131. Mayr, C., Hemann, M. T., and Bartel, D. P. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science*. **315**, 1576–1579

132. Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., and Song, E. (2007) let-7 Regulates Self Renewal and Tumorigenicity of Breast Cancer Cells. *Cell*. **131**, 1109–1123

133. Anders, C., and Jinek, M. (2014) In Vitro Enzymology of Cas9. in *The Use of CRISPR/cas9, ZFNs, TALENs in Generating Site Specific Genome Alterations*, 1st Ed., pp. 1–20, Elsevier Inc., **546**, 1–20

134. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D., and Serebrov, V. (2015) Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell*. **162**, 84–95

135. Jo, M. H., Shin, S., Jung, S.-R., Kim, E., Song, J.-J., and Hohng, S. (2015) Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. *Mol. Cell*. **59**, 117–124

136. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J., and Joo, C. (2015) A Dynamic Search Process Underlies MicroRNA Targeting. *Cell*. **162**, 96–107

137. Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*. **115**, 209–216

138. Okamura, K., Liu, N., and Lai, E. C. (2009) Distinct Mechanisms for MicroRNA Strand Selection by Drosophila Argonautes. *Mol. Cell*. **36**, 431–444

139. Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J., and Brennecke, J. (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature*. **453**, 798–802

140. Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L. W., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science*. **320**, 1077–1081

141. Steiner, F. a, Hoogstrate, S. W., Okihara, K. L., Thijssen, K. L., Ketting, R. F., Plasterk, R. H. a, and Sijen, T. (2007) Structural features of small RNA precursors determine Argonaute loading in Caenorhabditis elegans. *Nat. Struct. Mol. Biol.* **14**, 927–933

142. Jannot, G., Boisvert, M.-E. L., Banville, I. H., and Simard, M. J. (2008) Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in C. elegans. *RNA*. **14**, 829–835

143. Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., and Qi, Y. (2009) Rice MicroRNA effector complexes and targets. *Plant Cell*. **21**, 3421–3435

144. Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q., and Tomari, Y. (2010) ATP-dependent human RISC assembly pathways. *Nat. Struct. Mol. Biol.* **17**, 17–23

145. Willkomm, S., and Restle, T. (2015) Conformational Dynamics of Ago-Mediated Silencing Processes. *Int. J. Mol. Sci.* **16**, 14769–14785

146. Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012) Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* **31**, 3588–3595

147. Bramsen, J. B., Laursen, M. B., Nielsen, A. F., Hansen, T. B., Bus, C., Langkjoør, N., Babu, B. R., Højland, T., Abramov, M., Van Aerschot, A., Odadzic, D., Smicius, R., Haas, J., Andree, C., Barman, J., Wenska, M., Srivastava, P., Zhou, C., Honcharenko, D., Hess, S., Müller, E., Bobkov, G. V., Mikhailov, S. N., Fava, E., Meyer, T. F., Chattopadhyaya, J., Zerial, M., Engels, J. W., Herdewijn, P., Wengel, J., and Kjems, J. (2009) A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res.* **37**, 2867–2881

148. Allerson, C. R., Sioufi, N., Jarres, R., Prakash, T. P., Naik, N., Berdeja, A., Wanders, L., Griffey, R. H., Swayze, E. E., and Bhat, B. (2005) Fully 2‘-Modified Oligonucleotide Duplexes with Improved in Vitro Potency and Stability Compared to Unmodified Small Interfering RNA. *J. Med. Chem.* **48**, 901–904

149. Rashid, U. J., Paterok, D., Koglin, A., Gohlke, H., Piehler, J., and Chen, J. C. H. (2007) Structure of Aquifex aeolicus argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. *J. Biol. Chem.* **282**, 13824–13832

150. Gu, S., Jin, L., Huang, Y., Zhang, F., and Kay, M. a. (2012) Slicing-independent RISC activation requires the argonaute PAZ domain. *Curr. Biol.* **22**, 1536–1542

151. Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014) Direct observation of R-

loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9798–803

152. Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S. J. J., and Joo, C. (2015) Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol. Cell*. **58**, 60–70

153. Abbondanzieri, E. a, Greenleaf, W. J., Shaevitz, J. W., Landick, R., and Block, S. M. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature*. **438**, 460–465

154. Soifer, H. S., Rossi, J. J., and Saetrom, P. (2007) MicroRNAs in disease and potential therapeutic applications. *Mol. Ther.* **15**, 2070–2079

155. Singh, R., Kuscu, C., Quinlan, A., Qi, Y., and Adli, M. (2015) Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res.* 10.1093/nar/gkv575

