Aus dem Institut für Molekulare Medizin

der Universität zu Lübeck

Direktor: Prof. Dr. Georg Sczakiel

Insights into the functioning of Argonaute proteins with the aid of Molecular Dynamics simulations

Dissertation

zur

Erlangung der Doktorwürde

der Universität zu Lübeck

Aus der Sektion Naturwissenschaften

vorgelegt von

Munishikha Kalia

aus Shimla, India

Lübeck, 2014

First referee: Prof. Dr. Tobias Restle

Second referee: Prof. Dr. Christian G. Hübner

Date of oral examination: 15, October 2014

Table of Contents

GLOSSARY

3’-UTR	3’-UnTranslated Region
A	Adenine
Ago	Argonaute
Agos	Argonuates
Ago2	Argonaute 2
AaAgo	*Aquifex aeolicus* Argonaute
AMP	Adenosine monophosphate
A. thaliana	*Aribdopsis thaliana*
β2AR	β2-Androgenic Receptor
BPTI	Bovine Pancreatic Trypsin Inhibitor
C	Cytosine
CASP	Critical Assessment of Protein Structure-Prediction
CED	C-terminal Effector Domain
C. elegans	*Caenorhabditis elegans*
Cl	Chlorine
CHS	Chalcone Synthase
CMP	Cytosine monophosphate
CN	Compute Node
CPU	Computer Processor Unit

D_3	1, 25-dihydroxyvitamin D_3
D	Aspartic acid
DGCR8	DiGeorge syndrome critical region gene 8
D. melanogaster	*Drosophila melanogaster*
DNA	Deoxyribonucleic acid
DOPE	Discrete Optimized Protein Energy
ds	Double stranded
E	Glutamic acid
E_{kin}	Kinetic energy
EM	Energy minimization
et al.	And others
F	Phenylalanine
FAM	5/6-Carboxyfluoresceine
FRET	Fluorescence Resonance Energy Transfer
fs	Femtosecond
G	Guanine
GHz	Giga Hertz
GMP	Guanosine monophosphate
GPCR	G-Protein Coupled Receptors
GW repeat	Glycine Tryptophan repeat
H	Histidine

hAgo1	Human Argonaute 1
hAgo2	Human Argonaute 2
HIV	Human Immunodeficiency Virus
HLRN	Norddeutscher Verbund zur Förderung des Hoch- und Höchstleistungsrechnens
HMM	Hidden Markov Models
HPC	High Performance Computer
I	Isoleucine
IMM	Institute of Molecular Medicine
IN	Infiniband Network
K	Lysine
k	Rate constant
K_d	Dissociation constant
kDa	Kilo Dalton
KpAgo	*Kluyveromyces polysporus* Argonaute
M	Mole per Liter
$M^{-1}s^{-1}$	Mole per Liter per Second
MD	Molecular Dynamics
Mid	Middle
Mg	Magnesium
miRNA	Micro RNA

MPI	Message Passing Interface
mRNA	Messenger RNA
ms	Millisecond
N	Asparagine
Na	Sodium
N. crassa	*Neurospora crassa*
NED	N-terminal Effector Domain
nm	Nanometer
nM	Nano Mole per Liter
NMR	Nuclear Magnetic Resonance
NPT	Number of particles, Pressure, Temperature
NS loop	Nucleotide specificity loop
nt	Nucleotide
NVT	Number of particles, Volume, Temperature
PABP	Poly-A Binding Protein
PAGE	Polyacrylamide gel electrophoresis
PAZ	PIWI Argonaute Zwille
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PCA	Principal component analysis
PfAgo	*Pyrococcus furiosus* Argonaute

PIWI	*P-element induced Wimpy testis*
PME	Particle Mesh Ewald
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
ps	Picosecond
PTGS	Post Transcriptional Gene Silencing
R	Arginine
RISC	RNA-Induced Silencing Complex
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
RRM	RNA recognition motif
SARA	Stichting Academisch Rekencentrum Amsterdam; Amsterdam Foundation for Academic Computing
siRNA	Small interfering RNA
SOL	Solvent
ss	Single stranded
TIP3P	Transferable Intermolecular Potential 3P
TGS	Transcription Gene Silencing

TRBP	Trans-activation Response RNA-binding Protein
TtAgo	*Thermus thermophilus* Argonaute
Q	Glutamine
U	Uracil
UMP	Guanosine monophosphate
vdW	van der Waal
VMD	Visual Molecular Dynamics
Å	Angstrom
μM	Micromole per Liter

SUMMARY

Argonaute proteins are indispensable for the process of RNA interference (RNAi). In humans, the human Argonaute 2 (hAgo2) is the catalytic engine of RNAi as it possesses a catalytic slicer/endonucleolytic function. hAgo2 has a bilobal structure composed of four domains; N (N-terminal), PAZ (PIWI Argonaute and Zwille), Mid (Middle) and PIWI (P-element induced wimpy testes) inter-connected by the L1 and L2 linker domains. hAgo2 interacts closely with small RNAs of 21-25 nucleotides in length. The so-called guide strand derived from this small double stranded RNAs is tethered on both ends to hAgo2.

The 5'-end of the guide RNA binds the Mid domain through multiple interactions. It has been suggested that these interactions are important for proper positioning of the guide RNA in the protein nucleic acid binding channel. It has been observed that hAgo2 prefers a guide RNA with U or A at the 5'-position over C or G. However, the basis of this discrimination is not well understood. To shed light onto this subject extensive long timescale molecular dynamics (MD) simulations of hAgo2-guide RNA complexes with different 5'-bases were performed. It was observed that especially the presence of a G at the 5'-end considerably increases the flexibility of the Mid domain. In addition, the 5'-G induces a conformational change in the Mid domain which leads to the formation of a novel interaction between 5'-G and helix 7 of the L2 linker domain. This observation is rather interesting since such a 5'-G-L2 interaction has not been reported prior to this study. In summary, contrary to published data, hAgo2 is capable of functionally accommodating guide RNAs with all four possible nucleotides at the 5'-end.

The 3'-end of the guide RNA attaches to the PAZ domain of hAgo2. Recent kinetic data of our group suggest that dissociation of the guide RNA from the PAZ domain is obligatory for the formation of an active ternary complex, which is capable of cleaving the target messenger RNA (mRNA). Therefore, PAZ domain flexibility is deemed important for the catalytic function of hAgo2. To further explore the importance of these conformational changes lengthy simulations of hAgo2-guide RNA complexes were performed. It was observed that the PAZ domain is highly flexible. In addition to this, a novel interaction between the nucleotide specificity (NS) loop of the Mid domain and helix 7 of the L2 linker domain was identified; i.e. salt bridge between K525 and D358. *In vitro* studies with a recombinant D358A-hAgo2 mutant, performed by a colleague,

showed that this single point mutation abolishes the cleavage activity of the enzyme by affecting the guide-hAgo2 interaction. Further MD simulations suggest that the D358A mutation increases the flexibility of the PAZ domain manifold and thus causing changes in the relative positioning of the guide RNA in the hAgo2 nucleic acid binding channel.

The course of the guide RNA inside the nucleic acid binding channel is interrupted by protruding sidechains of neighbouring protein residues at two positions. A first major destacking or kinking occurs between nucleotides 6 and 7. This kink is caused by the I365 sidechain present at the bottom of helix 7 of the L2 linker domain, which also hosts the D358 residue at its pinnacle. It was observed during MD simulations of a hAgo2-guide RNA complex that this kink is preserved during the entire length of the simulations performed. Interestingly, a I365A mutation considerably increases the flexibility of nucleotides 6 and 7, thereby abolishing the kink while at the same time the flexibility of the guide RNA increased substantially. *In vitro* studies with a recombinant I365A-hAgo2 mutant showed that two of three phases of hAgo2-guide RNA binary complex association as well as dissociation are slowed down. Moreover, it was observed that the I365A mutation decreases the cleavage efficiency of hAgo2 by a yet unknown mechanism.

Another member of the human Argonaute family is hAgo1, which as well closely interacts with small RNAs while its role in RNAi is less well understood. Although, hAgo1 is strikingly similar to hAgo2 structurally, it does not possess a endoneucleolytic slicer activity. To obtain insights into potential differences concerning the modus operandi of hAgo1 versus hAgo2 series of MD simulations were performed. It was observed that hAgo1 is as flexible as hAgo2 and again the PAZ domain is the most flexible part of the protein. In addition, an interaction between the NS loop and helix7 of the L2 linker domain was identified. However, particular domain motions are reversed in comparison to hAgo2 which could give a first hint towards functional differences.

To investigate whether the unique L2-Mid interaction observed in human Argonautes during this thesis might be an evolutionarily conserved regulatory mechanism, series of MD simulations were performed in all Argonautes for which a full-length X-ray structure was available. It was observed that such a L2-Mid interaction occurs in all Argonautes investigated with the exception of the prokaryotic *Thermus thermophilus* Argonaute (TtAgo). Hence, this L2-Mid interaction seems to be a universal feature in Argonaute proteins and it is concluded that this inter-domain interaction is a regulatory mechanism conserved throughout evolution of Argonaute proteins.

ZUSAMMENFASSUNG

Argonaute Proteine sind unabdingbar für die RNA Interferenz (RNAi). Im Menschen ist das humane Argonaute 2 (hAgo2), welches eine endonukleolytische *slicer* Funktion besitzt, die katalytische Komponente der RNAi. hAgo2 besitzt eine zweigeteilte Struktur, die aus vier Domänen besteht; der N-terminalen-, der PAZ- (PIWI Argonaute *and* Zwille), der Mid- (*middle*) und der PIWI- (*P-element induced wimpy testis*) Domäne, welche durch die L1 und L2 Linker-Domänen verbunden sind. hAgo2 interagiert mit kurzen RNAs von 21- 25 nt Länge. Aus diesen kurzen doppelsträngigen RNAs entstehen die sogenannten einzelsträngigen *guide* RNAs, welche mit beiden Enden an hAgo2 gebunden sind.

Das 5‘-Ende der *guide* RNA bindet die Mid-Domäne über vielfältige Interaktionen. Es wurde postuliert, dass diese Interaktionen wichtig für die korrekte Positionierung der *guide* RNA im Nukleinsäurebindungskanal des Proteins sind. Beobachtungen haben gezeigt, dass hAgo2 eine *guide* RNA mit entweder einem U oder einem A gegenüber einer *guide* RNA mit einem C oder G an der 5‘-Position präferiert. Dennoch ist der zu Grunde liegende Mechanismus der Diskriminierung zwischen verschiedenen *guide* RNAs noch nicht komplett verstanden. Um diesen Aspekt näher zu beleuchten, wurden umfangreiche Moleküldynamik (MD) Simulationen über lange Zeiträume mit hAgo2-*guide* RNA Komplexen mit unterschiedlichen 5‘-Basen durchgeführt. Es konnte gezeigt werden, dass insbesondere die Anwesenheit eines G am 5‘-Ende zu einer beträchtlichen Erhöhung der Flexibilität der Mid-Domäne führt. Weiterhin induziert das 5‘-G eine konformationelle Änderung in der Mid-Domäne, welche zur Ausbildung einer neuen Interaktion zwischen dem 5‘-G selbst und der Helix 7 der L2 Linker-Domäne führt. Diese Beobachtung ist sehr interessant, da eine solche 5‘-G-L2 Interaktion vor dieser Studie nicht beschrieben wurde. Zusammenfassend kann hAgo2, was im Gegensatz zu bereits publizierten Daten steht, mit *guide* RNAs mit allen vier möglichen Nukleotiden am 5‘-Ende funktionelle hAgo2-*guide* RNA Komplexe bilden.

Das 3‘-Ende der *guide* RNA bindet an die PAZ-Domäne von hAgo2. Neueste kinetische Daten aus unserer Arbeitsgruppe zeigen, dass die Dissoziation der *guide* RNA von der PAZ-Domäne eine notwendige Voraussetzung für die Bildung aktiver ternärer Komplexe, welche in der Lage sind *target* RNAs zu spalten, darstellt. Daher wird die Flexibilität der PAZ-Domäne als

unabdingbar für die katalytische Funktion von hAgo2 erachtet. Um die Bedeutung dieser konformationellen Änderungen weiter zu untersuchen, wurden längere Simulationen von hAgo2-*guide* RNA Komplexen durchgeführt. Es konnte beobachtet werden, dass die PAZ-Domäne sehr flexibel ist. Zusätzlich konnte eine neue Interaktion zwischen dem Nukleotid-Spezifitäts *loop* (NS *loop*) der Mid-Domäne und der Helix 7 der L2 Linker-Domäne identifiziert werden; das heißt eine Salzbrücke zwischen K525 und D358. *In vitro* Studien mit einer rekombinanten D358A-hAgo2 Mutante, durchgeführt von einer Kollegin, zeigten, dass diese einzelne Punktmutation die Spaltungsaktivität des Enzyms aufhebt, indem die *guide* RNA-hAgo2 Interaktionen verändert werden. Weitere MD Simulationen deuten darauf hin, dass die D358A Mutation die Flexibilität der PAZ-Domäne um ein Vielfaches erhöht, was zu Veränderungen in der relativen Positionierung der *guide* RNA im hAgo2 Nukleinsäurebindungskanal führt.

Der Verlauf der *guide* RNA innerhalb des Nukleinsäurebindungskanals wird durch die hervorstehenden Seitenketten der benachbarten Protein Reste an zwei Positionen unterbrochen. Dies ist zum Einen ein Abknicken der Nukleinsäure zwischen Nukleotid 6 und 7. Dieser Knick wird durch die I365 Seitenkette, welche sich am unteren Ende der Helix 7 der L2 Linker-Domäne befindet, verursacht. An der Spitze derselben Helix befindet sich auch die D358 Seitenkette. Während der MD Simulationen des hAgo2-*guide* RNA Komplexes konnte beobachtet werden, dass dieser Knick während der gesamten Länge der Simulation aufrecht erhalten wird. Interessanterweise führt eine I365A Mutation zu einer beträchtlich erhöhten Flexibilität der Nukleotide 6 und 7, wodurch der Knick entfernt wird, was wiederum die Flexibilität der *guide* RNA substantiell erhöht. *In vitro* Studien mit einer rekombinanten I365A-hAgo2 Mutante, die wiederum von einer Kollegin durchgeführt wurden, zeigen, dass zwei der drei Phasen der hAgo2-*guide* RNA Komplex-Assoziation und –Dissoziation verlangsamt sind. Weiterhin konnte beobachtet werden, dass die Spaltungseffizienz von I365A-hAgo2 durch einen bislang unbekannten Mechanismus verringert ist.

Ein weiteres Mitglied der humanen Argonaute Familie ist hAgo1, welches ebenfalls mit kurzen RNAs interagiert, dessen Rolle in der RNA Interferenz jedoch schlechter verstanden ist. Obwohl hAgo1 eine auffällige strukturelle Ähnlichkeit zu hAgo2 aufweist, besitzt es keine endonukleolytische *slicer* Aktivität. Um Aufschluss über die potentiellen Unterschiede des *modus operandi* von hAgo1 gegenüber hAgo2 zu erlangen, wurden mehrere MD Simulationen

durchgeführt. Es konnte beobachtet werden, dass hAgo1 genauso flexibel ist wie hAgo2 und auch bei hAgo1 die PAZ-Domäne den flexibelsten Teil des Proteins darstellt. Zusätzlich wurde eine Interaktion des NS *loops* mit der Helix 7 der L2 Linker-Domäne identifiziert. Allerdings sind bestimmte Domänenbewegungen im Vergleich zu hAgo2 gegensätzlich, was erste Hinweise auf die funktionellen Unterschiede zwischen hAgo2 und hAgo1 geben könnte.

Um zu untersuchen, ob die einzigartige L2-Mid Interaktion, die in humanem Argonaute im Zuge dieser Arbeit gefunden wurde, möglicherweise einen evolutionär konservierten regulatorischen Mechanismus darstellt, wurden weitere MD Simulationen mit allen Argonaute Proteinen, für die eine komplette Röntgenkristall-Struktur vorhanden ist, durchgeführt. Es konnte beobachtet werden, dass so eine L2-Mid Interaktion in allen untersuchten Argonaute Proteinen, mit Ausnahme des prokaryotischen *Thermus thermophilus* Argonaute (TtAgo), stattfindet. Somit scheint die L2-Mid Interaktion eine universelle Eigenschaft in Argonaute Proteinen zu sein und es kann der Schluss gezogen werden, dass diese Inter-Domänen Wechselwirkung einen regulatorischen Mechanismus, der im Rahmen der Evolution der Argonaute Proteine konserviert ist, darstellt.

Chapter 1. INTRODUCTION

1.1. RNA interference (RNAi)

RNAi is a conserved mechanism, which regulates gene expression on the post transcriptional level. Andrew Fire and Craig Mello originally discovered the RNA interference pathway in the nematode *Caenorhabditis elegans* (*C. elegans*) (1). The discovery of RNAi resulted in a Nobel Prize in Medicine or Physiology in 2006. RNAi has been historically known as 'co-suppression' (2), 'quelling' (3) and 'post transcriptional gene silencing (PTGS)' (4). Most importantly, RNAi further highlighted the regulatory role of double stranded small non-coding RNAs that control the expression of genetic information.

The first double stranded RNA (dsRNA) with a regulatory role was reported in *C. elegans* (5). This dsRNA is complementary to a section of its target mRNA within the 3'- untranslated region, thereby inhibiting translation. Since then the regulatory role of small RNAs in plants and animals has been associated with diverse expression patterns and has been verified in some cases by genetic studies in model organisms and humans (6-8). In multicellular organisms, there are three major types of small RNA: microRNAs (miRNA), small interfering RNAs (siRNA) and piwi interfering RNAs (piRNA). These small non-coding RNAs are unique as they are double stranded and usually 19-22 nt in length.

Although Fire and Mello presented the first conclusive model for RNAi, Napoli and Jorgesen first reported RNAi like phenomena in 1990, during their studies on chalcone synthase (CHS), a key enzyme in flavonoid biosynthesis (2). Their goal was to determine the role of CHS as a rate-limiting enzyme in anthocyanin biosynthesis, which is responsible for the deep violet coloration in the petunias. In order to attain deep blue coloration they over expressed CHS, surprisingly they obtained white colored petunias. They noticed the level of introduced CHS was 50-fold lower than in wild type (wt) petunias. This led them to propose that the introduced transgene caused 'co-suppression' of the endogenous CHS gene. A similar phenomena was observed in *Neurospora crassa* (*N. crassa*) by Romano and Macino, as they observed 'quelling' of the endogenous gene on introduction of a homologous RNA sequence (3).

Guo and Kemphues were the first to document RNAi in animals. They observed in *C. elegans* that the introduction of an RNA complementary to par-1 mRNA blocked par-1 expression (9). This technique of introducing large amounts of nucleic acids with a sequence complementary to the target mRNA into the cytoplasm of a cell is known as 'antisense-mediated silencing' (10). It was believed that the base pairing would occur between the 'sense' mRNA sequence and the 'antisense' complementary nucleic acid. This would eliminate gene expression by either passively blocking mRNA translation or by destruction of the mRNA by cellular ribonucleases (11). However, surprisingly Guo and Kemphues observed that both antisense and control sense RNA induced silencing (9). Since control sense RNA and target mRNA are identical to each other it is impossible for them to pair, posing the question how the control sense RNA could cause silencing?

In 1998 Fire and Mello explained RNAi, they illustrated that gene silencing in *C. elegans* was caused by a dsRNA (1). In addition, they observed that a dsRNA was considerably more efficient in causing interference than 'sense' or 'antisense' strands alone. Furthermore, they observed that the effect of the interference was evident not only in the worms injected with dsRNA but also in their progeny. They also reported that just a small amount of dsRNA was required per affected cell, suggesting that RNAi could have a catalytic or amplification component.

The overall mechanism of RNAi is now fairly well established. It is a multi step process which involves the formation of an RNA Induced Silencing Complex (RISC), first reported in *Drosophila* by Hannon and coworkers (12, 13). They demonstrated that RISC is a multiprotein nucleic acid complex with ~500 kilodaltons (kDa) in size. Further biochemical studies reported that a protein called Dicer is a component of the RISC (14, 15). It was also established that another protein named 'Argonaute' (Ago) is a component of RISC (16). It was later reported that trans-activation response RNA-binding protein (TRBP) is also a part of the RISC (17). It has also been demonstrated that a catalytically functional 'minimal RISC' which can accurately cleave substrate RNAs is composed of recombinant human Ago 2 (hAgo2) and a single stranded siRNAs (18). More recently, it could be shown that recombinant hAgo2 can be functionally loaded with a double stranded siRNA (19).

1.2. MicroRNAs (miRNAs)

Amboros and coworkers discovered the first miRNA 'lin-4' in *C. elegans*. This lin-4 miRNA is essential for proper timing during *C. elegans* larval development (5). Since then the presence of miRNAs was shown to be ubiquitous across plants and animal species. In humans approximately 5% of the entire genome is dedicated to encoding and producing >1,800 miRNAs according to the miRNA database miRbase (http://www.mirbase.org/), which may regulate up to 30% of human genes (20, 21). miRNAs can target messenger RNAs (mRNA) through imperfect binding; therefore, they can target a wide range of mRNAs potentially regulating a large fraction of the human genome (22-24). miRNAs control several vital processes such as cell growth, tissue differentiation, heterochromatin formation, and cell proliferation. Aberrant miRNA expression has been associated with several diseases in humans such as cardiovascular diseases, neurological disorders and different types of cancer (25).

1.2.1. miRNA Biogenesis

A large number of miRNA genes are present in intergenic regions or in antisense orientation to annotated genes suggesting that they might form independent transcription units (26-29). In animals, miRNA biogenesis embarks in the nucleus. RNA Polymerase II (Pol II) generates primary miRNA (pri-miRNA) transcripts, these pri-miRNAs have 5'-methylated caps and 3'-polyadenylated tails typical to most products of Pol II (30-34). These pri-miRNA transcripts can be several thousand bases in length, for example the full length of pri-miR-21 RNA is ~3433-nt (33, 35). The pri-miRNAs form stem-loop structures, which bear the characteristic bulges and mismatches within the folded molecule. These bulges and mismatched nucleotides have been perceived to play an important role in the miRNA biogenesis (36, 37).

The process of miRNA biogenesis can be broadly classified into three major steps. In the first step, the long pri-miRNA transcripts are excised into hairpin shaped precursor-miRNAs (pre-miRNAs) by a protein complex called Microprocessor illustrated in the Figure 1-1 of this thesis (38, 39). The Microprocessor protein complex is composed of Drosha an RNase III enzyme and the double stranded RNA-binding protein DiGeorge syndrome critical region gene 8 (DGCR8) in vertebrates (40) or Pasha in invertebrates. The deletion of DGCR8 gene causes the DiGeorge syndrome (41, 42). The pre-miRNAs have distinct stem and loop structures and are 60 - 90 nt long (26-28, 43, 44).

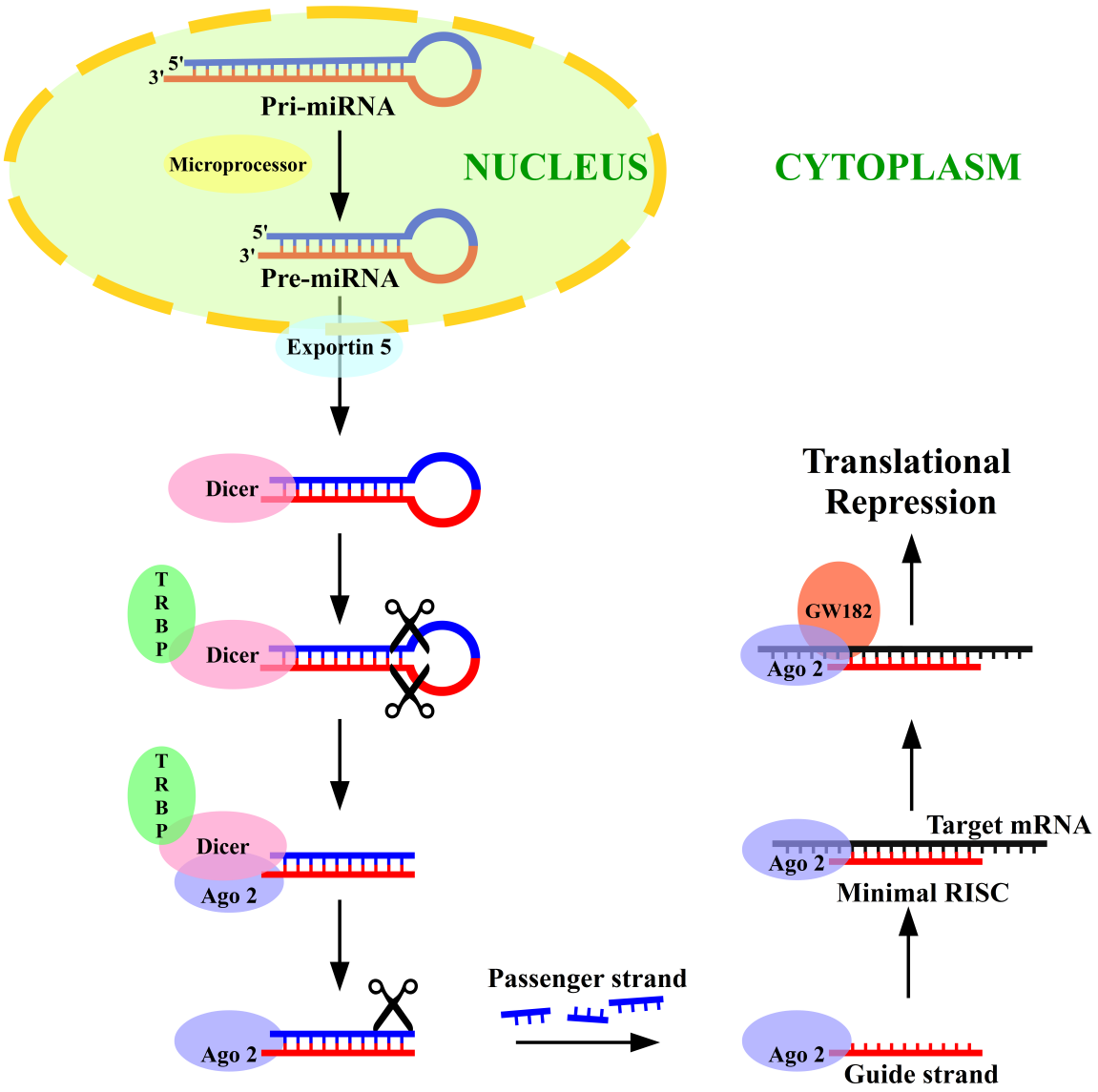

Figure 1-1: A simplified representation of miRNA biogenesis and the miRNA mediated RNAi pathway. The pri-miRNA is cleaved to form pre-miRNA, which is transferred to the cytoplasm with the aid of Exportin 5. In the cytoplasm, Dicer cleaves the pre-miRNA and forms mature miRNA. The guide strand is loaded to Ago2 and eventually translational repression occurs.

In the second step of miRNA biogenesis, the pre-miRNAs are then exported to the cytoplasm by a dsRNA binding protein called Exportin 5 (Figure 1-1) (45, 46). In the final step of miRNA biogenesis, Dicer a large multidomain RNase III endoribonuclease with the aid of TRBP cleaves the precursor miRNAs into mature miRNAs (14, 43, 47, 48). Dicer produces miRNAs that are 19-22 nt in length with characteristic 5'-phosphate and 2 nt overhang at the 3'-end (Figure 1-1) (49, 50).

1.2.2. miRNA mediated RNAi pathway

The mature miRNAs produced by Dicer are double stranded structures, however only one strand of the miRNA is incorporated into the RISC. Therefore, the two strands of mature miRNA must unwind first. The two miRNA strands are thermodynamically dissimilar to each other (37, 51, 52). The thermodynamically less stable 5'-end strand is loaded into the Ago and is called the 'guide strand'. Whilst the other strand known as the 'passenger strand' is believed to be rapidly discarded (Figure 1-1) (52). However, there are examples where the passenger strand is loaded onto the Ago and direct repression of target mRNAs (53). There is evidence for two distinct miRNA mediated silencing mechanisms; mRNA cleavage and translational repression, also known as slicer dependent and slicer-independent silencing (54-59). The extent and nature of the complementarity between the guide and target mRNAs determines the gene silencing mechanism (60-62). The animal miRNAs contain bulges and mismatches and are not fully complementary to their corresponding target mRNAs. However, they are fully complementary to the target mRNAs in the 'seed region', which occurs close to the 5'-end of the miRNA, usually between the nucleotides at position 2 to 8 (22).

In the slicer dependent silencing the Ago cleaves the target mRNA. This occurs when there is an extensive base pairing between the guide and target mRNA over the seed region and the nucleotides at position 10 and 11 of the guide RNA (62-66). Nevertheless, it has been observed that certain additional requirements might be necessary for the slicer-dependent cleavage (67, 68). In the slicer-independent silencing mechanism the slicer activity of Ago is prevented by the bulges formed due to imperfect base pairing between the guide and target mRNAs (5, 60, 65). However, there is no concrete model so far for the inhibition of translation through miRNAs. It has been observed in the experiments performed on mammalian cells that the miRNA induced inhibition of translation can occur both at initiation and elongation steps of the translation (69, 70). Recent data suggests that the promoter of the transcribed target mRNA determines the mechanism employed for the translational repression (71).

It has also been demonstrated that miRNA mediated translational repression requires the GW182 scaffold protein. Three functional regions have been mapped in *Drosophila melanogaster (D. melanogaster)* and mammalian GW182 proteins, which cause mRNA translation repression to a similar extent (72). The three domains are; N effector domain (NED) which has multiple glycine

and tryptophan (GW) repeats, the middle Glutamine rich domain called the Q-rich domain and the C-terminal effector domain (CED). The CED domain contains the cytoplasmic polyA-binding protein (PABP)-interacting motif 2 (PAM2) and the RNA recognition motif (RRM). The NED of GW182 interacts closely with Ago, whilst the CED interacts with PABP and adenylases (73-75). The interactions between GW182 and the PABC are required for the miRNA target adenylation and degradation (75). The mechanism by which the GW182 domains cause translational repression seems to be evolutionary conserved, as GW182 can repress mRNA function in mammalian cells and human TNRC6 proteins can act as repressors in *D. melanogaster* cells (72, 76-78).

1.3. Small interfering RNAs (siRNAs)

siRNAs were initially perceived to be primarily exogenous in origin, introduced experimentally as dsRNA or through viral infections (49, 79). However, later several endogenous siRNAs were identified in yeasts, plants, *C. elegans* and mammals (80-83). The exogenous siRNAs are commercially available and can be synthesized by solid-phase. These siRNAs can be ordered in the form of pre-formed duplexes or as ssRNA oligonucleotides that can be annealed. The synthetic siRNAs can be delivered into cells and even tissues with the aid of carriers, such as the biodegradable nanoparticles and lipids (84).

1.3.1. siRNA biogenesis

The endogenous siRNAs originate from transposon transcripts, sense-antisense transcript pairs and long stem-loop structures (85, 86). The biogenesis of the endogenous siRNAs is determined by the activity of RNA-dependent RNA polymerase (RdRP) which catalyzes the replication of RNA from an RNA template (87-89). The precursors of endogenous siRNAs are mostly produced from sense-antisense derived from transposons. They might also originate from convergent transcription of protein-coding genes and from unannotated parts of the genome (85). Since these precursors are not always transcribed from the same loci, they are likely to have bulges and mismatches. Another kind of endogenous siRNA precursors are the single-stranded, self-hybridizing, which form an elongated stem-loop structure, however they are disparate from the miRNA precursors due to the increased length of the stems (86, 90).

The precursors of exogenous siRNAs are long, linear, perfectly base paired dsRNAs, which can be introduced into the cytoplasm artificially (91). Dicer cleaves the siRNA precursors into double stranded siRNA duplexes (19 – 23 nt) with a 2 nt 3'-overhang (Figure 1-2) (92, 93). It was initially believed that the siRNA processing in animals mostly occurs in the cytoplasm (30). This view has been recently confronted by studies on *D. melanogater,* which found that Dicer 2 is predominately found in nucleus (94). It is however confirmed that siRNA processing in *C. elegans* occurs in the cytoplasm (95).

1.3.2. siRNA mediated RNAi pathway

The most significant difference between miRNAs and siRNAs is their complementarity with their corresponding target mRNAs. miRNAs are rarely 100 % complementary to the target mRNAs outside the seed region, whilst siRNA are fully complementary to their target mRNA (37, 52, 93). It is fairly established now, that the degree and nature of the complementarity between the guide and target mRNAs determines the gene silencing mechanism (60-62).

It has been observed that a single stranded siRNA can be directly loaded onto recombinant hAgo2 (18), however a siRNA duplex generated by Dicer cannot and requires the siRISC assembly pathway. This view was recently disproved by Deerberg *et al.* (19) who showed functional loading of ds siRNA to recombinant hAgo2. The pathways have been reasonably well characterized in vivo in humans and *D. melanogaster*. The siRISC pathway in *D. melanogaster* is mediated by the R2D2/Dicer 2 heterodimer, which binds an siRNA duplex and then leads to the formation of RISC loading complex (RLC) (93).

In humans, the RLC is formed of Dicer, TRBP and hAgo2. However, it has been suggested that Dicer is not essential for RISC loading in mammals, by studies performed on mouse cells in which siRNA loading occurs in the presence of a null allele of Dicer (96, 97). Similar to miRNAs, one strand of the siRNA duplex binds the Ago known as the 'guide strand'. The other strand called 'passenger strand' is degraded (Figure 1-2). The strand selection process in siRNAs is also similar to miRNAs.

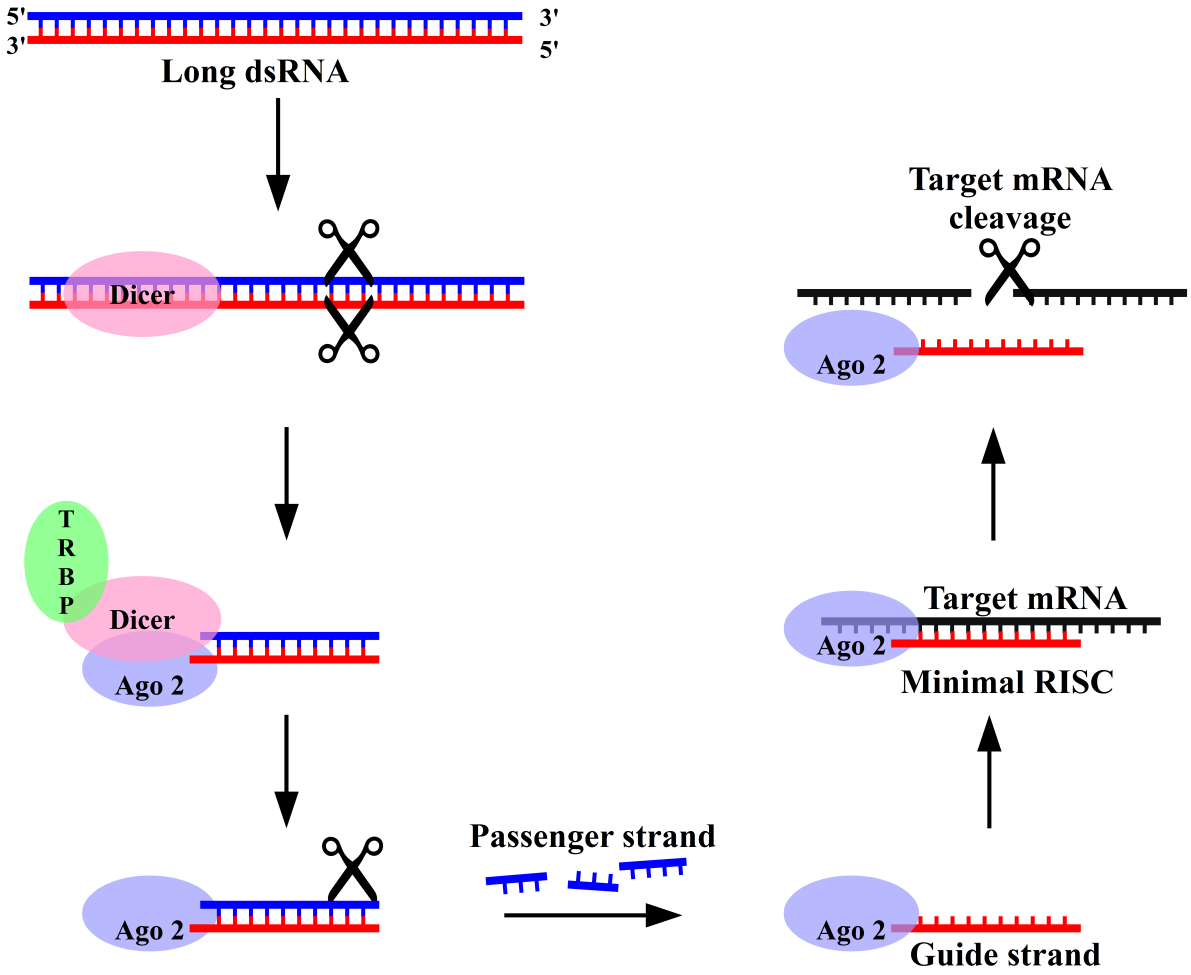

Figure 1-2: The schematic representation of a simplified siRNA mediated RNAi pathway. The long ds siRNA is cleaved to form mature siRNA duplex. The mature siRNA duplex is loaded to Ago2 with the help of TRBP and Dicer. The guide strand bound to Ago2 targets the mRNA to be cleaved. Further, the target mRNA is cleaved by Ago2.

In vivo and *in vitro* experiments have explained strand selection occurs based on relative thermodynamic stability of the duplex ends. The strand which has the less thermodynamically stable 5'-end is preferred as the guide strand (93), it binds Ago and forms a binary complex. The target mRNA then binds this binary complex, and the whole setup is then called the 'ternary complex'. Ago then cleaves the target mRNAs at a single phosphodiester bond linking nucleotides at positions 10 and 11(Figure 1-2) (49, 64).

1.4. Argonaute proteins

Argonaute proteins are a highly conserved family that are involved in RNAi across all three kingdoms of life (25). The Ago1 mutant in *Arabdopsis thaliana* (*A. thaliana*), caused different phenotypic abnormalities, such as radicalized leaves and infertile flowers that resembled a squid (*Argonaute argo*), hence the name Argonaute (98). The functional role of prokaryotic Argonautes (Agos) has remained elusive for a long time. However, recent studies on *Thermus thermophilus* (*T. thermophilus*) bacteria have demonstrated that TtAgo acts as a barrier against uptake and propagation of foreign DNA (99-101).

In comparison to the prokaryotic counterparts, the functional role of eukaryotic Agos is well established. The eukaryotic Ago family is composed of the Ago and the PIWI proteins. The Ago proteins are also known as the Ago clade and are similar to *A. thaliana* Ago1. The PIWI proteins are homologous to the *D. melanogaster* and are referred to as the PIWI clade. The Ago proteins mostly interact with miRNAs and siRNAs causing post transcriptional gene silencing, whereas the PIWI proteins mainly bind the piRNAs (102).

The mammalian Ago family consists of Ago1, Ago2, Ago3 and Ago4. They are all expressed ubiquitously and are known to associate with small RNAs (103). The characteristic feature of the Agos is the presence of PAZ (PIWI-Argonaute-Zwille) and PIWI (P-element induced wimpy testes) domains. Hannon and coworkers demonstrated that Ago2 is necessary for mouse development and cells deficient in Ago2 are unable to load guide RNAs. They also observed that mutations within the PIWI domain inactivates RISC (64). Hence, it was established that Ago2 is the only member of the mammalian Ago family, which possesses a 'slicer activity' (102, 104).

The first full-length crystal structure of an Ago protein (PfAgo) illustrated the overall structural organization of Agos (105). Most importantly the PfAgo structure revealed that the tertiary structure of the PIWI domain in PfAgo resembled the RNase H enzymes (105). After the initial PfAgo structures, full-length structures of archae *Aquifex aeolicus* Ago (AaAgo) were reported (106). A comparison of AaAgo structures pointed towards a rather large flexibility of the PAZ domain. In addition a short MD simulation of the AaAgo structure was performed and it was hypothesized that the PAZ domain flexibility could be a potential regulator of the RISC function (106, 107). Later, a range of TtAgo structures were reported in different arrangement of guide

and target strands (108-110). These TtAgo structures provided crucial insights into mode of guide and target strand binding. It was observed in the TtAgo binary complex that the guide strand is tethered on its both ends. The 5'-end of the guide strand is attached to the Middle (Mid) domain and the 3'-end binds the PAZ domain. The structures of the TtAgo ternary complex provided 'high-resolution snapshots' of the important steps involved in the Ago mediated mRNA cleavage. It demonstrated the relative positioning of the catalytic residues and also revealed the presence of Mg^{2+} cations (99). Two TtAgo structures with different lengths of target strand demonstrated that the PAZ domain undergoes conformational changes during a catalytic cycle (109).

Further insights into the structural architecture of the eukaryotic Agos was provided by the recently reported crystal structures of the full-length human Ago 1 (hAgo1), human Ago 2 (hAgo2) and yeast *Kluyveromyces polysporus* Ago (KpAgo) proteins (111-115). Surprisingly, despite of low sequence identity (~12%) between hAgo2 and TtAgo, the overall architecture of the Ago is highly conserved. This is one of the finest examples of domain conservation over sequence conservation (112). The details of the hAgo2 structure are explained in Chapter 1.5.

1.5. Human Argonaute 2 protein

Ago 2 is known to be the catalytic engine of RNAi in humans as it embodies an endonucleolytic "slicer" activity (116, 117). It has been demonstrated by gene inactivation experiments in mouse that the Ago2 is critical for embryonic development, stem cell maintenance and cell differentiation (118, 119). Additional studies in mouse models have indicated that the Ago2 is a key regulator of B lymphoid and erythroid development and function (120). Moreover recent studies in human myeloid cell lines and primary blast cells have illustrated that hAgo2 is crucial for human monocytic cell fate determination and LPS-induced inflammatory response of 1,25-dihydroxyvitamin D_3 (D_3)-treated myeloid cells. hAgo2 depletion changes the accurate modulation of transcription factors and miRNAs involved in monocyte cell fate determination, causing a decrease of D_3-induced monocyte differentiation and activation (121). Further studies revealed that hAgo2 plays a critical role in maintaining miRNA homeostasis in human T cells. It was observed that hAgo2-deficient T cells have decreased miRNA levels and higher probability of differentiating into cytokine-producing effectors (122).

The significance of hAgo2 in a diverse range of biological functions has only started to unravel, which makes hAgo2 a very lucrative subject of research. It is however necessary to have more insights into the functioning of hAgo2. A better understanding of the hAgo2 on atomic level has been provided by two recently reported full-length crystal structures of hAgo2 in complex with guide RNAs of varying lengths; PDB codes: 4F3T.pdb, 4OLA.pdb (113, 123).

1.5.1. General structural architecture

The first crystal structure of hAgo2 in complex with a guide RNA was reported by Schirle and MacRae (113). The resolution of the structure was 2.30 Å with a guide RNA bound to the Mid domain with its 5'-end and extending up to 8 nucleotides in length inside the nucleic acid binding channel. Elkayam and coworkers subsequently reported another crystal structure (123). This structure was reported in the presence of miR-20a at a resolution of 2.25 Å (Figure 1-3). The overall structural architecture and relative positioning of individual domains is similar in both cases.

hAgo2 is composed of four domains; N (N-terminal), PAZ (PIWI, Argonaute and Zwille), Mid (Middle) and PIWI (P-element induced wimpy testes), tethered by two linker regions L1 and L2 (123). The structure is bilobed with a central cleft, which forms a nucleic acid binding channel. One lobe consists of the N and PAZ domains and the other is formed by the Mid and PIWI domains (Figure 1-3). The mode of small RNA binding to hAgo2 is comparable to the prokaryotic Agos: both ends of the guide RNA are fixed, the 5'-end by the Mid domain and the 3'-end by the PAZ domain (108-110, 113, 123). The miRNA sits in a nucleic acid binding groove and in its course across the protein, the guide RNA interacts with all four hAgo2 domains together with the two linkers (113, 123).

One of the intriguing features observed in the guide RNA bound to hAgo2 is the presence of prominent destacking or kinks. The first noticeable destacking is observed between nucleotides 6 and 7 of the guide RNA. The protruding I365 sidechain is placed between bases 6 and 7, which causes the kink. I365 is part of the L2 linker domain, connecting the PAZ and Mid domains. A stacking interaction occurs between the I365 sidechain and the base G7. It has been hypothesized that this kink might play a role in the target recognition and could also be important for the release of the cleaved RNA (113). Another major destacking occurs between nucleotides 9 and

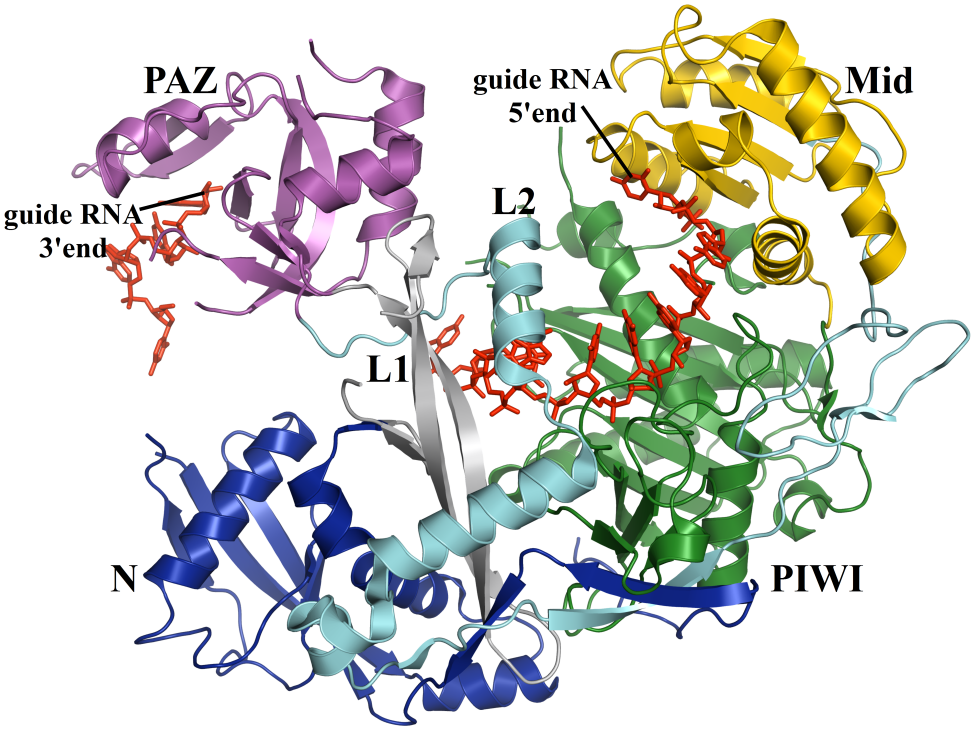

Figure 1-3: Structural organization of the hAgo2 in complex with a guide RNA (4F3T.pdb). Individual domains are color-coded: N (blue), L1 (silver), PAZ (magenta), L2 (pale cyan), Mid (yellow), PIWI (green). The 5'-end of the guide RNA (red sticks) binds Mid domain and 3'-end is attached to the PAZ domain.