7. Abbreviations

A_{260}	absorbance at 260nm
A_{280}	absorbance at 280nm
aa	amino acid
Ago	Argonaute
APS	ammonium persulfate
AtAgo	*Arabidopsis thaliana* Argonaute
ATP	adenosine triphosphate
bp(s)	base pair(s)
BSA	bovine serum albumin
BPB	bromophenol blue
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR-associated systems
Cas9-sgRNA	binary complex of Cas9 and sgRNA
Cas9-sgRNA-dsDNA	ternary complex of Cas9, sgRNA and dsDNA
C. elegans	*Caenorhabditis elegans*
CTD	carboxy-terminal domain
C-terminus	carboxy-terminus
CTP	cytidine 5'-triphosphate
Ci	Curie
cm	centimeter
crRNAs	CRISPR RNAs
dCas9	nuclease-deficient Cas9
D. melanogaster	*Drosophila melanogaster*
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNAi	DNA interference
dNTP	deoxynucleoside triphosphate
ds	double-strand
DTT	dithiothreitol
eAgo	eukaryotic Argonaute
E. coli	*Escherichia coli*

ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
et al.	*et alia*: and others
EtBr	ethidium bromide
FAM	6-Carboxyfluorescein
FPLC	fast protein liquid chromatography
Fwd	forward
G	glycine
h	hour
GST	Glutathion-S-Transferase
GTP	guanosine-5'-triphosphate
hAgo(s)	human Argonaute protein(s)
k	rate constant
K_d	dissociation constant
LB	Luria-Bertani
M	Mol l^{-1}
min	minutes
miRNA(s)	microRNA(s)
mRNA	messenger RNA
N-terminus	amino-terminus
nt	nucleotide(s)
NTP	triphosphate
OD_{600}	optical density at 600nm
OH	hydroxyl
PAA	Polyacrylamide
PAM	protospacer adjacent motif
PAGE	polyacrylamide gel electrophoresis
pAgo	prokaryotic Argonaute
PAZ	PIWI Argonaute Zwille
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
pH	*potentium hydrogenii*
PIWI	P-element induced wimpy testis

PVDF	polyvinyl difluoride
R	arginine
Rev	reverse
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
rpm	rotations per minute
RT	room temperature
s	second
SDS	sodium dodecylsulphate
sgRNA	single guide RNA
siRNA(s)	small interfering RNA(s)
SpyCas9	*Streptococcus pyogenes* Cas9
sRNA	small RNA
ss	single-stranded
T	thymidine
TAE	Tris-acetate-EDTA
Taq	*Thermophilus aquaticus*
TBE	tris-borate-EDTA
TBS	tris-buffered saline buffer
TCA	trichloroacetic acid
TEMED	N,N,N,N,-tetramethylethylendiamin
TtAgo	*Thermus thermophilus* Argonaute
tracrRNA	transactivating crRNA
Tris	Tris(hydroxymethyl)aminomethane
u	unit, enzymatic activity
UTP	uridine-5'-triphosphate
UV	ultraviolet
V	Volt
v/v	volume/volume
Wt	wild type
w/v	weight/volume
XC	Xylene cyanol

8. Acknowledgements

At the end of my thesis I would like to thank all those people who helped me in the past three years. It will be an unforgettable experience for me.

First, I would like to express my deepest sense of gratitude to my advisor Prof. Dr. Tobias Restle for the continuous support of my Ph.D study. I could not have finished this thesis without his patience, motivation, and immense knowledge; right up to the correction of grammatical errors. I thank him for all his help and great effort he put into training me in the scientific field.

Besides my advisor, I would like to thank Prof. Dr. Georg Sczakiel, Dr. Sarah Willkomm, Dr. Sonja Petkovic and Felix Niesar for their insightful comments and encouragement, which inspired me to widen my research regarding various perspectives. In particular, I am grateful to Dr. Sonja Petkovic for the great work on the correction of this thesis and Felix Niesar for helping me since day 1 I came to Lübeck. My sincere thanks also go to other colleagues at the Institut für Molekulare Medizin, Petra Höltig, Michael Schütt, Dr. Alessandra Mescalchin, Dr. Rosel Kretschmer KazemiFar and Kirsten Frank. Without their precious support it would not have been possible to conduct this research.

I acknowledge my sincere gratitude to the China Scholarship Council (CSC) and the Universitätsklinikum Schleswig-Holstein (UKSH), for the financial support of my PhD project.

I am also thankful to my landlord, the family Carriere, who offered me more than a place to stay, but precious memories and friendship.

Last but not the least, I would like to thank my family and friends, especially my girlfriend. They did support me spiritually throughout my PhD study in Germany and my life in general.