10. This destacking in caused by protruding arginine sidechains; R710 stacks against U9 and R635 stacks against U10. Another arginine is deemed to be important, R351 which interacts with the backbone of U9. I353 also contributes to this destacking through van der Waals interactions.

The comparative analysis of eukaryotic and prokaryotic Agos revealed the presence of eleven eukaryotic specific insertions (124). Three of these insertions are located in the nucleic acid binding channel, which causes a lengthening of the nucleic acid binding channel relative to the prokaryotic counterparts. A specific eukaryotic C-terminal helical insertion was also revealed, which carries several phosphorylation sites and is believed to play some role in discriminating between single stranded guide and a guide-target duplex (125). A detailed description of the individual domains of hAgo2 and their interactions with the guide RNA is described in the following subsections of this chapter.

1.5.2. Guide RNA 5'-end & the Mid domain interaction

The 5'-end of the guide RNA forms a multitude of specific interactions with several residues in a tight binding pocket of the Mid domain (112, 126). The full-length structure of hAgo2 also showed that residues from the PIWI domain complete this pocket. The three non-bridging atoms of the 5'-phosphate of the guide RNA interact closely with neighboring protein residues (Figure 1-4). The first phosphate atom of the guide RNA interacts with Y529 and K533 from the Mid domain and R812 from the PIWI domain. The second oxygen interacts with the sidechains of K570, K566, and Q545, whereas the third oxygen interacts with C546 and K570 (Figure 1-4). These residues in turn interact very closely with their neighbouring residues, hence forming a close network of (very) firm interactions. It has been proposed that these firm and extensive interactions between the guide RNA 5'-phosphate and the Mid binding pocket define the position of the guide RNA relative to the active site in the PIWI domain. This ensures that the cleavage of target RNA occurs at a fixed and predictable position. It has also been proposed; in the absence of these strong interactions, the guide RNA might adopt different conformations inside the nucleic acid binding channel, which might affect the fidelity of the complex (18, 123).

Yet another important feature of the Mid binding pocket is a distinct loop known as the nucleotide specificity (NS) loop, which harbors the base of the 5'-nucleotide (Figure 1-4) (126). Sequence analysis performed on various nematode, fly, plant and conserved human miRNAs reveal that there is a strong bias for a U or a A at the 5'-position of the guide strand (126-130). The crystal structures of the hAgo2 Mid domain in the presence of isolated 5'-mononucleotides sheds some light over this bias. The hydrogen-bonding patterns of 5'-C and 5'-G are completely opposite to those of 5'-A and 5'-U, furthermore the amide group present in 5'-G collides with the carbonyl group of G524 present in the NS loop. NMR titration experiments of a hAgo2 Mid domain with isolated 5'-nucleotides further confirm this bias, as 5'-UMP and 5'-AMP have been shown to have 30 fold higher binding affinity in comparison to 5'-CMP and 5'-GMP (126). This bias plays a role in the loading of miRNA into the Ago with a strong preference for a miRNA with 5'-U (131). The specific interactions between the 5'-nucleotide and the Mid binding pocket seem to be important for the RISC fidelity.

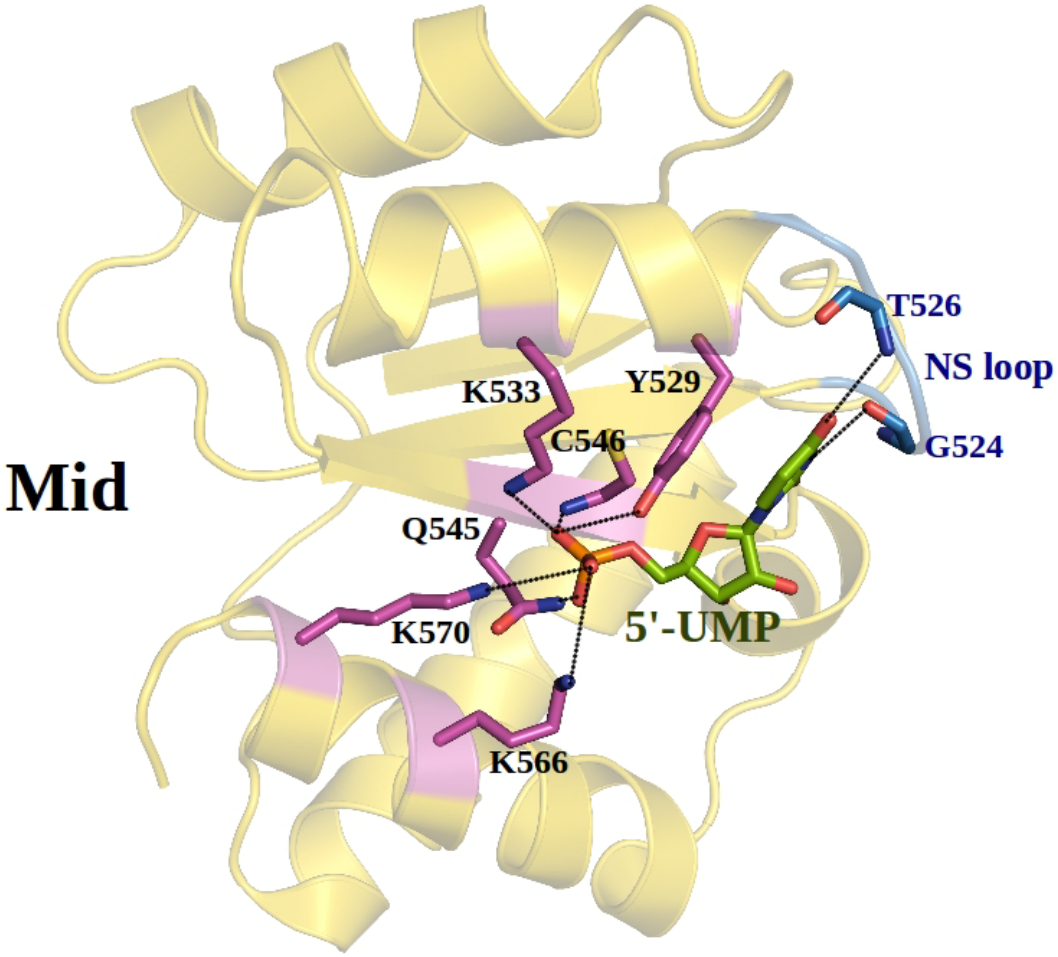

Figure 1-4: Close view of the Mid binding pocket represented in cartoon (yellow). The 5'-UMP (green) and the interacting protein residues (pink) are represented in sticks. The NS loop is highlighted in blue cartoon. Black dotted lines represent hydrogen bond interactions between the 5'-UMP and the residues in the Mid binding pocket.

1.5.3. Guide RNA seed region & PIWI domain interaction

The PIWI domain is the largest domain of hAgo2 and the one, which embodies the catalytic/slicer function. In the hAgo2 structure, the active site lies opposite to nucleotides 9 and 10 (Figure 1-5). It is in contrast to the TtAgo were the active site is situated opposite to nucleotides at position 10 and 11. The organization of the catalytic site in hAgo2 is similar to the enzymes of the RNase H family. The active site of RNase H enzymes has a conserved DEDX motif, which is comparable to the human Ago DEDH catalytic tetrad (111, 132, 133). The

hAgo2 catalytic site is composed of two canonical acidic residues D597 and D669, protruding into the central β sheet of the PIWI domain as illustrated in Figure 1-5. The active site also contains the H807 residue, which is present in a neighboring α helix.

Although the hAgo2 structures reported the D597, D669 and H807 residues in the catalytic site, a glutamate residue, which would complete the catalytic tetrad, was not indicated. A crucial insight into this missing glutamate residue was provided by the yeast KpAgo structure (111). It was observed that E1013 'glutamate finger' is inserted into the catalytic pocket, which is

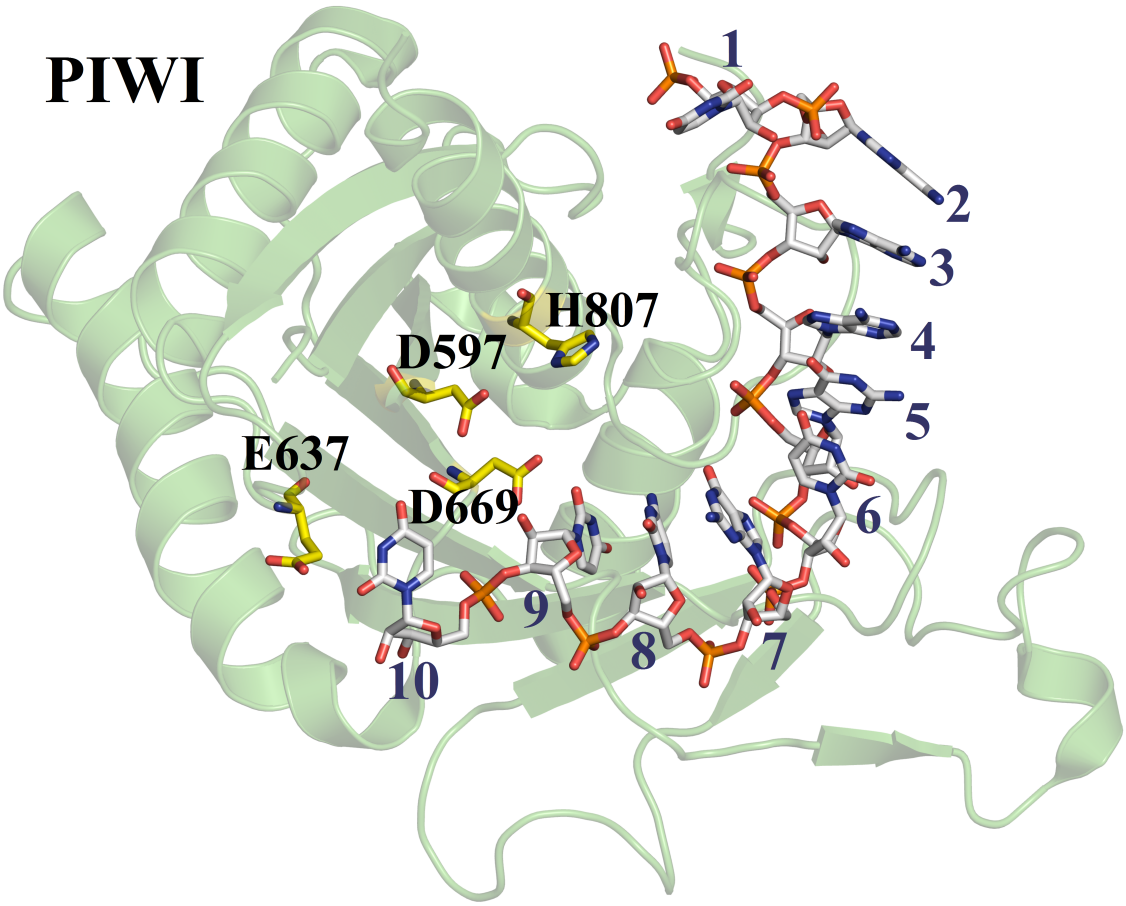

Figure 1-5: Close view of the catalytic tetrad represented by sticks (yellow), embodied in the PIWI domain represented by transparent cartoons (green). The guide RNA is represented by sticks (white and blue). Although the structure of target RNA was not reported in the crystal structure, it is known that the catalytic site of hAgo2 is placed opposite to the scissile phosphate between nucleotides 9 and 10 of the target RNA.

stabilized by an extensive hydrogen bond network.

In hAgo2, the E637 residue was identified as the 'glutamate finger', which completes the DEDH catalytic tetrad. A comparative analysis of Agos revealed that the plugged in conformation of the glutamate finger is observed only when the 3'-end of the guide RNA is not bound to the PAZ domain. It is fairly established now that the release of guide 3'-end from the PAZ domain is mandatory for the target RNA cleavage (19). Therefore, it was proposed that the Ago is in a catalytically 'active state' when the glutamate finger is inserted into the catalytic pocket. However, in the 'inactive state', the glutamate finger is not inserted into the catalytic pocket. The PIWI domain also forms one-half of the narrow nucleic acid binding channel, whilst L1/L2 linker domains and N-domain form the other half. The guide RNA sequence follows a kinky path during its course across the nucleic acid binding channel. It is interesting to note that backbone phosphates within the seed region interact very closely with neighboring protein residues present in the nucleic acid binding channel, whilst the bases of the seed region have very limited interactions.

1.5.4. Guide RNA 3'-end PAZ domain interaction

The 3'-end of the guide RNA binds to the PAZ domain. In the hAgo2 structure, the PAZ domain appears to be more elongated than in the prokaryotic counterparts. It makes the hAgo2 structure most open relative to the other known Ago structures. In the crystal structure reported by Elkayam and coworkers (4F3T.pdb) only the last four nucleotides were observed. The last two nucleotides occur in a cleft of the PAZ domain. The last nucleotide A20 stacks against F294, the backbone phosphate joining the last base also makes multiple contacts. One of the nonbridging oxygens coordinates with H271 and H316. The bridging oxygen interacts with Y311. Out of the four terminal bases, only A18 and G19 appear to be stacked (Figure 1-6).

1.6. Molecular Dynamic simulations

Molecular dynamic (MD) simulations are a computational tool used to study the dynamics of biomolecules. It acts like a computational microscope that captures the behaviour of biomolecules in full atomic detail. Static structures determined through crystallography and other techniques provide a wealth of knowledge about the structural organization of biomolecules such

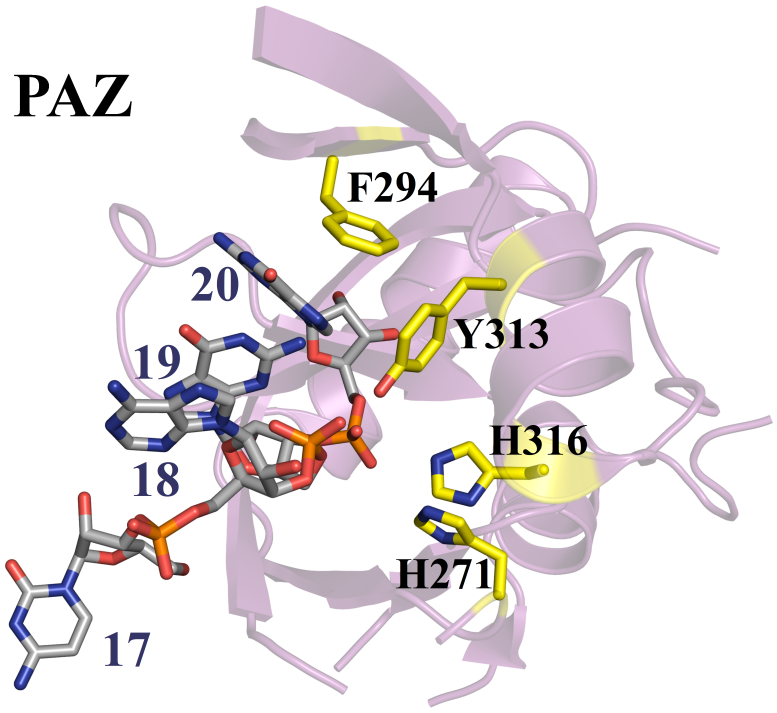

Figure 1-6: Close view of the PAZ domain represented by transparent cartoons (pink). The last four nucleotides of the guide RNA (grey) and the residues interacting with the guide RNA (yellow) are represented by sticks.

as proteins, nucleic acids etc. However, proteins are highly flexible and dynamic, which is crucial for their catalytic function.

Proteins are involved in a diverse range of catalytic functions and mechanisms. They act as signaling molecules, transporters, catalysts, mechanical effectors and sensors. Proteins also interact with drugs, hormones, nucleic acids and one another. These diverse functions require a high degree of agility; therefore, proteins have to undergo conformational changes to fulfill these functions. Static structures are comparable to a photograph, however to get deeper insights kind of a time-resolved microscopy is needed to observe these dynamic motions.

Several experimental techniques such as NMR structures or single molecule FRET analysis offer the opportunity to observe such dynamic motions; however, there are limitations as to the spatial and temporal resolution. Moreover, these techniques usually provide ensemble average

properties, instead of a motion of the individual molecule. Hence, computational tools such as MD simulations represent a promising alternative to provide insights at a atomic level (134).

1.6.1. Background

The first ever MD simulations were performed by Alder and Wainwright in 1957 (135). An assembly of hard spheres was placed in a rectangular box with periodic boundary conditions (PBC). The hard spheres moved at a constant velocity in a straight line unless interrupted by occasional collisions. The collisions were perfectly elastic and the principle of conservation of linear momentum was applied to calculate new velocities of colliding spheres. The first simulation of a liquid using continuous potential was performed by Rahman in 1964 on liquid argon at 94.4K (136). This was soon followed by the MD simulation of water at 307.45K (137).

However, none of these simulations were performed on biomolecules. The first MD simulation of a folded protein was reported by the Karplus group (138). The study was performed on bovine pancreatic trypsin inhibitor (BPTI), a globular protein of 58 amino acids in length. The actual simulation was carried out for just 8.8 picoseconds (ps), which revealed an internal 'fluid-like' motion of BPTI. This study showed the dynamic behaviour of folded proteins.

Since the first MD simulation study of a protein, the field of biomolecular simulation has progressed manifold. Duan and Kollman beautifully illustrated the process of protein folding in the villin headpiece, by performing the first microsecond long simulation (139). They started out with an unfolded state of the villin headpiece subdomain and observed a hydrophobic collapse and helix formation in the initial phase, followed by subsequent conformational changes. Recently, Shaw and coworkers reported a BPTI simulation in the millisecond range (140).

There are several instances where MD simulations have provided deeper insights into the understanding of a biological mechanism. One such example is the G-protein-coupled receptors (GPCRs) transmembrane proteins. The superfamily of GPCR proteins contains at least 800 seven-transmembrane receptors, which are involved in a diverse range of physiological processes. Approximately 36% of the marketed targeted cancer pharmaceuticals interfere with human GPCRs (141). GPCRs recognize different extracellular signals and transfer them to G proteins, which then transduce these signals to intracellular effectors.

The structure of β2-adrenergic receptor (β2AR) a member of class A, rhodopsin like GPCRs, is reported in both active and inactive states (142-144). However, the mechanistic details of the activation and deactivation process are still lacking. Various MD simulation studies have enlightened the process of conformational changes and mechanisms of receptor activation (145-149).

In addition to providing crucial insights into several biological mechanisms, MD studies have also made a crucial contribution to drug development. In 2004 Schames and coworkers performed MD simulations of the HIV-1 integrase protein, a drug target not very susceptible to structure based drug design (150). Through MD simulations a hidden trench was identified which could not be observed in the crystal structure. This so-called hidden trench was termed as a 'cryptic binding site'. It was later shown through X-ray crystallography experiments that the previously known inhibitors of HIV-1 integrase could bind to this cryptic binding site. These results were further validated through experiments by Merck and co., which eventually led to the production of the antiretroviral drug raltegravir (151). This compound was the first drug against HIV-1 integrase approved by the US Food and Drug Administration. Although the field of biomolecular simulations has come a long way from its initial stages, it still faces important challenges such as computational cost and limitations due to imprecise physical models.

1.6.2. MD simulations of Argonaute proteins

Conformational flexibility within hAgo2 plays a crucial role in its slicer/endonucleolytic activity (19, 106, 152, 153). The 3'-end is anchored in the PAZ domain in the binary complex, on target binding the 3'-end of guide RNA dissociates from the PAZ domain (19). The PAZ domain is known to undergo considerable domain motions during the catalytic cycle, as seen in the TtAgo structures with varied substrate lengths (109, 154).

The conformational flexibility has been further highlighted by several MD studies performed on various TtAgo structures (155-157). One of the studies showed that disruptive mutations (G→C) in the seed region create a large bending motion of the PAZ domain across the L1/L2 hinge region which led to an opening of the nucleic acid binding channel (157). These studies have generated interesting insights into the dynamic behaviour of prokaryotic Agos, however a better understanding of the eukaryotic counterparts especially in humans is still lacking.

1.7. Aims of the thesis

hAgo2 is the catalytic lynchpin of the RNAi pathway. A recent study, in the Restle group established minimal mechanistic model of hAgo2, based on pre-steady state kinetic studies, alludes to the fact that certain conformational changes within the protein are crucial for its catalytic function (19). Additionally, single molecule FRET studies of two prokaryotic Ago proteins further point towards the significance of such conformational changes for enzyme mechanism (152, 153). Although, a wealth of knowledge has been provided by these biochemical studies insights into the working of hAgo2 at an atomistic level are still unavailable. Therefore, the computational technique 'MD simulations' has been employed to gain a deeper understanding of the conformational changes and potentially observe novel inter-atomic interactions at a microscopic level. The overall aim of this thesis was to provide insights and to identify novel residues crucial for the functioning of Ago proteins.

In the Chapter 3.1, a homology based 3D model of hAgo2 is presented. During initial stages of the thesis there was no X-ray structure of hAgo2 available. In Chapter 3.2, a series of lengthy MD simulations on hAgo2/guide RNA complexes are described in order to understand the alleged discrimination among different guide RNA 5'-bases. Chapter 3.3 describes MD experiments which lead to the identification of a novel hAgo2 L2-Mid interaction crucial for catalytic function. In Chapter 3.4 experiments are listed to address the significance of destacking or kinking of the guide RNA within the complex. Chapter 3.5 describes studies addressing a potential function of the L2-Mid interaction for hAgo1. Finally, Chapter 3.6 describes a comparative analysis of the discovered L2-Mid interaction in all eukaryotic and prokaryotic Agos for which full-length X-ray structures were available.

Chapter 2. METHODS

2.1. Hardware used

All the computations were performed on Intel® Core™ i7-2600 Processor, 3.6 GHz, workstation at the Institute of Molecular medicine (IMM), University of Lübeck, Germany. The workstation runs on an LINUX base operating system (openSUSE 11.4). The final step of the MD simulations called the 'production run' was performed at the two High Performance Centers (HPC); "SARA", Amsterdam, the Netherlands and HLRN, Hannover, Germany.

2.1.1. SARA- HPC center, Amsterdam, The Netherlands

The access to SARA HPC center was acquired by applying to the European "*HPC Europa2"* grant. This grant was applied in collaboration with Prof. Alexandre M.J.J.Bonvin at the Bijvoet Center of Biomolecular Research, Utrecht University, The Netherlands. It is a translational access grant, which gives access to one of the seven HPC centers in Europe and an opportunity to the doctoral candidate for a funded research visit to the chosen host lab. The application process required a project proposal, statement of purpose and an extensive work plan of the project, with an estimate of the total number of computing hours required for the project.

In total 45,000 core-hours were awarded to run computations on the 'Huygens National Supercomputer'. The Huygens supercomputer is based on the IBM's next generation of POWER systems. The system consisted of the Compute Nodes (CNs) which were made up of 8 processors each, which are interconnected with each other through an Infiniband Network (IN), to increase the efficiency of computations involving Message Passing Interface (MPI). The system ran on Linux based operating system. The processor core of each CN could perform 4 floating point operations per clock cycle on 64-bit numbers. The Huygens supercomputer was only functional until the end of December 2012, therefore, the access to SARA HPC center was limited and all the computation had to be completed by the end of December 2012.

2.1.2. HLRN- HPC center, Berlin, Hannover, Germany

The North-German Supercomputing Alliance (Norddeutscher Verbund zur Förderung des Hoch- und Höchstleistungsrechnens - HLRN) is a distributed supercomputer system for science and research in seven north German states, with sites at two locations; ZIN in Berlin and RRLN in Hannover. The access to HLRN HPC center was gained by writing a short project proposal to the organization of the supercomputing center. The hardware is formed of CRAY XC30, which has 724 compute nodes, with 24 CPU core each. There are two CPU sockets per node, which are equipped with 12-core Intel Xeon IvyBridge processors. The total peak performance is 342.8 TFlop/s and the total memory 46.5 TByte. The system runs SuSE Linux Enterprise Server (SLES) operating system version 11.

2.2. Softwares used

During the course of this thesis, a wide range of softwares were used locally and remotely. These softwares are readily used in the field of computational biochemistry and biophysics. The following softwares were used in the present thesis:

2.2.1. UniProt

Universal Protein Resource (UniProt) is a database, which provides detailed protein sequence and functional information (158-160). It can be accessed freely at http://www.uniprot.org/. The website has multiple tools, which can be utilized to perform full text, and field based text research, similarity search and multiple sequence alignment. The detailed information of a protein can be obtained by submitting important relevant keywords or a specific identifier, which then directs the search to a web page which a series of proteins from which a desired protein can be selected based on the organism. It can be used to obtain the sequence of a protein in all official formats such as the FASTA format (161), which is often used for homology modelling.

2.2.2. HHPred

HHPred is a fast homology detection and structure prediction toolkit which is based on the pairwise comparison of Hidden Markov Models (HMM) (162). A wide range of databases such as PDB, SCOP, Pfam, CDD etc. can be accessed through it. A single query sequence of a single protein or multiple query sequences can be submitted to HHPred for homology detection and

structure prediction. It was used to detect the amino acid sequences homologous to the hAgo2 sequence. It is freely available software, accessible through a web server http://toolkit.tuebingen.mpg.de/hhpred or used on a local computer by a simple download process.

2.2.3. SWISS-MODEL

SWISS-MODEL is an automated server for comparative or homology modelling of 3D protein structures (163-165). All the steps involved in homology modelling are completely automated; template identification, alignment and model building. The 'SWISS-MODEL workspace' can be used to perform homology modelling. It can also be used for modelling in missing residues in a crystal structure in addition to the homology modelling. SWISS-MODEL is freely accessible at http://swissmodel.expasy.org/.

2.2.4. SALIGN

SALIGN software is used to align multiple sequences of protein structures, by dynamic programming optimization of a scoring function. This scoring function is a sum of an affine gap penalty and terms dependent various sequence and structure features (166). The features taken into consideration include the type of amino acid residue, the position of the residue, the accessible surface area of the residue, the secondary structure state of the residue and the conformation of a short segment centered on the residue.

A matrix of all pairwise protein alignment scores is constructed which builds the 'guide tree' used for the multiple sequence alignment. SALIGN has been benchmarked against other softwares available for multiple sequence alignment, it was demonstrated that SALIGN on average has 15% improvement in structural overlap (167). SALIGN can be freely accessed at http://modbase.compbio.ucsf.edu/salign/ or can be downloaded on a local computer.

2.2.5. MODELLER

MODELLER is a program, which is used for comparative or homology modelling of three-dimensional (3D) protein structures. Comparative or homology modelling predicts the 3D structure of a 'target' protein sequence based on alignment or homology to a 'template' sequence or sequences. The prediction of 3D structures is a multi-step process, which involves fold assignment, target-template alignment, model building and model evaluation. In addition to the

prediction of complete homology models, it is also readily used for modelling missing residues in the X-ray crystal structures of proteins.

The first version of MODELLER was reported in 1993, which has since been updated and is under constant up-gradation. For this thesis, the MODELLER 9v10 version was used to predict the 3D structure of hAgo2 before the X-ray crystal structures of hAgo2 were reported. MODELLER is freely available for academic users through a simple registration process available at https://salilab.org/modeller/about_modeller.html (168-171).

2.2.6. CHARMM

CHARMM is an acronym for Chemistry at HARvard Macromolecular Mechanics. It is a MD simulation package used to perform simulations of biomolecules (172). It is not freely available and a license is required (http://www.charmm.org/). It was used to perform implicit solvent simulations of the hAgo2-guide RNA complex.

2.2.7. Gromacs

It is a MD simulation package (http://www.gromacs.org/) freely available for all the users (173). It was developed in the early 1990s, as a collaborative group project between the Department of Chemistry and Computer Science at the University of Groningen. Gromacs is an abbreviation of GROningen Machine for Chemical Simulation. It was initially designed in FORTRAN, but has since been upgraded and rewritten in C. It is a parallel message-passing MD program. It is widely used for the simulation of biomolecules in aqueous and membrane settings. Gromacs supports different force fields such as Gromacs force field (based on GROMOS-87), GROMOS-96, Charmm 27, Charmm 36, OPLS, AMBER etc. It was used for the system preparation of the simulation system and for the final production runs. It has several in built functions that were used for the extensive analysis of the simulation trajectories.

2.2.8. MolProbity

It is an open access web server available at http://molprobity.biochem.duke.edu/. It is employed for all-atom quality validation of 3D structures of proteins, nucleic acids and complexes (174, 175). It evaluates the quality of models obtained through homology modelling, in addition could also be utilized for the analysis of models from X-ray crystallography. MolProbity gives a detailed all-atom analysis of the steric hindrances within molecules. It utilizes Ramachandran

and rotamer distributions for the analysis of sidechain and mainchain outliers. One of the most important aspects of the analysis is the addition and complete optimization of polar and nonpolar hydrogen atoms. The results of quality analysis by MolProbity are reported in multiple forms; as overall numeric scores, lists or charts of local problems, downloadable PDB files and 3D kinemage graphics shown online in the KiNG viewer (175).

2.2.9. DynDom

It is a domain motion analysis program (176, 177). The underlying idea behind Dyndom is that domains can be determined by their differing rotational properties (176, 177). The two conformations from principal component analysis are superimposed and then the rotation vectors from the rigid-body movement between the two conformations of each main-chain are calculated. The κ-means clustering algorithm (178) is then used to cluster these rotation vectors, which is the basis of defining the dynamic domain.

Next, the inter-domain screw axis is determined between the domains based on the Chasles' theorem, which states that the general displacement of a rigid body is a screw motion about a unique screw axis (179). This is used to determine the 'fixed domain' and the 'dynamic domain' based on six different parameters, orientation and location of the axes, in addition to the magnitude of rotation and translation of the domain. Finally an 'effective hinge axis' is determined, which is an inter-domain screw axis that passes near any of the inter domain residues. There are two types of effective hinge axes, twist axes parallel to the line joining the centers of mass of the two domains and closure axes perpendicular to this line. DynDom is freely available; it can be downloaded or accessed at http://fizz.cmp.uea.ac.uk/dyndom/.

2.2.10. VMD

Visual Molecular Dynamics (VMD) is a molecular visualization package (180). It can be downloaded at http://www.ks.uiuc.edu/Research/vmd/. It was used to display protein/RNA 3D structures. It was further utilized to visualize the MD simulation trajectories. It has several inbuilt plugins, which were used to perform preliminary analysis of the simulation trajectories. VMD was also used to generate the figures and movies of the simulation trajectories. It is also freely available for academic users and requires a simple registration.

2.2.11. PyMOL

It is a molecular visualization package available at http://www.pymol.org/ (181). It was used to display the protein/RNA 3D structures. It was used to display the electron density of X-ray crystal structures. It was also used to mutate the protein residues for the MD simulations. PyMOL was also for figure preparation. Although the entire PyMOL is commercial, however there is an academic version of the software available for the academic users.

2.2.12. COOT

It is a macromolecule model-building program, used primarily for protein model building using X-ray data (182, 183). It can also be utilized for model completion and model validation. It is available freely for academic users and is available at http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/. COOT is used to display maps and models, it also allows model manipulations like idealization, real space refinement, manual rotation/translation, rigid-body fitting, ligand search, rotamers, mutations etc. COOT was used to replace the 5'-nucleotides of the guide RNA for the MD simulations.

2.2.13. Xmgrace

It is a plotting tool and is freely available for UNIX systems at the following web site: http://plasma-gate.weizmann.ac.il/Grace/. The data can be imported into the software in ASCII or xvg format, the plots can be customized by using different options present in the axis properties. It was used to generate plots obtained from the analysis of the MD simulations. The plots can be saved and then printed in eps or ps format.

2.2.14. MATLAB

MATLAB® is an interactive environment used for numerical computation, visualization and programming. It can be used to analyze data, develop algorithms and to create models and applications. MATLAB® is not freely available and a license is required for access, it can be obtained at http://www.mathworks.com/products/matlab/. It was used for the data analysis of the MD simulations for calculating the total variance represented by individual principal components in addition to several other calculations.

2.3. Homology modelling of hAgo2

2.3.1. Template identification

To identify the templates which are homologous to hAgo2, first the protein sequence of hAgo2 was retrieved from the UniProt database (160) in the FASTA format (http://www.uniprot.org/uniprot/Q9UKV8.fasta) under the accession number Q9UKV8. The entire protein sequence of hAgo2 was used to identify homologous protein sequences by Hidden Markov Model Comparison using HHpred web server (162). This FASTA sequence of hAgo2 was then uploaded to the HHPred homology detection server. Altogether, 27 sequences were identified with sequence identity ranging between 3% - 99%. Five sequences were selected, *viz* 4ncb_A, 1u04_A, 2yhb_A, 2yha_A & 3luc_A. The sequences were selected on the basis of high sequence identity and the length of the sequences. The sequence '3luc_A' belongs to the Mid domain of hAgo2 for which the crystal structure was available. Although, the sequence identity of 4ncb_A, 1u04_A was very low 15% and 13% respectively, they were selected as they had sequence identity over a large segment of the hAgo2 sequence. 2yhb_A, 2yha_A sequences were only identical to the Mid and PIWI domains of hAgo2.

2.3.2. Multiple sequence alignment

In the next step, the query sequence (hAgo2) and the template sequences were aligned by the SALIGN program of the Modeller9v10 in a two step process (169-171). In the first step, the template sequences were aligned with each other using the Python script shown in Appendix 7.1, the alignment output is directed to the filename 'aligns.ali'. In the last step the query sequence (hAgo2) and the template sequences stored in the file align.ali are aligned with each other using another python script shown in the Appendix 7.2 and the final alignment is directed to the filename 'hAgo2_mult.ali'.

2.3.3. Model generation

The homology model of the full-length hAgo2 was generated by the Modeller9v10 program (169-171). The python script employed to generate the models is listed in Appendix 7.3. Altogether 100 models were generated, with a thorough optimization using the variable target function method (automodel.library_schedule = autosched.slow) with a maximum number of

conjugate gradient iteration of 300, followed by a slow refinement with simulated annealing MD (automodel.md_level = refine.slow).

2.3.4. Assessment and refinement

The initial assessment of the hAgo2 homology models generated was done on the basis of the normalized Discrete Optimized Protein Energy (DOPE) scores (184). Out of the 100 models initially generated, ten models with lowest DOPE scores were selected. The models were then assessed with the MolProbity (174). The MolProbity score has been reported to be a good measure of the protein-likeness in an analysis of all-atom accuracy of models from the Critical Assessment of Protein Structure-Prediction (CASP8) exercise (185). The final ten models were uploaded to the MolProbity web server (http://molprobity.biochem.duke.edu/), one at a time in standard PDB format. In the following step, the hydrogen atoms with no flips were added to the models. The 'all atom' contacts and geometry of the models were analyzed. The model with the lowest MolProbity score was selected.

2.4. MD system preparation: hAgo2

The system preparation step was performed on the workstation at the IMM. The recently reported crystal structure of hAgo2 (4F3T.pdb) in complex a guide RNA was used to perform MD simulations. To begin with the PDB file (4F3T) of the hAgo2-guide RNA complex was obtained from the Protein Data Bank (186). In this crystal structure the guide RNA was bound at its 5'-end (1-10 nucleotides) to the Mid domain and its 3'-end (17-20 nucleotides) was attached to the PAZ domain. However, the intermediating six nucleotides (11-16 nucleotides) were not observed in the crystal structure. Therefore, a truncated version (1-10 nucleotides) of the guide RNA was used. Gromacs simulation package was then used to prepare the chemical system and run MD simulations. The protocol developed in the Computational Structural Biology laboratory, at the Utrecht University, The Netherlands was utilized to prepare the MD system of hAgo2. However, the parameters and the particulars were changed accordingly, which were more suitable for the MD simulations of hAgo2. The protocol is available at the following web address http://nmr.chem.uu.nl/~adrien/course/molmod/md.html. The following protocol was employed for preparing and running the MD simulations:

2.4.1. Building missing residues

In the downloaded hAgo2 structure (4F3T.pdb) the following residues were missing 1–22, 120–126, 186–188, 245–247, 273–275, 603–606, 821–836. The system was prepared by first modeling the missing loop residues using a python script called Loopmodel.py (Appendix 7.4) with Modeller9v10 (171). The unstructured N (1-22) and C (821-859) termini of hAgo2 were excluded from the simulations. The missing residues of the hAgo2 were modelled using a python script Loopmodel.py (Appendix 1).

2.4.2. Structure conversion and topology generation

The PDB file of hAgo2-guide RNA complex only contains the structural coordinates. In order to perform MD simulations topology was constructed first, which includes the information of the parameters of the chemical system such as atom types, charges, bonds etc. The topology generated is specific to the force field used. Here, Amber Parm bsc0 force field was used which has been shown to have the best parameters for protein-nucleic acid interactions (173, 187). The structure was converted to the topology using the pdb2gmx program, in the presence of the TIP3P (Transferable Intermolecular Potential 3P) water model. The following command was used:

```
pdb2gmx -f ago_rna.pdb -o ago_rna.gro -p ago_rna.top -i ago_rna.itp
```

The chemical coordinates are stored in ago_rna.gro, topology of the system is contained in ago_rna.top and the energy coordinates are present in the ago_rna.itp.

2.4.3. Periodic boundary conditions

The simulations were performed under periodic boundary conditions (PBC), to avoid edge effects due to the walls of the simulation volume. It involves defining a cell of a specific shape, which is stacked in a space filling way; therefore, an infinite system can be simulated. In Gromacs simulation package, various shapes can be selected such as cubic, square etc. However, for this thesis a rhombic dodecahedron cell was used to set up the PBC, as it corresponds to the optimal packing of a sphere and allows greater rotation for the molecules. To prevent direct contact with the periodic images in the neighbouring cells the minimal distance was set up to 1.3 nanometer (nm), therefore the two neighbors were not allowed to get closer than 2.6 nm. The editconf program set up these PBC:

edticonf –f ago_rna.gro –o ago_rna-PBC.gro –bt dodecahedron –d 1.3

2.4.4. Energy minimization of the structure

The format in which the structure and topology of the system is generated corresponds to the force field used. This process involves addition or deletion of hydrogen atoms and sometimes these atoms might be positioned in a close proximity of the other atoms, thereby introducing a strain on the bonds. Thus, energy minimization of the system is of paramount importance, which allows the system to relax and remove clashes between the atoms. It is a two-step process; first, a run input file was created which combines the structural and topology information. This input file was created using a set of parameters, enclosed in the following parameter file called EM_VAC.mdp (Appendix 7.5).

The grompp program created the input file, the energy minimization in vacuum was performed over 50,000 steps, using steepest descent:
grompp –v EM_VAC.mdp –c ago_rna-PBC.gro –p ago_rna.top –o ago_rna_EM_VAC.tpr

In the second step, the input file (ago_rna_EM_VAC.tpr) was then submitted to the mdrun simulation program:
mdrun –f ago_rna_EM_VAC.tpr –o ago_rna_EM_VAC.gro

2.4.5. Solvent addition

The minimized structure inside the unit cell was then utilized to add the solvent. The solvent added was TIP3P water model. The solvent was added by the genbox program:
genbox –cp ago_rna_EM_VAC.gro –cs spc216.gro –p ago_rna.top –o ago_rna_water.gro

2.4.6. Addition of ions: counter charge and concentration

Counterions were added to neutralize the chemical system. First, an input file was created containing the structure and the topology in presence of a parameter file called EM_SOL.mdp (Appendix 7.6). The program grompp was used to create the input file:
grompp –f EM_SOL.mdp –c ago_rna_water.gro –p ago_rna.top –o ago_rna_water.tpr

The input file (ago_rna_water.tpr) was then utilized to add the counterions. The concentration was specified to be 0.15M and the ions added were Na^+ and Cl^-. The program genion was used to add the counterions and the group 'SOL' (solvent) was chosen to add these counterions:

Genion –s ago_rna_water.tpr –o ago_rna_solvated.gro –conc 0.15 –neutral –pname Na^+

–nname Cl^-

2.4.7. Energy minimization of the solvated system

Next, the solvated system was energy minimized, which gets rid of the unfavorable interactions such as overlapping atoms, equal charges too close together etc. due to the addition of the solvent and the counterions. The solvated system is minimized in two steps, first the input file generation, which combines the structure and the topology using grompp in presence of a parameter file EM_SOL.mdp used in the previous step:

grompp –f EM_SOL.mdp –c ago_rna_solvated.gro –p ago_rna.top –o ago_rna_EM_solvated.tpr

The input file (ago_rna_EM_solvated.tpr) was then energy minimized over 50,000 steps, under steepest descent method, using the program mdrun:

mdrun –f ago_rna_EM_solvated.tpr –o ago_rna_EM_solvated.gro

2.4.8. Position restrained MD simulations

In this step, the solvent settles in with the protein, which is accomplished by the allowing free movement of the solvent, whilst keeping the non-hydrogen atoms of the protein relatively fixed to the reference position. This step ensures that the solvent configuration is similar to that of the protein. This is also the first step of equilibration, which is carried out in the canonical or isothermal-isochoric ensemble. The NVT [Number of particles (N), Volume (V), Temperature (T)] equilibration stabilizes the temperature of the system. The control parameters for input file generation are contained in the NVT_PR.mdp file (Appendix 7.7). The input file was produced with the program grompp:

grompp –f NVT_PR.mdp –c ago_rna_EM_solvated.gro –p ago_rna.top –o ago_rna_nvt_pr.tpr

The input file (ago_rna_nvt_pr.tpr) was then utilized to perform restrained MD over 50,000 steps with mdrun:

mdrun –s ago_rna_nvt_pr.tpr –o ago_rna_nvt_pr.gro

2.4.9. Releasing the restraints

During this step the pressure and density of the system is stabilized. Isothermal- isobaric (NPT) ensemble was used, to equilibrate the pressure, keeping the number of atoms, temperature and pressure constant. The restraints were released slowly in the presence of control parameters placed in NPT_PR.mdp file (Appendix 7.8).

The input file was produced with grompp program:

grompp –f NPT_PR.mdp –c ago_rna_nvt_pr.gro –p ago_rna.top –o ago_rna_npt_pr1000.tpr

This input file (ago_rna_npt_pr1000.tpr) was equilibrated over 50,000 steps with mdrun program:

mdrun –s ago_rna_npt_pr1000.tpr –o ago_rna_npt_pr1000.gro

This equilibration step was, further repeated with a progressive release of constraints over 50,000 steps.

sed –e ‘s/1000 1000 1000/ 100 100 100/g’ ago_rna.itp> tmp.itp

mv tmp.itp ago_rna.itp

grompp –f NPT_PR.mdp –c ago_rna_npt_pr1000.gro –p ago_rna.top –o ago_rna_npt_pr100.tpr

mdrun –s ago_rna_npt_pr100.tpr –o ago_rna_npt_pr100.gro

More restraints were released by the final equilibration carried over 50,000 steps:

sed –e ‘s/100 100 100/ 10 10 10/g’ ago_rna.itp> tmp.itp

mv tmp.itp ago_rna.itp

grompp –f NPT_PR.mdp –c ago_rna_npt_pr100.gro –p ago_rna.top –o ago_rna_npt_pr10.tpr

mdrun –s ago_rna_npt_pr10.tpr –o ago_rna_npt_pr10.gro

2.4.10. Unrestrained MD simulations

At the final step of equilibration, the position restraints were released from the previously equilibrated system. The control parameters were used from the file NPT_NO_PR.mdp (Appendix 7.9). The input file for the final equilibration step was produced by grompp program:

grompp –f NPT_NO_PR.mdp –c ago_rna_npt_10.gro –p ago_rna.top –o ago_rna_npt_no_pr.tpr
This input file (–o ago_rna_npt_no_pr.tpr) was then equilibrated over 50,000 steps with mdrun program:
mdrun –s ago_rna_npt_no_pr.tpr –o ago_rna_npt_no_pr.gro

2.4.11. Production simulation

The equilibrated system was then employed to run long timescale simulations, under control parameters contained in the MD.mdp file (Appendix 7.10).
The input file for the production simulation was generated by grompp:

grompp –f MD.mdp –c ago_rna_npt_no_pr.gro –p ago_rna.top –o ago_rna_md.tpr

The input file (ago_rna_md.tpr) was then run for 100 ns with mdrun. The file was then transferred to a Supercomputer.

The solvated system was minimized over 50000 phase. The system was then equilibrated in two 100 ps simulations under NVT and NPT conditions, respectively, with position restraints on the solute. The position restraints of the system were then released to start the production run. The non-bonded interactions were calculated by the particle mesh ewald (PME) method with an order of 4 and a Fourier spacing of 0.16. Non-bonded cutoff for the van der Waal (vdw) interactions was set to 1.0 nm. The production run for each system was performed at 2 femtosecond (fs) time step, generating 100 ns in each system. The production run of the MD simulations was performed at SARA supercomputing facilities using script Gromacs 4.6.3 (Appendix 7.11).

2.4.12. Generation of the guide RNA mutants

These subsequent simulations were performed on the hAgo2- guide RNA complex in which the 5’-U of the guide RNA was replaced to 5’-A, 5’-C and 5’-G respectively. The simulations containing the different 5’-bases are referred to as hAgo2_5’-U, hAgo2_5’-A, hAgo2_5’-C and hAgo2_5’-G respectively throughout this thesis (Table 2-1).

To change the 5'-base of the guide RNA for MD simulations, COOT software was used (182, 183). The starting structure used for the simulations was uploaded to the COOT program and the 'simple mutate' option from Model/Fit/Refine dialog box was selected. The nucleotide to be replaced was then selected and changed to the desired nucleotide. A similar MD system preparation protocol was followed for the subsequent simulations of hAgo2.

2.4.13. Generation of the D358A and I365A hAgo2 mutants

For the D358A and I365A hAgo2 mutant simulations, the protein mutants were generated using the PyMOL software (181). The starting structure used in the previous simulations was uploaded to PyMOL molecular graphics system. In the wizard menu, mutagenesis was selected and then the D358 protein residue was selected. The option 'No selection' was used and the D358 was chosen. The rotamer with the best fit to an alanine residue was then selected. The simulations corresponding to these two mutants are referred to as D358A-hAgo2 and I365A-hAgo2 in Table 2-1.

2.5. MD system preparation: hAgo1

The system preparation step was performed on the workstation at the IMM. Two crystal structures of the hAgo1 in complex a guide RNA were recently reported 4KRE.pdb (104) and 4KXT.pdb (115). For this thesis, the 4KRE.pdb crystal structure was used for performing MD simulations since the structure was reported at a higher resolution of 1.75 Å and an R-value of 0.172. The PDB file (4KRE.pdb) of the hAgo1-guide RNA complex was obtained from the Protein Data Bank (186). The crystal structure had the following missing residues: 1-16, 82-83, 117-123, 240-244, 272-273, 330-333, 601-604, and 818-834. The unstructured residues at the N (1-16) and C (818-857) termini of hAgo1 were excluded from the simulations.

The missing residues were modelled using the SWISS-MODEL automated server (http://swissmodel.expasy.org/). A truncated version (1-9 nucleotides) of the guide RNA was used; the two nucleotides attached to the PAZ domain were removed. Gromacs simulation package was then used to prepare the chemical system and run MD simulations. The simulation protocol described above in the Chapters 2.4.2 - 2.4.7 was employed for preparing the chemical system MD simulations. The simulations were performed under PBC conditions using a rhombic dodecahedron box.

The system was solvated using TIP3P water model and Na+ and Cl- ions were added to neutralize the system. Altogether, the solvated system with the ions was formed of 156234 atoms. The system was further energy minimized and equilibrated using the protocol mentioned in Chapters 2.4.8 - 2.4.11 of this thesis. For the subsequent simulations of the D356A-hAgo1, the D356 protein residue was mutated to an alanine following the protocol described above in Chapter 2.4.13. These simulations are referred to as wt-hAgo1 and D356A-hAgo1 simulations in the Table 2-1and the entire thesis. The production run was performed at 2 fs time step, generating 100 ns in total. The production run of all the simulations was run at the HLRN HPC center. The simulations were run on 16 nodes (128 CPU's) in parallel.

2.6. MD system preparation: KpAgo

The crystal structure of KpAgo (4F1N.pdb) was used (111). The PDB file (4F1N.pdb) of the KpAgo/guide RNA complex was obtained from the Protein Data Bank (186). The structure was reported at a resolution of 3.19 Å. However, the structure had long loops of missing residues: 1-216, 283-299, 351-381, 573-586, 603-627, 713-720, 837-864 and 1220-1229. The unstructured N-terminal (1-216) of the KpAgo structure was excluded from the simulations. The other missing loops were modelled using automated SWISS-MODEL server (http://swissmodel.expasy.org/). The guide RNA had 1-9 nucleotides, which were bound to the Mid domain.

Gromacs simulation package was then used to prepare the chemical system and run MD simulations. The simulation protocol described above in the Chapters 2.4.2 - 2.4.11 was employed for preparing and running the MD simulations. The simulations are referred to as KpAgo in the Table 2-1and throughout the thesis. The production run was performed at 2 fs time step, generating 100 ns trajectories. The production run of the simulations was performed at the HLRN HPC center and the simulations were run on 16 nodes (128 CPU's) in parallel.

2.7. MD system preparation: TtAgo

There are a series of crystal structures available for TtAgo in various combinations of guide DNA and target RNA. Please note: prokaryotic Agos use DNA oligonucleotides instead of RNA as guide strands. For this thesis, a binary complex of TtAgo (3DLH.pdb) was selected (110). The PDB file (3DLH.pdb) of the TtAgo/guide DNA complex was obtained from the Protein Data

Bank (186). The structure was reported at a resolution of 3.0 Å. The structure had only a few missing residues: 273-275, 496-497 and 508-513. The missing loops were modelled using automated SWISS-MODEL server (http://swissmodel.expasy.org/). The guide DNA was bound to the Mid domain (nucleotides 1-12), only a stretch of nucleotides (18-21) was bound to the PAZ domain; a fragment of the guide DNA (nucleotides 13-17) was missing. For the simulations a guide with nucleotides 1-12 was used.

Gromacs simulation package was then used to prepare the chemical system and run MD simulations. The simulation protocol previously described in the Chapters 2.4.2 - 2.4.11 was employed for preparing and running the MD simulations. The production run was performed at 2 fs time step, generating 100 ns trajectory. The production run of the simulations (Table 2-1) was performed at the HLRN HPC center and the simulations were run on 16 nodes (128 CPU's) in parallel.

2.8. MD system preparation: PfAgo

There are three crystal structures available for PfAgo, for this thesis, 1U04.pdb structure of PfAgo was selected as it had the highest resolution (105). The PDB file (1U04.pdb) of the PfAgo was obtained from the Protein Data Bank (186). The structure was reported at a resolution of 2.25 Å. The structure had the following missing residues: 27-38, 253-257, 278-281, 347-354 and 414-442. The missing loops were modelled using automated SWISS-MODEL server (http://swissmodel.expasy.org/). There was no guide DNA bound to this structure; hence, the simulations were performed in the absence of the guide.

Gromacs simulation package was then used to prepare the chemical system and run MD simulations. The simulation protocol described above in the Chapters 2.4.2 - 2.4.11 was employed for preparing and running the MD simulations. The production run was performed at 2 fs time step, generating 100 ns trajectory. The production run of the simulations (Table 2-1) was performed at the HLRN HPC center and the simulations were run on 16 nodes (128 CPU's) in parallel.

2.9. MD system preparation: AaAgo

For this thesis, 2NUB.pdb structure of AaAgo was selected (106). The PDB file (2NUB.pdb) of the AaAgo was obtained from the Protein Data Bank (186). The structure was reported at a resolution of 3.20 Å. The structure had the following missing residues: 134, 176-177, 245-247 and 262-271. The missing loops were modelled using automated SWISS-MODEL server (http://swissmodel.expasy.org/). Gromacs simulation package was then used to prepare the chemical system and run MD simulations. The simulation protocol described above in the Chapters 2.4.2 - 2.4.11 was employed for preparing and running the MD simulations. The production run was performed at 2 fs time step, generating 100 ns trajectory. The production run of the simulations (Table 2-1) was performed at the HLRN HPC center and the simulations were run on 16 nodes (128 CPU's) in parallel.

Table 2-1: List of MD simulations performed in the present thesis.

Chemical System	Force field	Timescale	Repeats
hAgo2 alone	Amber03 ILDN	100ns	2
hAgo2_5'-U	AMBER Parmbsc0	100ns	2
hAgo2_5'-A	AMBER Parmbsc0	100ns	2
hAgo2_5'-C	AMBER Parmbsc0	100ns	2
hAgo2_5'-G	AMBER Parmbsc0	100ns	3
D358A-hAgo2	AMBER Parmbsc0	100ns	2
I365A-hAgo2	AMBER Parmbsc0	100ns	1
wt-hAgo1	AMBER Parmbsc0	100ns	2
D356A- hAgo1	AMBER Parmbsc0	100ns	2
KpAgo	AMBER Parmbsc0	100ns	1
TtAgo	AMBER Parmbsc0	100ns	1
PfAgo	AMBER Parmbsc0	100ns	1
AaAgo	AMBER Parmbsc0	100ns	1

2.10. Analysis of the simulations

2.10.1. Root mean square deviation (RMSD)

The RMSD of Cα atomic coordinates is a standard measure of determining dissimilarity between protein structures (188). To calculate the RMSD the structures to be compared are superimposed. The RMSD is calculated as follows:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i^X - r_i^Y)^2}. \quad [1]$$

Here, N is the number or atoms, r^X is the target structure and r^Y is the reference structure. The RMSD is calculated by selecting a reference frame, usually the starting structure of the MD simulations and then superimposing all the succeeding frames to this reference structure. The RMSD is also used as a measure for calculating the convergence of the MD simulations.

The RMSD calculations of the trajectories from MD simulations were performed using the program 'g_rms' (189) built in the Gromacs simulation package (173). Another method of identifying the transitions between structures is to create a RMSD matrix. In order to create the RMSD matrix 'g_rms' module is called with two different simulation trajectories. A 'cross-RMSD' matrix can also be created to individualize groups (clusters) and transitions between the different trajectories. In order to create this cross-RMSD matrix all four trajectories can be concatenated into one trajectory. It results in a grayscale matrix, which can further be transformed into the rainbow gradient.

2.10.2. Root mean square fluctuations (RMSF)

The root mean square fluctuations (RMSF) are useful in distinguishing the local stability and fluctuations along the protein chain. The RMSF of a residue can be computed as follows:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < \left(r_i'(t) - r_i(t_{ref})\right)^2 > } \; . \qquad [2]$$

Here *T* is the time of the trajectory over which the *RMSF* is calculated, t_{ref} is the reference time, r^i is the position of the residue *i* and *r'* is the position of atoms in residue *i* after superposition on the reference. The angle brackets show the average of the square distance is calculated over the selection of atoms in the residue.

The RMSF of a protein can also be used to compute the B-factors (B) of the protein as follows:

$$RMSF = \sqrt{\frac{3B}{8\pi^2}} . \qquad [3]$$

The 'g_rmsf' module of Gromacs simulation package was used to calculate the RMSF and B factors of the simulations.

2.10.3. Principal component Analysis (PCA)

Principal component analysis (PCA) is one of the crucial techniques employed commonly to investigate the conformational changes in proteins (190). PCA is carried out on an ensemble of conformations obtained from the MD simulations. The outcome of the Principal components (PC's) is classified according to their individual contribution to the total fluctuation of the entire ensemble. It has been demonstrated that a small number of these PC's can actually represent a large amount of the total displacement (191).

PCA is performed in two major steps: 1) Calculation of the covariance matrix (C) of the positional deviations 2) Diagonalization of this covariance matrix. The covariance matrix is calculated based on the number of protein structures in the ensemble.

The covariance matrix 'C' is calculated as follows:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle. \quad [4]$$

Here, x_i and x_j are the atomic coordinates, while the brackets denote the ensemble average. The diagonalization of this covariance matrix is similar to equating the eigenvalue problem:

$$A^T C A = \lambda. \quad [5]$$

Here A denotes the Eigenvectors and λ represents the associated Eigenvalues.

The PCA calculations in the present thesis were performed with the Gromacs simulations package. The module 'g_covar' was used to construct and diagonalize the covariance matrix. The Eigenvectors were calculated with the module 'g_anaeig'.

2.10.4. Distance calculations

The distance between center of mass of specific atoms as a function of time was calculated using the 'g_dist' module of the Gromacs simulation package. The desired atoms were first placed in specific groups using the 'make_ndx' module of the Gromacs package.

The mean distance between specific atoms and residues was also calculated for which the 'g_mindist' module of Gromacs was used. First, the desired atoms or residues were placed into individual groups with the 'make_ndx' module.

2.10.5. Hydrogen bond interactions

Hydrogen bond formation is one of the most significant interactions in molecular biology. A hydrogen bond formation occurs when a proton (H) covalently attached to an electronegative atom donor (D) is shared with another electronegative atom acceptor (A). The hydrogen bond formation has been related with the protein stability and secondary structure formation (192). MD simulations has often been utilized as a tool to unravel novel hydrogen bond interactions in proteins, which were not observed in the crystal structure (193).

The initial hydrogen bond identification during the simulations was performed by 'VMD hydrogen bond' plugin. The hydrogen bonds were further identified the 'g_hbond' module of

Gromacs. The hydrogen bond formation cutoff distance of 3.5 Å between the donor and acceptor was used and an angle of 30° between Donor-Hydrogen-Acceptor.

2.10.6. Salt bridge interaction

Salt bridge interactions occur frequently in proteins providing conformational stability and contribute to molecular recognition (194). Salt bridge formation occurs between oppositely charged protein residues within hydrogen bonding distance. The protein residues, which frequently make salt bridges, are Aspartate (Asp/D), Glutamic acid (Glu/E), Histidine (His/H), Arginine (Arg/N, and Lysine (Lys/K).

The preliminary salt bridge identification was performed with the VMD 'Salt bridge' plug-in. The Gromacs 'g_saltbr' module was used to identify the salt bridges in the proteins. The cutoff distance of <3.5 Å between the heavy atoms was used for salt bridge calculation.

2.10.7. Movie making with VMD

VMD was utilized to analyze the trajectories from MD simulations. VMD possesses a very important 'movie making' plugin which can be utilized to create movies of the MD simulation trajectories. The protocol to generate the high-resolution movies is as follows:

1. In the first step, the trajectory of the MD simulation is loaded using the "stride" feature in the VMD trajectory loader dialog. It is important to note that the option "load all frames at once" should be selected rather than "background".
2. To prevent the protein structure from wobbling around the screen, a part of the structure was "fixed". To fix a part of the structure 'RMSD Tool' was clicked and the specific protein residues were selected. These residues were then aligned using the 'align' feature of VMD, which aligns the entire structure through all frames. This makes sure that these specific residues remain as stationary as possible.
3. In the next step, a desired representation of the structure can be selected with "Create Rep" feature. The selection box can be used to specify the representations. In addition, different types of views (coloring, style, transparency) for each representation can be selected individually.

4. It is important to smooth the trajectory over several frames for each representation. A 4 ns smoothing window reduces the distracting movements in the protein without eliminating interesting side chain motion.
5. To render the movie the ‘Rendermode’ was set to OpenGL/CUDA.
6. For movie making, the ‘movie maker’ feature was selected. The movie settings were set to trajectory, which prevents the structure to wobble.
7. Next, the smoothing window of 4 was selected. The internal Tachyon ray tracer was used.

Chapter 3. RESULTS

3.1. Homology modelling of hAgo2

The crystal structures of hAgo2 were reported in 2012 (113, 123). However, there were no insights into the structural organization of full-length hAgo2 prior to these crystal structures. Therefore, initially homology modelling of full-length hAgo2 was performed with Modeller 9v10 (166).

3.1.1. Template identification

The template identification of the complete protein sequence of hAgo2 was performed with the HHpred web server (162). In total, 27 sequences were identified with sequence identity ranging between 3% - 99%. Five sequences were selected, *viz* 4ncb_A, 1u04_A, 2yhb_A, 2yha_A & 3luc_A. The sequences were selected based on high sequence identity and the length of the sequences (Table 3-1). The sequence '3luc_A' belongs to the Mid domain of hAgo2 for which the crystal structure was available (126). Although, overall the sequence identity of 4ncb_A and 1u04_A was very low with 15% and 13%, respectively, they were selected since their sequence covered a large segment of the hAgo2 sequence. 2yhb_A and 2yha_A sequences only matched the Mid and PIWI domains of hAgo2.

Table 3-1: List of the template sequences identified with HHPred server, which were used for the homology modelling of the full-length hAgo2.

Sequence identifier	Organism	Sequence identity	Sequence length
4ncb_A	*Thermus thermophilus*	15%	685
1u04_A	*Pyrococcus furiosus*	13%	771
2yhb_A	*Neurospora crassa*	31%	437
2yha_A	*Neurospora crassa*	28%	388
3luc_A	*Homo sapiens*	99%	138

3.1.2. Homology model of hAgo2

Modeller9v10 was used to generate 100 initial models (166). The final model was selected based on the lowest DOPE score. The final model was then assessed with MolProbity (174, 175). Various criteria such as the clashscore, Ramachandran outliers, poor rotamers along with bad backbone bonds and angles were assessed, in the end a MolProbity score is marked. A low MolProbity score is indicative of a high quality structure. The results from the MolProbity analysis of the homology model in comparison to the crystal structure of hAgo2 (4F3T.pdb) are exhibited in the Table 3-2.

The overall structural organization exhibited by the homology model of hAgo2 is similar to the X-ray crystal structure (4F3T.pdb). However, there are major differences concerning secondary structures. These differences are most pronounced in the PAZ, N and the linker domains (Figure 3-1). An overall RMSD of ~24 Å was observed between the homology model and the X-ray structure of hAgo2.

Table 3-2: Comparison of MolProbity assessment of the hAgo2 homology model (HM-hAgo2) and X-ray crystal structure of hAgo2 (X-ray-hAgo2).

Property	HM-hAgo2	X-ray-hAgo2 (4F3T.pdb)
MolProbity score	4.11	1.42
Clashscore, all atoms	269.58	5.38
Poor rotamers	5.72%	0.98%
Ramachandran outliers	5.13%	0.13%
Ramachandran favored	85.53%	97.33%
Cβ deviations >0.25Å	7.20%	0.0%
Bad backbone bonds	1.46%	0.0%
Bad backbone angles	4.46%	0.0%

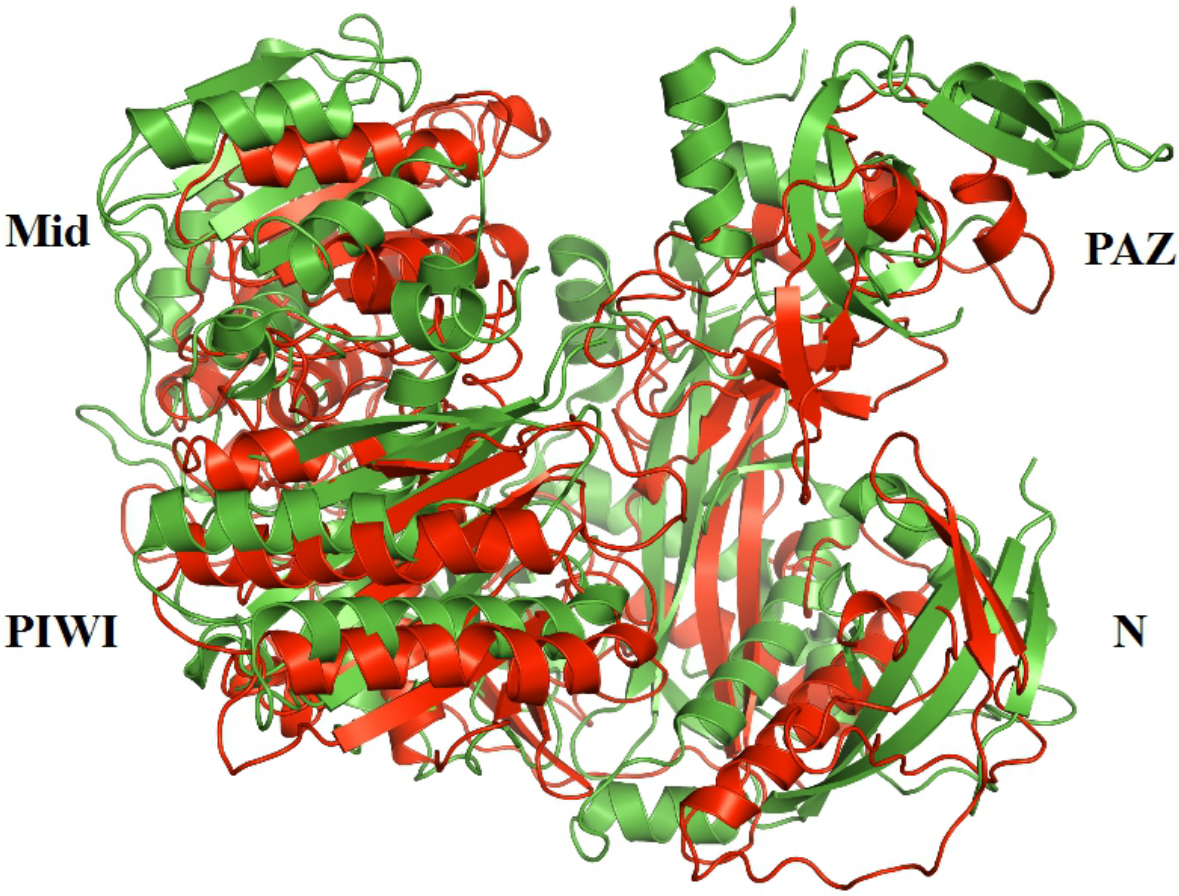

Figure 3-1: Superposition of the hAgo2 homology model and the X-ray crystal structure. Red cartoons represent the homology model of hAgo2 and the X-ray crystal structure of hAgo2 is shown in green cartoons.

3.2. The effect of different guide RNA 5'-bases on the dynamic behaviour of the hAgo2

3.2.1. hAgo2 is stabilized by protein-RNA interactions

hAgo2 is a multidomain protein (Figure 3-2 (a)) which undergoes considerable conformational changes in order to achieve slicer activity (19, 152, 153). Thus far, all the insights into these conformational changes have been provided by a series of TtAgo structures. However, insights into hAgo2 are limited, therefore to investigate these conformational changes in the hAgo2, MD

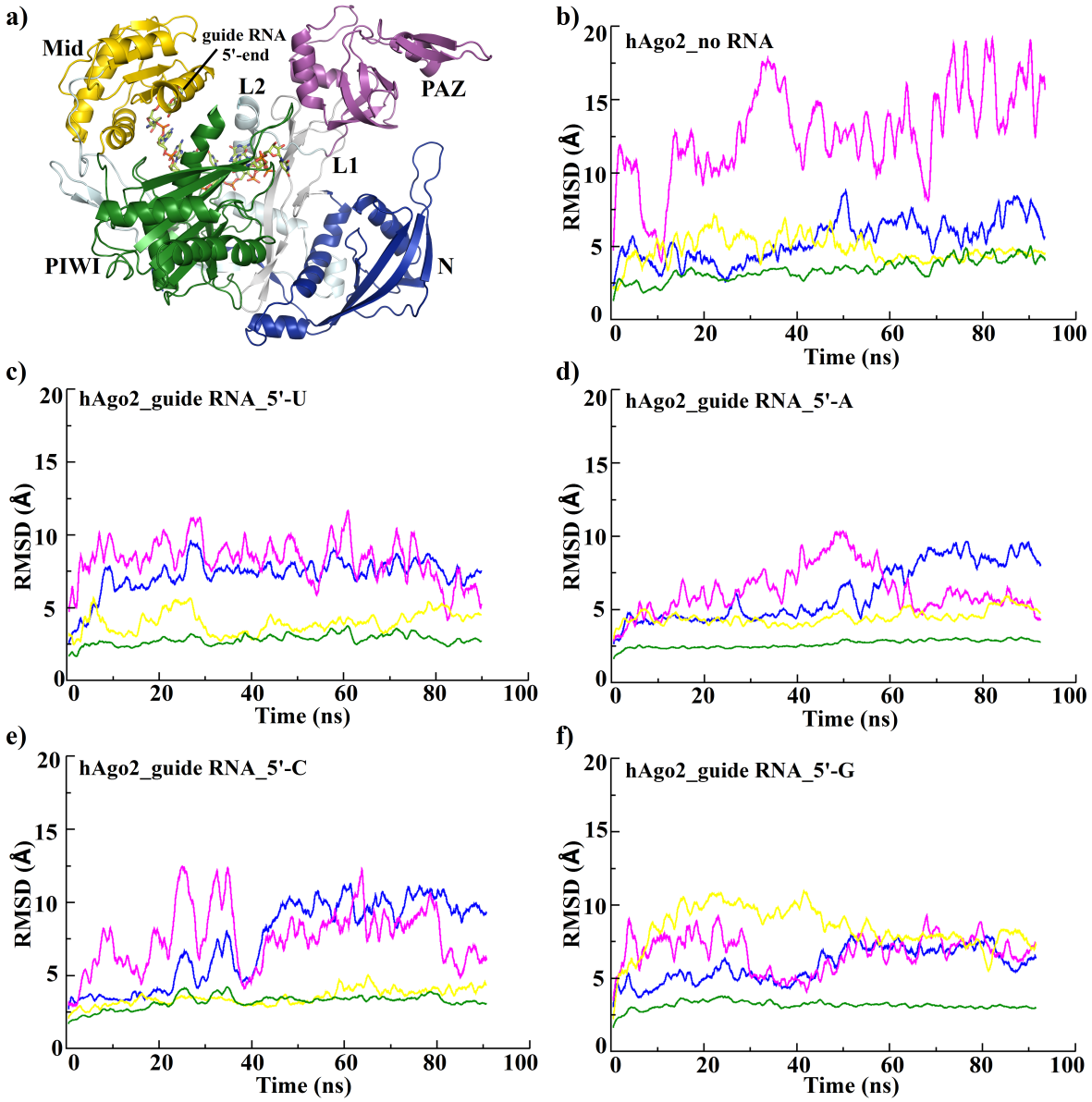

Figure 3-2: Analysis of backbone RMSD as a function of the simulation time for various hAgo2/guide RNA complexes. (a) Full-length structure of hAgo2 in complex with a truncated (1-10nt) guide RNA (PDB 4F3T). The individual domains are labeled and color-coded: N (blue), PAZ (magenta), Mid (gold) and PIWI (green) joined by two linker regions, L1 & L2 (grey). The protein is represented in cartoon and the guide RNA in licorice. (b) Free hAgo2 in the absence of bound guide RNA. hAgo2-guide RNA complexes with U (c), A (d), C (e) or G (f) at the 5'-end of the guide RNA.

simulations of the hAgo2-guide RNA complex (PDB ID code 4F3T) were performed for 100 ns, using the Gromacs simulation package (http://www.gromacs.org/). This study illustrates that the overall fold of the protein and individual domains is conserved during the timescale of the

simulations. The hAgo2 backbone exhibits small RMSD changes (~3-4 Å) in comparison to the structure at t=0 ns of the simulation (Figure 3-2 (c)). It was further observed, that the overall positioning of the guide RNA bound to hAgo2 is also preserved. MD simulations of the hAgo2-guide RNA complex revealed a breathing motion of the protein caused by large movements of the N and PAZ domains. The movement and flexibility of the PAZ domain dominates this breathing motion, although the L1 and L2 linker regions seem to regulate and orchestrate it. The PAZ domain behaves like a pulley, tethered on both sides by the two-linker regions acting like a fictitious rope, which pulls the N domain on one side and the Mid/PIWI domains on the other. Despite of the large motions in the protein, the PIWI domain remains stable throughout, acting like the core of the protein. Moreover, MD simulations of hAgo2 in the absence of bound RNA confirm that the PAZ domain is the major contributor towards global protein flexibility. The RMSD of the PAZ domain increased from ~ 5 Å to 8 Å in the absence of RNA (Figure 3-2 (b)).

3.2.1. 5’-bases induce different inter-domain motion in hAgo2

To study the effect of the four possible 5’-bases on the inter-domain motion of the hAgo2-guide RNA complex, MD simulations were performed where the 5’-U was subsequently replaced by A, C and G, respectively. The simulations revealed that individual 5'-base interactions differently affect the observed inter-domain motions of the hAgo2. In the presence of a 5’-U guide RNA the hAgo2 domain motion is dominated by movements of the PAZ and N domains, whilst the Mid and PIWI domains are relatively less flexible (Figure 3-2 (c)). When a 5’-U was replaced with a 5’-A, it was observed that hAgo2 becomes more flexible and consequently the domain motions for the PAZ and N domains become more pronounced (Figure 3-2 (d)). Nevertheless, overall the inter-domain motion pattern of hAgo2 is comparable to the situation with 5’-U. In the presence of a 5’-C the inter-domain motion pattern further changes and the protein becomes even more flexible. The motion of the N domain noticeably becomes more explicit as seen by high RMSD values (Figure 3-2 (e)). Though, hAgo2 retains its usual alternating breathing motion caused by large movements of the PAZ domain. The Mid domain appears less flexible than observed for 5'-U or 5'-A while the PIWI domain remains to be the least flexible of all the domains, retaining its function as the core of the protein. Surprisingly, in the presence of 5’-G the inter-domain motion of hAgo2 is unique as compared to the other three bases showing an antagonistic domain motion pattern (Figure 3-2 (f)). Here, especially extensive movements of the Mid domain replace the

conventional pattern dominated by movements of the PAZ domain. The Mid domain becomes very flexible attaining RMSD values of up to ~ 8 Å (Figure 3-2 (f)). The PAZ and N domains also contribute to these domain movements, whereas the behavior of the PIWI domain remains unaltered.

The specific inter-domain protein motion pattern of hAgo2 is corroborated by PCA of the trajectories in the presence of the respective 5'-bases. For each of the different 5'-bases a distinct pattern is observable. In the presence of a 5'-U the most pronounced domain motion is observed for the PAZ domain (Figure 3-3 (a)). The domain motion of hAgo2 is similar in the presence of 5'-A; a correlated motion of PAZ and parts of the N domain is observed (Figure 3-3 (b)). In the

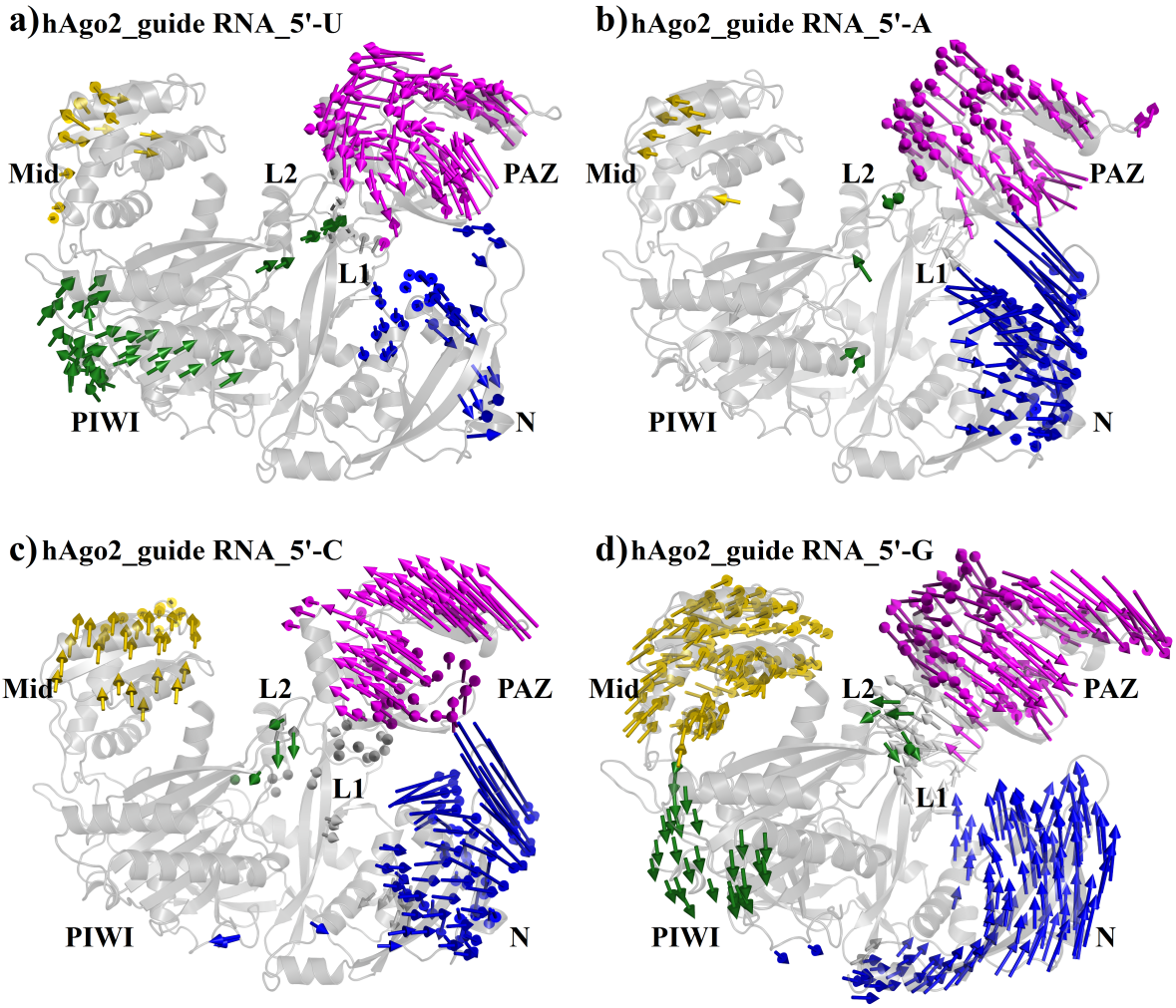

Figure 3-3: PCA analysis of the hAgo2-guide RNA complex in the presence of different bases at 5'-end of the guide RNA. Color-coded arrows represent domain motions obtained from PCA; N (blue), PAZ (magenta), Mid (gold) and PIWI (green). **a)** 5'-U **b)** 5'-A **c)** 5'-C **d)** 5'-G.

presence of 5'-C the domain motion of the N domain becomes more prominent than the PAZ domain (Figure 3-3 (c)). Yet, the most striking motion occurs in the presence of 5'-G, where the Mid domain shows an unusually high domain motion, not observed for any of the other 5'-bases (Figure 3-3 (d)).

Additionally, the inter-domain motion of hAgo2 is further investigated with the aid of cross-RMSD plots. The cross-RMSD plot for the Mid domain (Figure 3-4 (a)) shows that this domain is stable and does not show much domain motion, except in the presence of a 5'-G, as described above. However, a clear difference was observed for the N domain (Figure 3-4 (b)). The cross-RMSD values for the N domain constantly increase as we replace 5'-U with 5'-A, 5'-G and 5'-C. This implies that the 5'-bases directly influences the motion of the N domain.

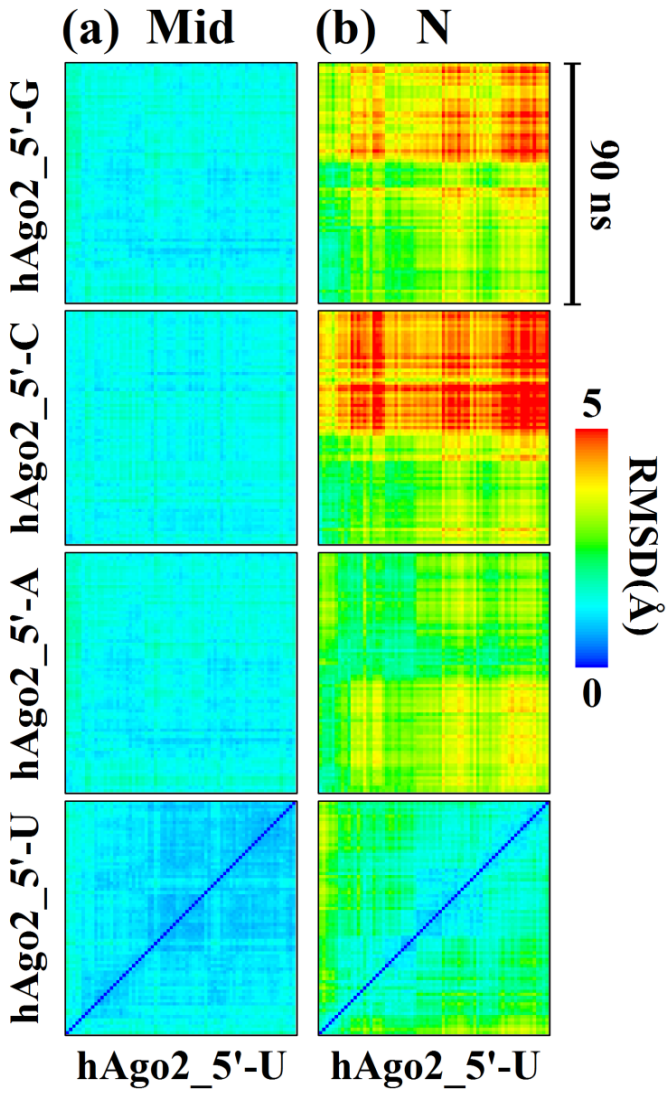

Figure 3-4: Cross-RMSD plots of hAgo2 Mid (a) and N (b) domain. The different 5'bases are indicated. The RMSD gradient (0-5 Å) reveals very little RMSD changes in the Mid domain independent of the guide 5'-end. Concerning the N domain there is clear difference observable which gradually increases when moving form 5'-U to 5'-G.

3.2.2. Novel insights into the 5'-base discrimination by hAgo2

The interaction network between 5'-A (Figure 3-5 (a, b)) and the NS loop was preserved throughout the simulations. In addition, it was observed through MD simulations that 5'-U (Figure 3-6 (a, b)) also retains the initial interaction pattern with the NS loop through the entire length of the simulation. The interaction pattern of 5'-C is stabilized by a hydrogen bond formation with Q548 sidechain (Figure 3-5 (d, e)). The most striking interaction pattern was observed between 5'-G and the neighbouring protein residues. It was noticed that the Mid domain undergoes a huge domain movement at ~20 ns of the simulation, tilting the Mid domain and bringing the NS loop in close proximity of the L2 region. This movement triggers hydrogen bond formation between the -NH2 and carbonyl group of the base 5'-G and OD1 and OD2 of ASP358 present in the helix7 of the L2 region (Figure 3-6 (e)). Once the hydrogen bond formation occurs, it is retained thereafter for the remaining time of the simulation.

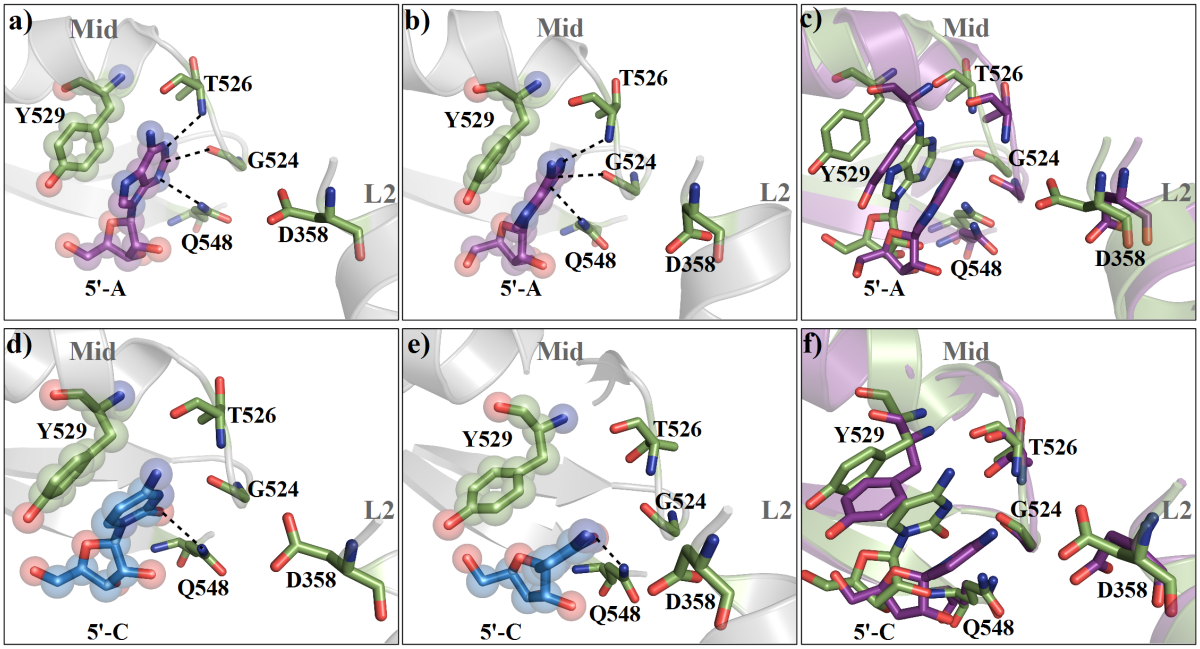

Figure 3-5: Close-up of the guide RNA 5'-end and relevant protein residues of the Mid domain and L2 linker region. The Mid domain and the L2 helix are represented in cartoon (grey), important protein residues (green) and 5'-A (magenta) or 5'-C (dark blue) are shown in sticks. For clarity merely the 5'base is shown. Hydrogen bonds are represented by black dotted lines. Transparent spheres highlight stacking interactions between protein residues and the corresponding bases. (a, d) show the initial situation and (b, e) after 20 ns of simulation of 5'-A and 5'-C terminated hAgo2 bound guide strands, respectively. (c, f) superposition of (a, b) and (d, e) with neighboring protein residues at t=0 ns are shown in green and at t=20 ns in magenta.

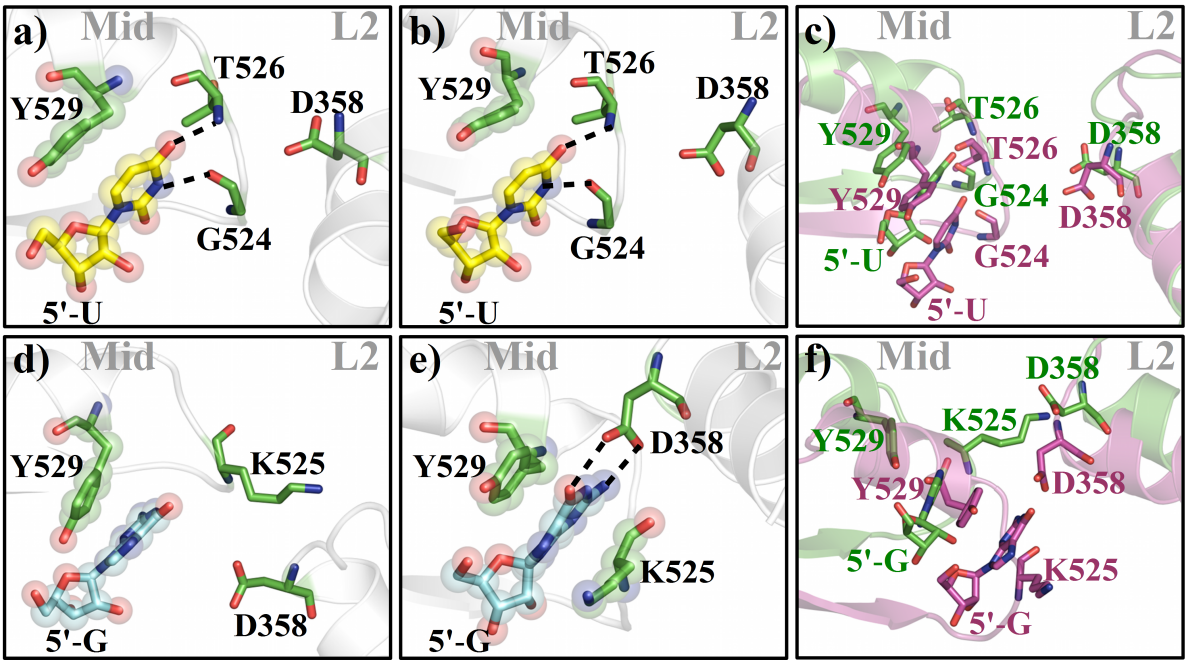

Figure 3-6: Close-up of the guide RNA 5'-end and relevant protein residues of the Mid domain and L2 linker region. The Mid domain and the L2 helix are represented in cartoon (grey), important protein residues (green) and 5'-U (yellow) or 5'-G (blue) are shown in sticks. For clarity merely the 5'-base is shown. Hydrogen bonds are represented by black dotted lines. Transparent spheres highlight stacking interactions between protein residues and the corresponding bases. (a, d) show the initial situation and (b, e) after 20 ns of simulation of 5'-U and 5'-G terminated hAgo2 bound guide strands, respectively. (c, f) superposition of (a, b) and (d, e) with neighboring protein residues at t=0 ns are shown in green and at t=20 ns in magenta.

In addition to this hydrogen bond interaction, perfect alignment of the aromatic ring of the 5'-G with K525 sidechain occurs, leading to a cation- π interaction. This cation- π interaction complements the base stacking that already exists between the aromatic ring of the 5'-G and Y529 sidechain (Figure 3-6 (e)). Y529 represents a highly conserved residue which has been shown to play a critical role in guide RNA binding and if phosphorylated drastically reduces such interactions (195). Therefore, this novel cation- π interaction in addition to the stacking interaction between Y529 and 5'-G pyrimidine ring could have potential biological consequences.

3.3. The biological role of D358 residue in hAgo2

3.3.1. MD simulations of hAgo2-guide RNA complex

The MD simulations of hAgo2-guide RNA complex revealed a salt bridge formation between the D358 residue present in the helix 7 of the L2 linker domain and the K525 sidechain present in the NS loop of the Mid domain (Figure 3-7). As the D358 protein residue interacts so intimately with the NS loop, which harbours the 5'-end of the guide RNA (Figure 3-8), it was anticipated to have a significant biological role. Therefore, *Sarah Willkomm a co-worker at the group of Prof. Tobias Restle performed in vitro studies of a recombinant D358A-hAgo2 to establish the biological significance of the D358 protein residue*.

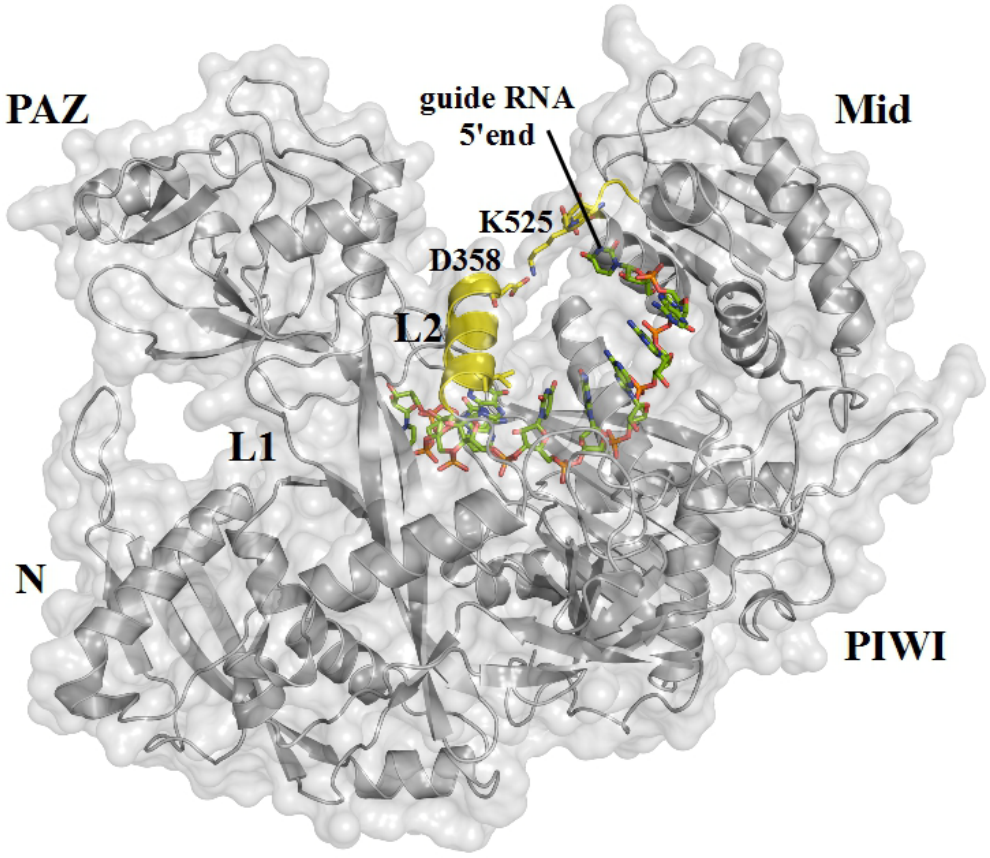

Figure 3-7: Structural organization of hAgo2-guide RNA complex represented by cartoons (grey). The α helix7 of the L2 linker domain and the NS loop of the Mid domain are highlighted in yellow. The D358 and K525 residues are represented by sticks (yellow). The guide RNA is shown in green sticks.

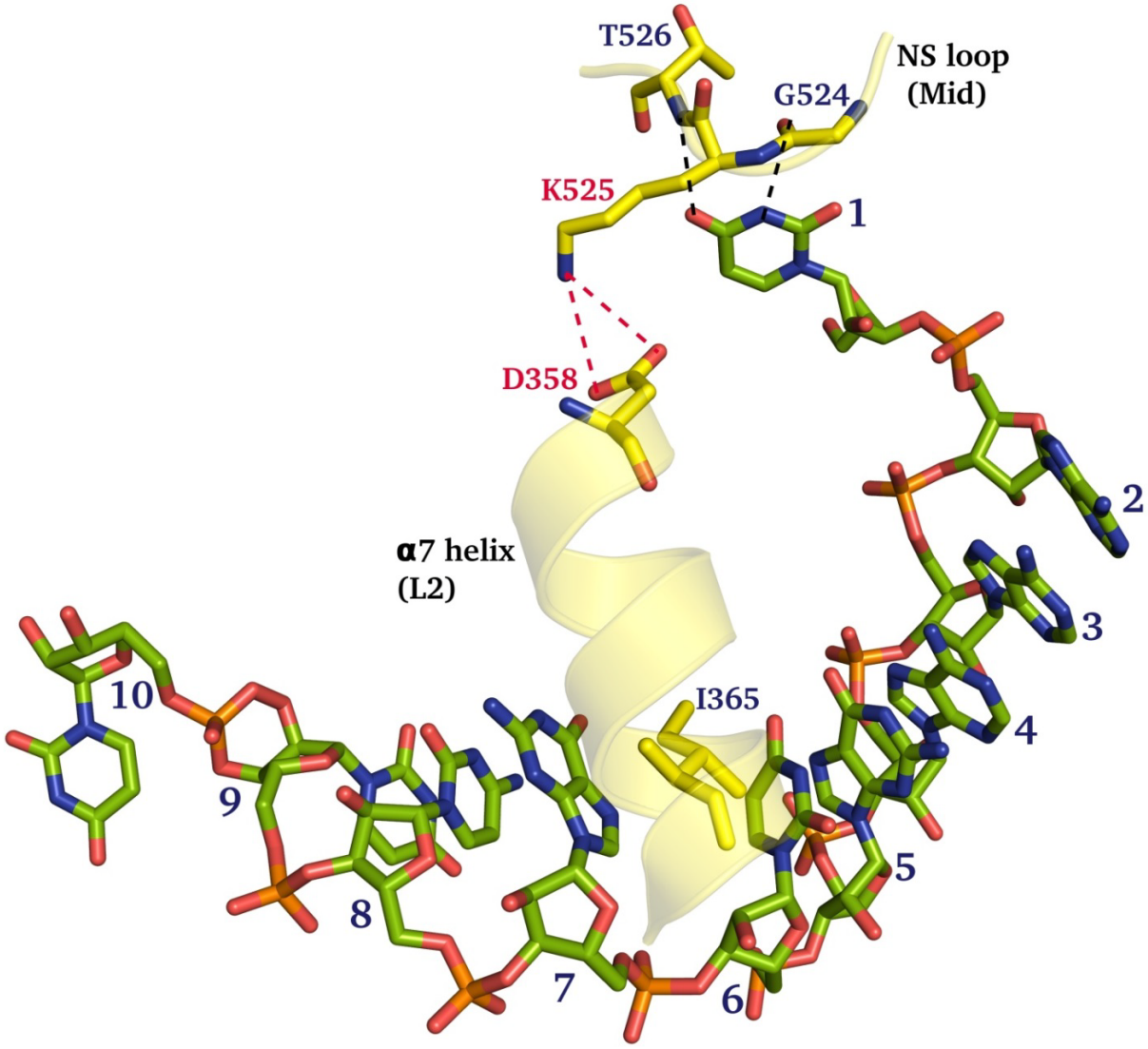

Figure 3-8: Close view of the L2-Mid salt bridge interaction. Helix7 and the NS loop are represented by cartoon (yellow), the interacting residues of the salt bridge and the NS loop are represented by sticks (yellow). Red dotted lines represent the salt bridge interaction between D358 and K525. Black dotted lines represent hydrogen bond interactions between 5'-U of the guide RNA and NS loop.

3.3.2. Biochemical characterization of the D358A- hAgo2 mutant

Biochemical experiments were performed by Sarah Willkomm. To analyze the effect of the D358A mutation on RNA binding under pre-steady state conditions stopped-flow experiments using an APP SX20 device were performed. On the basis of a minimal mechanistic model of siRNA-dependent hAgo2-mediated target RNA slicing developed earlier by the Restle group (19), data obtained with D358A-hAgo2 were compared to wt-hAgo2. To assemble binary

complexes, a 21-mer synthetic guide RNA with a FAM-label at position 14 was rapidly mixed with D358A-hAgo2 and the change in fluorescence was observed over time (Figure 3-9 (a)). Binding data are summarized in Table 3-3.

Next, cleavage assays were performed to corroborate the finding that the D358A-hAgo2 does not bind the guide RNA effectively, which prevents the formation of a functional ternary complex (Figure 3-10). wt-hAgo2 and guide RNA were preassembled in cleavage buffer and the reaction was started by adding radio-labeled target RNA (21-mer) to monitor target cleavage. A cleavage product could not be observed in the case of D358A-hAgo2. It suggests that the guide RNA does not bind properly to D358A-hAgo2, precluding the formation of functional ternary complexes.

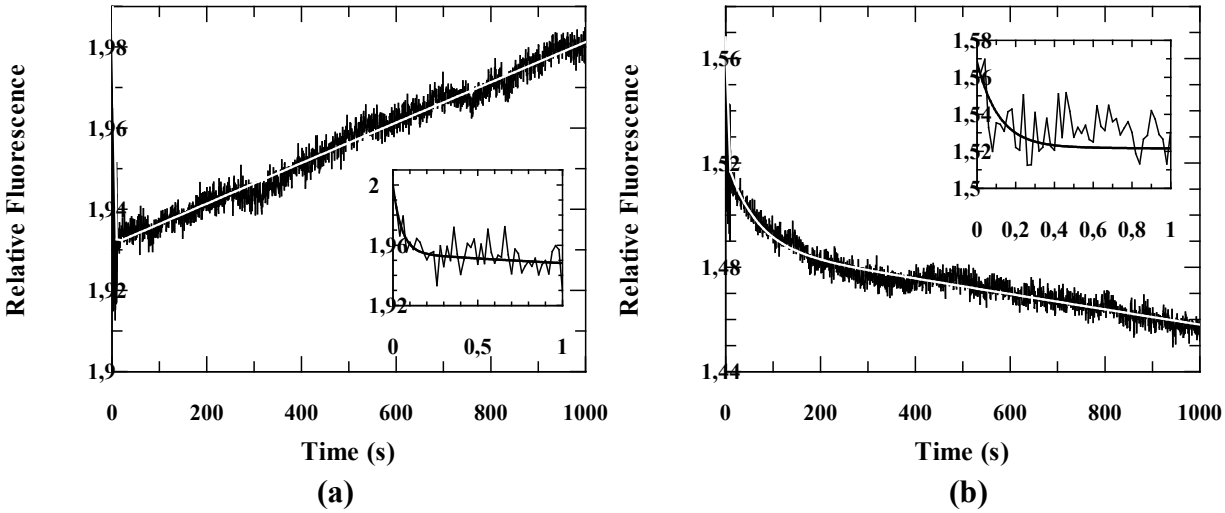

Figure 3-9: Pre-steady state kinetics of binary complex formation and dissociation with D358A-hAgo2. Representative stopped flow graphs are shown. The inserts show the reaction on a shorter time scale. (a) To analyze association of guide RNA and D358A-hAgo2, 20 nM of FAM-labeled guide RNA were rapidly mixed with 500 nM D358A-hAgo2. (b) Preformed binary complexes composed of 20 nM FAM-labeled guide RNA and 500 nM D358A-hAgo2 were rapidly mixed with 2 µM unlabeled guide RNA. In both cases data were fitted to an exponential equation with two terms yielding the following rate constants: k_1: 19.2 (± 2.1) s^{-1}; k_{-1}: 8.8 (± 0.9) s^{-1}; k_2: 0.34 (± 0.02) s^{-1}; k_{-2}: 0.02 (± 0.0007) s^{-1}. *Figure courtesy of Sarah Willkomm.*

Table 3-3: Summary of pre-steady state binding data for binary complex formation with D358A-hAgo2. *Table courtesy Sarah Willkomm.*

Protein	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})
wt-hAgo2 (19)	0.6×10^8	6.2	0.26	0.17	0.012	0.007
D358A-hAgo2	0.26×10^8	8.8	0.42	0.02	-	-

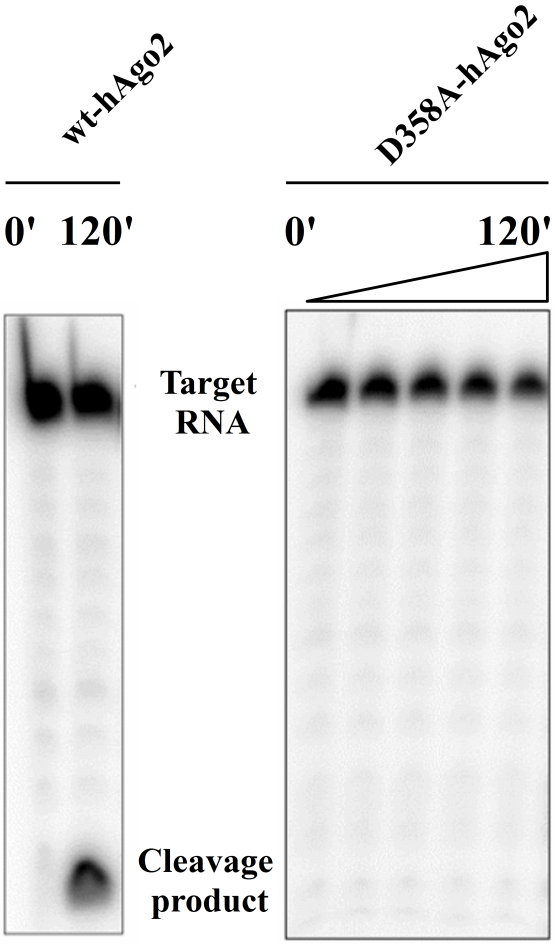

Figure 3-10: Cleavage activity of the wt-hAgo2 and the D358A-hAgo2. Cleavage assays were conducted using 2.5 µM hAgo2, 100 nM ss-siRNA and radio-labeled target RNA. Samples were collected at 0', 10', 30', 60' and 120'. Reactions were analyzed by 20% (w/v) PAGE and visualized using autoradiography. *Figure courtesy of Sarah Willkomm*.

3.3.3. MD simulations explain the change in binding pattern and loss of cleavage/slicer activity of hAgo2

During the simulations of the hAgo2-guide RNA complex, it was observed that the D358 protein residue forms a salt bridge with K525. This hinted that the D358 protein residue might have a significant biological role, which, indeed was shown by the *in vitro* studies performed on the D358A-hAgo2. It was observed that D358 forms the salt bridge with K525 early on during simulation, which was retained for almost ~60% of the 100 ns simulation performed. Interestingly, in either of the reported crystal structures (4F3T, 4EI1) D358 does not interact with K525 sidechain, although, there seems to be a propensity that D358 might interact with K525. On a closer analysis of the electron density of the reported crystal structure 4F3T, it was gathered that the D358 sidechain has a clear electron density; however, K525 sidechain seems to have rather very low electron density (Figure 3-11). It further explains why this interaction could not be observed in the crystal structures, however, it was very promptly observed during the MD simulations of the hAgo2-guide RNA complex.

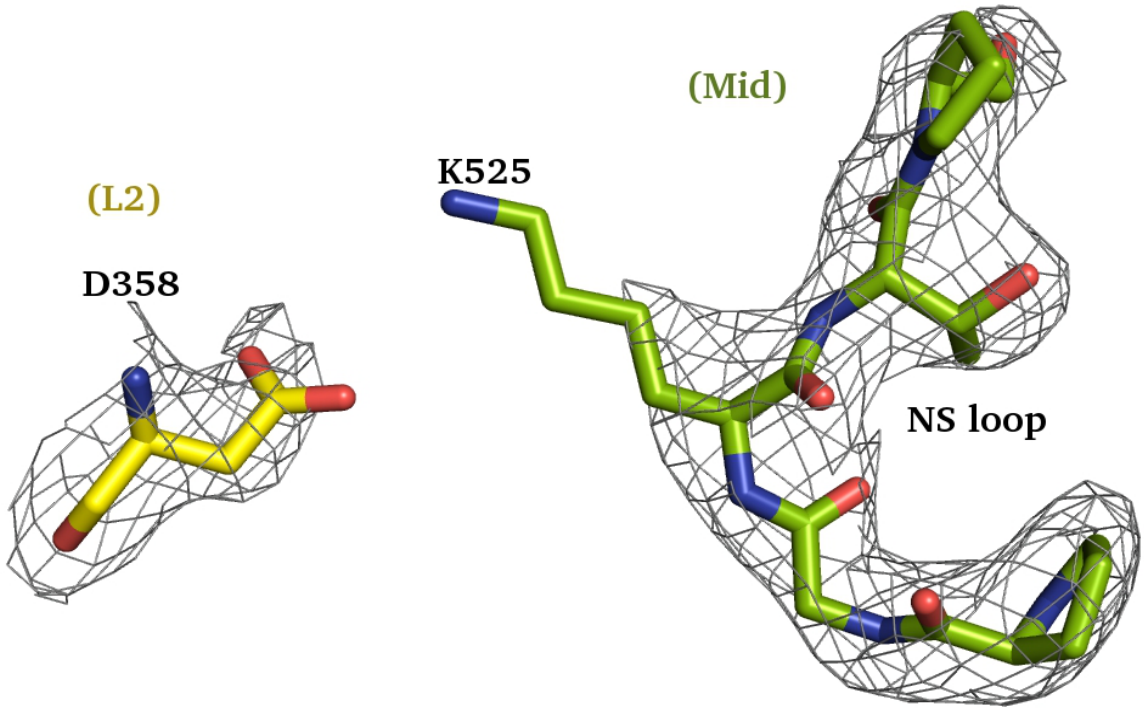

Figure 3-11: Electron density map of the D358 protein residue (yellow) which is present in the L2 domain and protein residues present in the NS loop (green) of the Mid domain, at sigma value 2.

The studies illustrate that the overall fold of the wt-hAgo2 and its individual domains is conserved during the timescale of the simulations. The wt-hAgo2 backbone exhibits RMSD values of ~3-4 Å compared to the backbone of the starting structure at t = 0ns of the simulations. The major changes were observed in the N and PAZ domains, the average RMSD of the backbone atoms of these domains was 2.7 and 1.4 Å, respectively, when the RMSD was calculated by aligning the individual domains against themselves (Table 3-4). However, when the individual domain backbones were aligned with the entire protein backbone, the average RMSD of the N and PAZ domains was observed to be 3.8 and 4.2 Å, respectively. This demonstrates that the change in RMSD is contributed by the displacement of the N and PAZ domains from their initial positions during the course of simulations. In comparison, the Mid and PIWI domains did not have major RMSD changes (Table 3-4).

Table 3-4: RMSD analysis of the wt-hAgo2. RMSD in the first column is calculated by fitting the backbone of individual domains with themselves at t=0ns. RMSD in the second column is calculated by fitting the individual domains with the entire protein backbone at t=0ns. The average and standard deviation of the RMSD measured over the backbone atoms are reported.

Domain	Domain fit domain (RMSD Å)	Domain fit protein (RMSD Å)
N	2.7 ± 0.5	3.8 ± 0.8
PAZ	1.4 ± 0.2	4.2 ± 1.5
Mid	1.6 ± 0.2	3.0 ± 0.7
PIWI	2.0 ± 0.2	2.8 ± 0.4

In addition to the L2-Mid interaction, a subtle breathing motion was also observed in wt-hAgo2 during the MD simulations. Further, PCA of the wt-hAgo2-guide RNA simulation was performed to effectively identify the significant domain motions in the wt-hAgo2 backbone. The first principal component (PC1) of the wt-hAgo2 backbone shows that this subtle breathing motion is caused by an inward motion of PAZ and N towards Mid, alternated by outward motion of PAZ and N away from Mid (Figure 3-12 (a)). The second principal component (PC2) of the wt-hAgo2 backbone shows that the second phase of the breathing motion involves the movement of the PAZ and Mid domains away from each other (Figure 3-12 (b)).

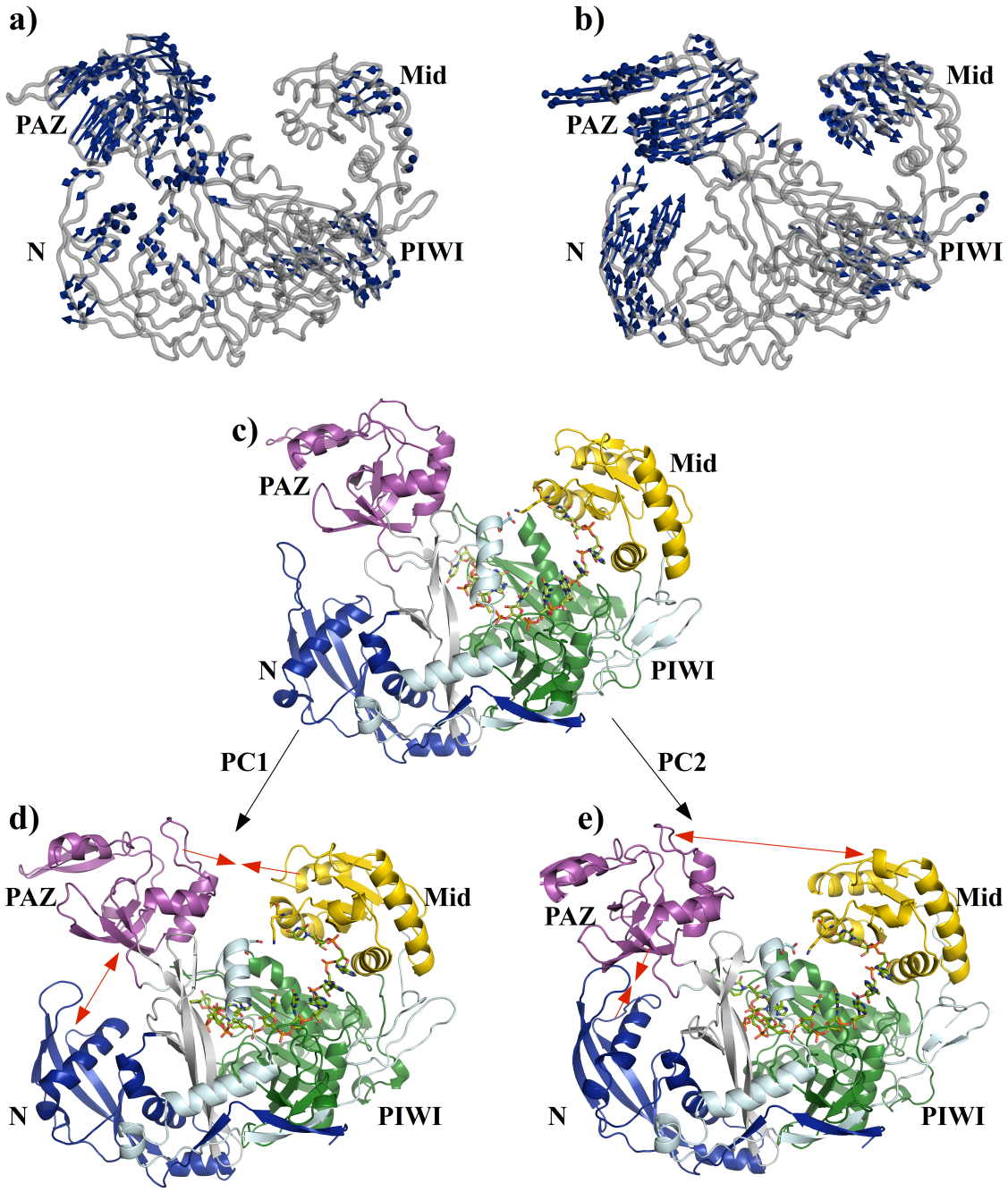

Figure 3-12: Illustration of the breathing motion observed in the wt-hAgo2. a) and b) correspond to the PC1 and PC2 obtained from the simulations of wt-hAgo2 protein backbone (grey cartoons), respectively. Blue arrows represent the direction and magnitude of the domain motion c) The wt-hAgo2 structure at t=0 ns of the simulations. d) and e) correspond to the structures representing the first and second breathing motions of wt-hAgo2 (cartoon). Each domain is color-coded; N (blue), L1 (silver), PAZ (magenta), L2 (pale cyan), Mid (yellow), PIWI (green).

To understand the effect the D358A mutation may have on the positioning of the guide RNA, MD simulations with D358A-hAgo2 were performed. It was observed during the MD simulations that the overall fold of the D358A-hAgo2 domains was retained; the RMSD of the protein backbone changes ~2 - 5 Å. The most pronounced difference was observed for PAZ domain. The average RMSD of the PAZ domain backbone, obtained by fitting it with the PAZ domain of the structure at t=0ns was 2.3 Å (Table 3-5). When the average RMSD of the PAZ domain backbone was calculated by aligning it against the entire protein backbone, a huge difference of 6 Å was observed (Table 3-5). This illustrates that the PAZ domain displaces immensely from its initial position during the simulations. A considerable difference is also noted for the N and Mid domains, when the average RMSD is obtained by aligning them respectively against the entire D358A-hAgo2 backbone (Table 3-5).

Table 3-5: RMSD analysis of the D358A-hAgo2. RMSD in the first column is calculated by fitting the backbone of individual domains with themselves at t=0ns. RMSD in the second column is calculated by fitting the individual domains with the entire protein backbone at t=0ns. The average and standard deviation of the RMSD measured over the backbone atoms are reported.

Domain	Domain fit domain (RMSD Å)	Domain fit protein (RMSD Å)
N	2.8 ± 0.7	4.0 ± 1.5
PAZ	2.3 ± 0.4	6.0 ± 1.9
Mid	1.5 ± 0.2	5.1 ± 1.5
PIWI	1.8 ± 0.2	2.6 ± 0.3

The PCA suggests that the PAZ and Mid domains move towards each other, whilst N, L1 and L2 move away from PAZ and Mid in a synchronous manner (Figure 3-13). It was observed that the subtle breathing motion of wt-hAgo2, was replaced by a prominent bending motion caused by movement of PAZ and Mid towards each other. The PC1 of the D358A-hAgo2 simulation, alone contributes ~93% of the variance. It reveals prominent domain motions; PAZ and Mid move inwards towards each other, whilst N, L1 and L2 move away from PAZ and Mid in a synchronous manner (Figure 3-13 (a)). The PC2, which contributes ~4% variance, shows that PAZ, Mid and N moves outwards in a synchronous manner (Figure 3-13 (b)).

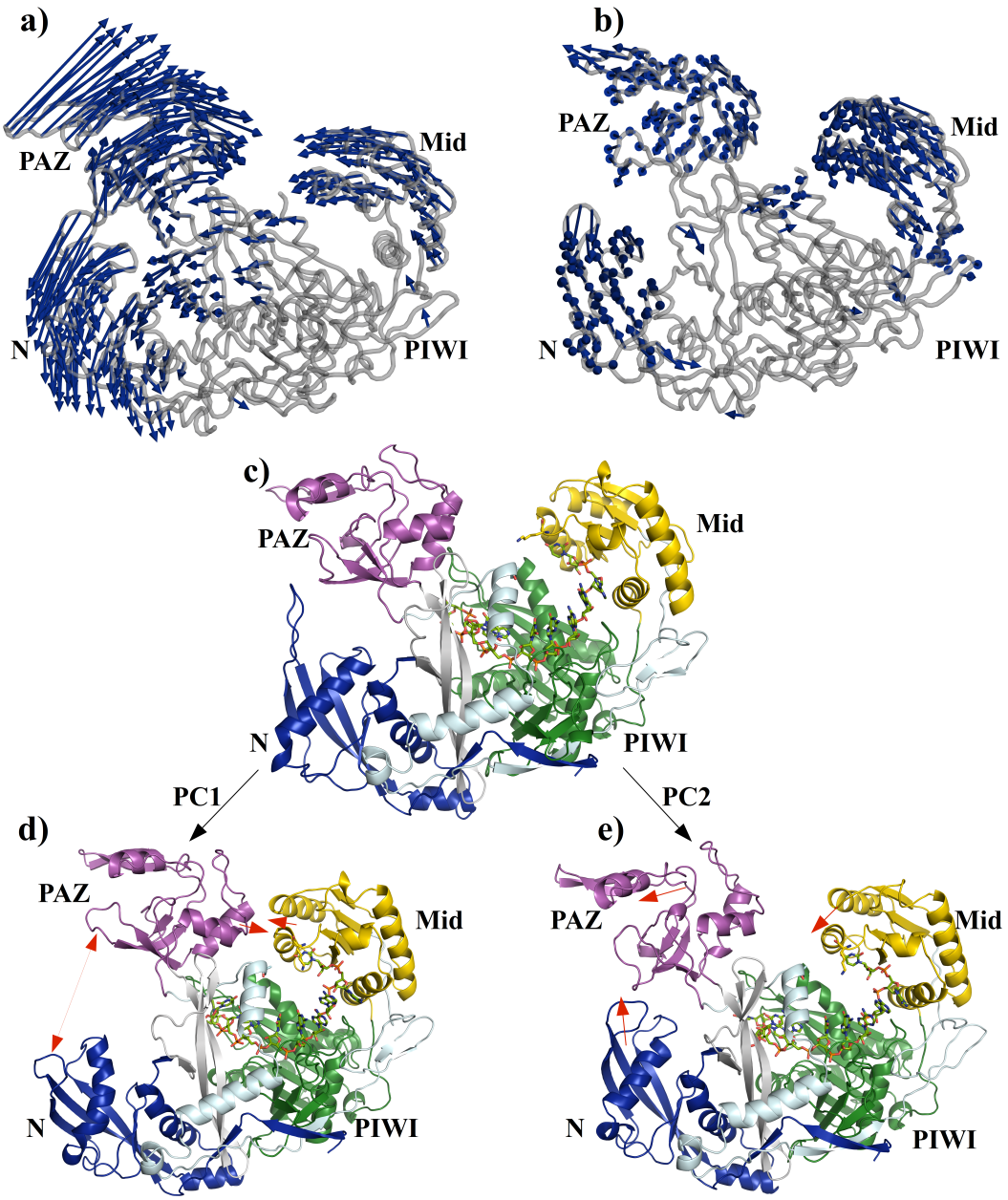

Figure 3-13: Illustration of the bending motion observed in D358A-hAgo2. a) PC1obtained from the simulations of the D358A-hAgo2 backbone (grey cartoons), clearly illustrates the concerted motion of the PAZ and Mid domains towards each other and the motion of the N domain away from the PAZ domain. The blue arrows represent the direction and magnitude of the domain motion. b) PC2 obtained from the simulations of D358A-hAgo2 backbone (grey) illustrates the motion of PAZ and Mid domains away from each other. c) Structure of the D358A-hAgo2 at t=0 ns of the simulations. d) Structure represents the bending motion of PAZ and Mid domains in the D358A-hAgo2 (cartoon), each domain is color-coded; N (blue), L1 (silver), PAZ (magenta), L2 (pale cyan), Mid (yellow), PIWI (green). e) Structure represents the movement of PAZ and Mid away from each other. The red arrow shows the extent of the bending motion.

Next, the first two principal components from each different simulation were submitted to the Dyndom package for domain motion analysis (177, 196). The Dyndom identifies the part of the protein, which is mobile called as the 'moving domain'. It also determines the inter-domain screw axis an 'effective hinge axis' and the residues involved in the bending motion of this hinge. The results of the Dyndom analysis are shown in Table 3-6; in both cases the PAZ domain is identified as the 'moving domain'. It also illustrates that the rotation angle of the PAZ domain in D358A-hAgo2 is almost doubled. Interestingly, the D358 residue is part of the bending residues, which form the L1/L2 hinge region (Table 3-6), both in the wt-hAgo2 and D358-hAgo2 simulations, clearly illustrating that D358 plays a vital role in the bending of the PAZ domain.

Table 3-6: The analysis of the wt-hAgo2 and D358A-hAgo2 domain motion with the Dyndom package. The PAZ domain is identified as the most flexible domain; its rotation is almost doubled in case of D358A-hAgo2.

Property	wt - hAgo2	D358A-hAgo2
Moving domain	PAZ	PAZ
Rotation Angle (deg)	17.8	30.3
Translation (A)	-1.5	-2.3
Closure (%)	99.1	29.3
Bending Residues	221 – 229 356 – 367	222 – 224 352 – 368

It was observed that the D358A single point mutation considerably increases the flexibility of the PAZ domain in hAgo2. A comparison of the RMSF of the C_α atoms in wt-hAgo2 and D358A-hAgo2 shows the flexibility increases in the PAZ domain from ~3.5 to ~8 Å due to the D358A single point mutation (Figure 3-14).

The effect of the D358A mutation was explored on the position of the guide RNA. In the D358-hAgo2, the NS loop becomes more flexible, which is demonstrated by its increased RMSF as seen in the Figure 3-15 (a). It was also observed that the D358A mutation affects the relative

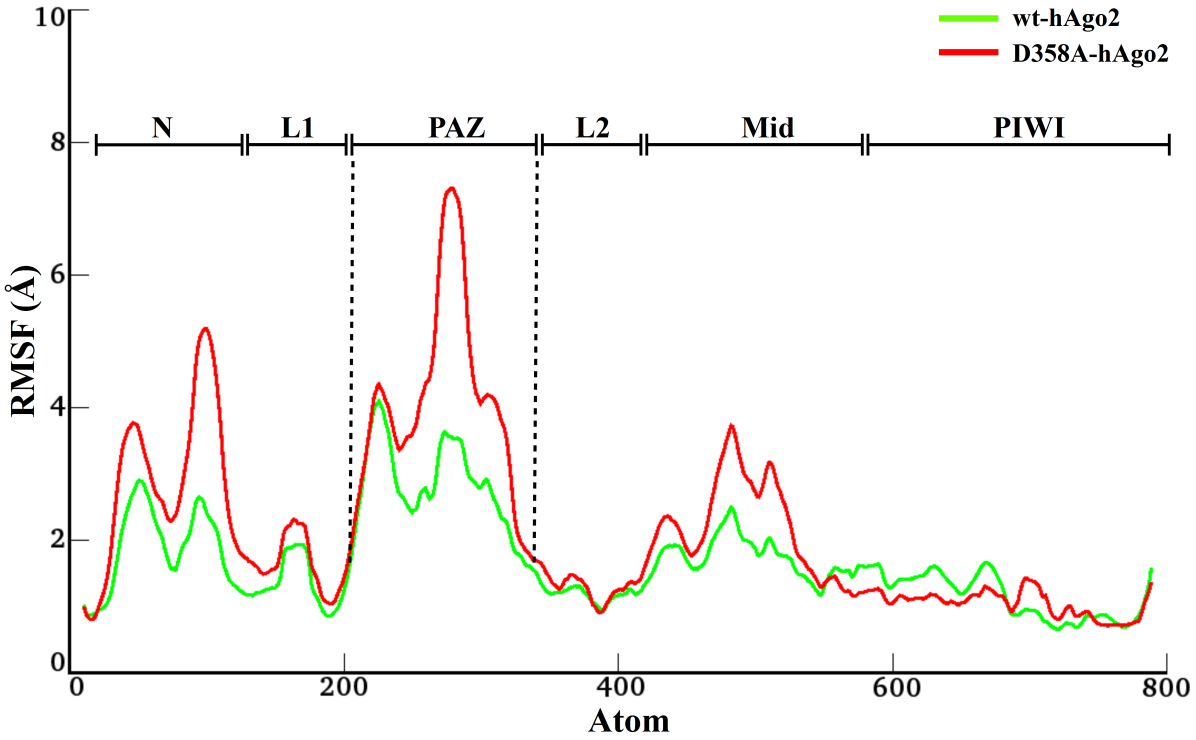

Figure 3-14: RMSF comparison of the protein backbone in case of the wt-hAgo2 (green) and D358A-hAgo2 (red). Black dotted lines highlight the greatest difference in the RMSF of PAZ domain.

positioning of the guide RNA. The RMSD of guide RNA bound to the Mid in the D358-hAgo2 shows a huge difference in comparison to the wt-hAgo2 (Figure 3-15 (b)). It makes it tempting to speculate that although the 5'-end of the guide RNA remains tightly bound to the Mid domain, the relative positioning of the entire guide is altered.

3.4. The role of the I365 residue in hAgo2

3.4.1. I365A mutation effects the flexibility of guide RNA

The L2 linker domain is the bridge interconnecting the PAZ and Mid domains. In the recently published hAgo2 structures (4F3T.pdb, 4OLA.pdb), it was observed that a prominent destacking occurs between the nucleotides at position 6 and 7. This destacking or the kink occurs due to the protruding I365 sidechain. This I365 sidechain also has a stacking interaction with the aromatic ring of 7G. Thus far, no biological significance has been associated with this destacking or this interaction between the I365 sidechain and base of 7G. The location of I365 is very noteworthy;

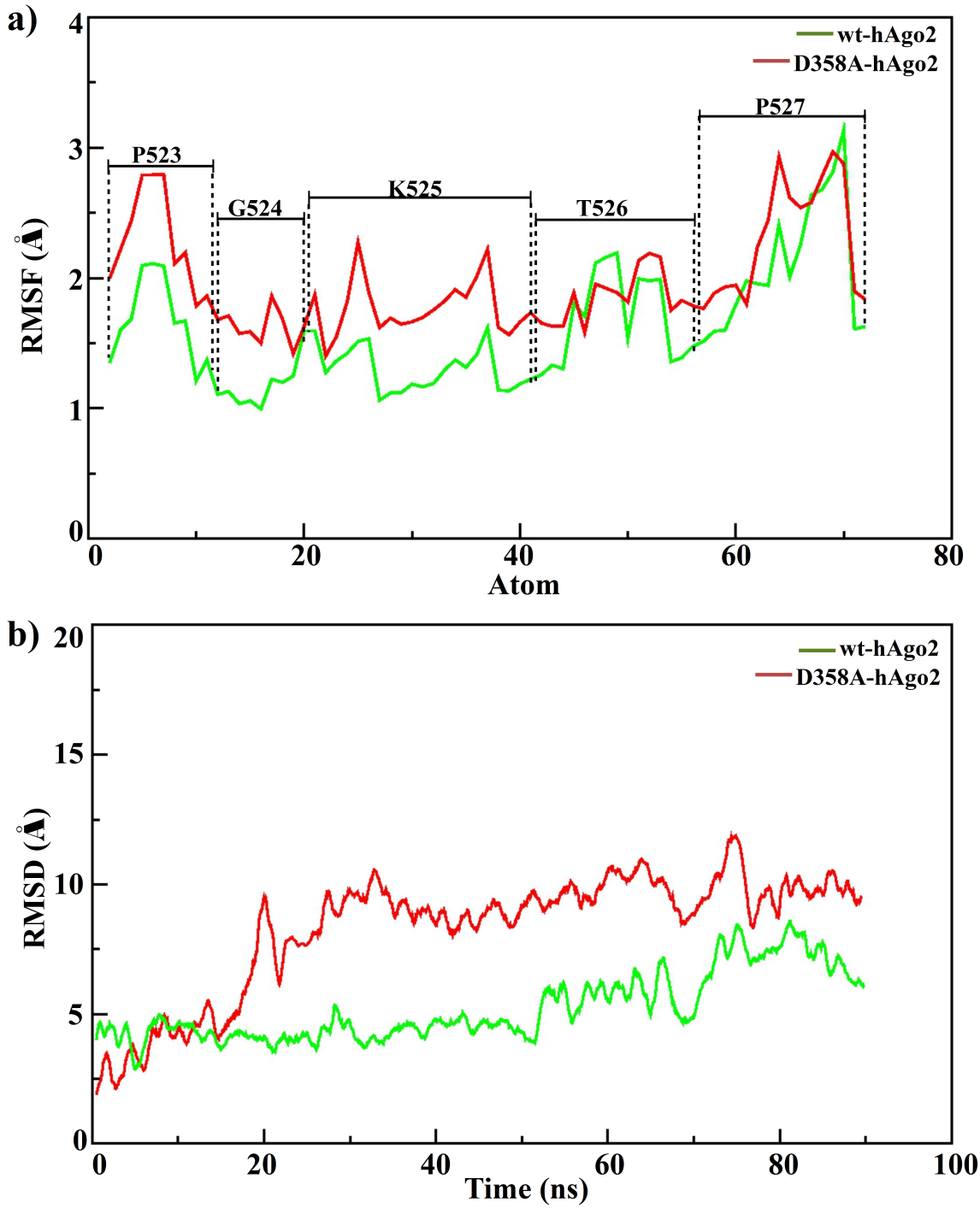

Figure 3-15: a) RMSF comparison between the residues present in the NS loop of the Mid domain present in the wt-hAgo2 (green) and D358A-hAgo2 (red). Black dotted lines separate the five residues present in the NS loop. b) RMSD comparison of the guide RNA bound to the wt-hAgo2 (green) and D358A-hAgo2 (red), calculated by fitting all structures during the simulations to the Mid domain of the initial structure at t=0 ns. The RMSD signifies the difference in the relative positioning of the guide RNA bound to the wt-hAgo2 and D358A-hAgo2.

it sits at the bottom of helix7. This helix 7 also hosts the D358 residue that forms a salt bridge with the Mid domain. Due to the unique and crucial location of I365, it was expected that it would play a significant role in pinning down the guide RNA inside the nucleic acid binding channel.

To validate this hypothesis, MD simulations of hAgo2-guide RNA complex were performed, in which I365 was mutated to alanine. It was observed that the I365A mutation in the hAgo2-guide RNA complex affected the flexibility of the guide RNA bound to hAgo2. The overall flexibility of the guide RNA increased from ~1.5 Å to ~3 Å. The RMSF comparison of the guide RNA bound to the wt-hAgo2 and I365A-hAgo2 demonstrates that the 5'-end of the guide RNA has increased mobility. The most noticeable difference was observed for the nucleotides at position 6 and 7 (Figure 3-16). Although the nucleotides at position 9 and 10 of the guide RNA bound to I365A-hAgo2 have higher RMSF in comparison to the wt-hAgo2, it can be attributed to the lack of interactions with neighbouring protein residues.

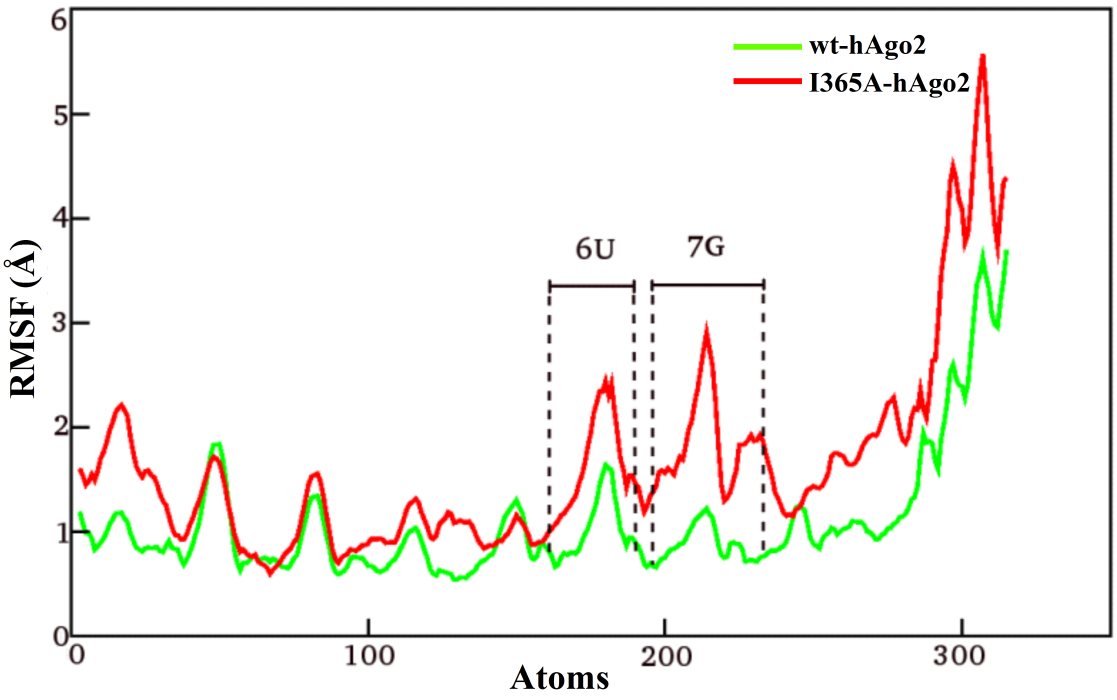

Figure 3-16: RMSF comparison of the guide RNA bound to the wt-hAgo2 (green) and I365A-hAgo2 (red). Black dotted lines highlight the greatest difference in the RMSF of nucleotides 6 and 7.

The I365A mutation also affects the destacking of the guide RNA between nucleotides at positions 6 and 7. Around ~ 50 ns into the simulation, a conformational movement was observed in the base 7G, the aromatic ring of the 7G moves and arranges itself parallel to the aromatic ring of the 6U base. This movement abolishes the prominent destacking between 6U and 7G and creates a perfectly stacked orientation.

This observation is further corroborated by the distance calculated between the center of mass of the 6U and 7G bases. It can clearly be seen in the Figure 3-17 that the distance between the bases 6U and 7G (colored in red) in case of I356A-hAgo2 reduces considerably from ~7 Å to ~4 Å due to the I365A mutation. Whereas, the distance remains constant between the 6U and 7G bases of the guide RNA bound to the wt-hAgo2.

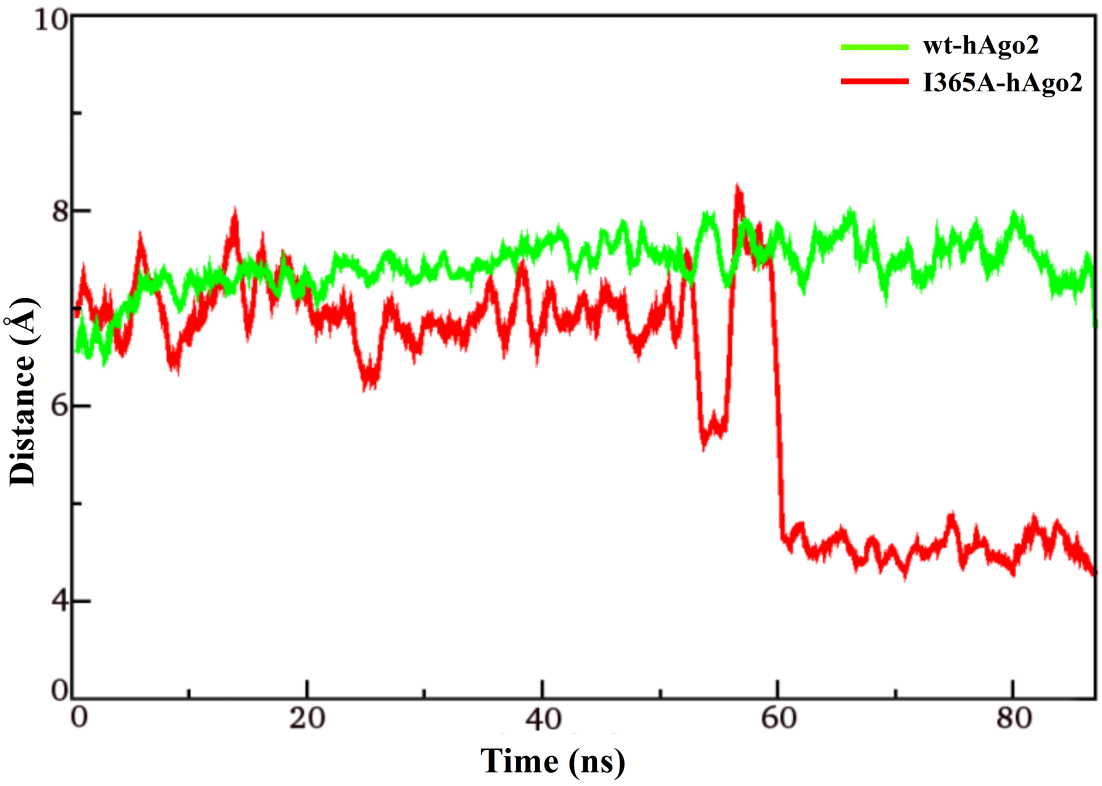

Figure 3-17: Plot of distance between the 6U and 7G bases present in the guide RNA's bound to the wt-hAgo2 (green) and I365A-hAgo2 (red).

3.4.2. Effect of the I365A mutation on the binding of guide RNA

Although the guide RNA is bound to the Mid binding pocket by its 5'-end, the L2 domain intervenes in the course of the guide RNA in the nucleic acid binding channel by introducing a prominent kink in the guide RNA. *Here further in vitro studies with a recombinant I365A-hAgo2 were performed by Sarah Willkomm.* It was observed that in the I365A-hAgo2 the second and third phase during association and dissociation of the binary complex were slowed down (Table 3-7). The most profound effect was observed in the second phase of the dissociation, which was reduced by a factor of >10. Furthermore, cleavage efficiency was decreased by up to 75% in case of I365A-hAgo2.

This suggests that the I365A mutation considerably increases the flexibility of guide RNA bound to the hAgo2. Through the MD simulations, it was observed that the I365A mutation further influences the relative positioning of the guide RNA relative to the active site. As seen in Figure 3-18, the RMSD of the guide RNA when fit to the PIWI domain, is considerably higher in case of the I365A-hAgo2 (red), than the wt-hAgo2 (green). This clearly illustrates the effect of I365A mutation on the guide RNA positioning relative to the PIWI domain.

Table 3-7: Summary of pre-steady state binding data for binary complex formation with the wt-hAgo2 and the I365A-hAgo2. *Table courtesy of Sarah Willkomm*

Protein	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})
wt-hAgo2 (19)	0.06	6.2	0.26	0.17	0.012	0.007
I356A-hAgo2	0.03	20.4	0.05	0.01	0.006	0.002

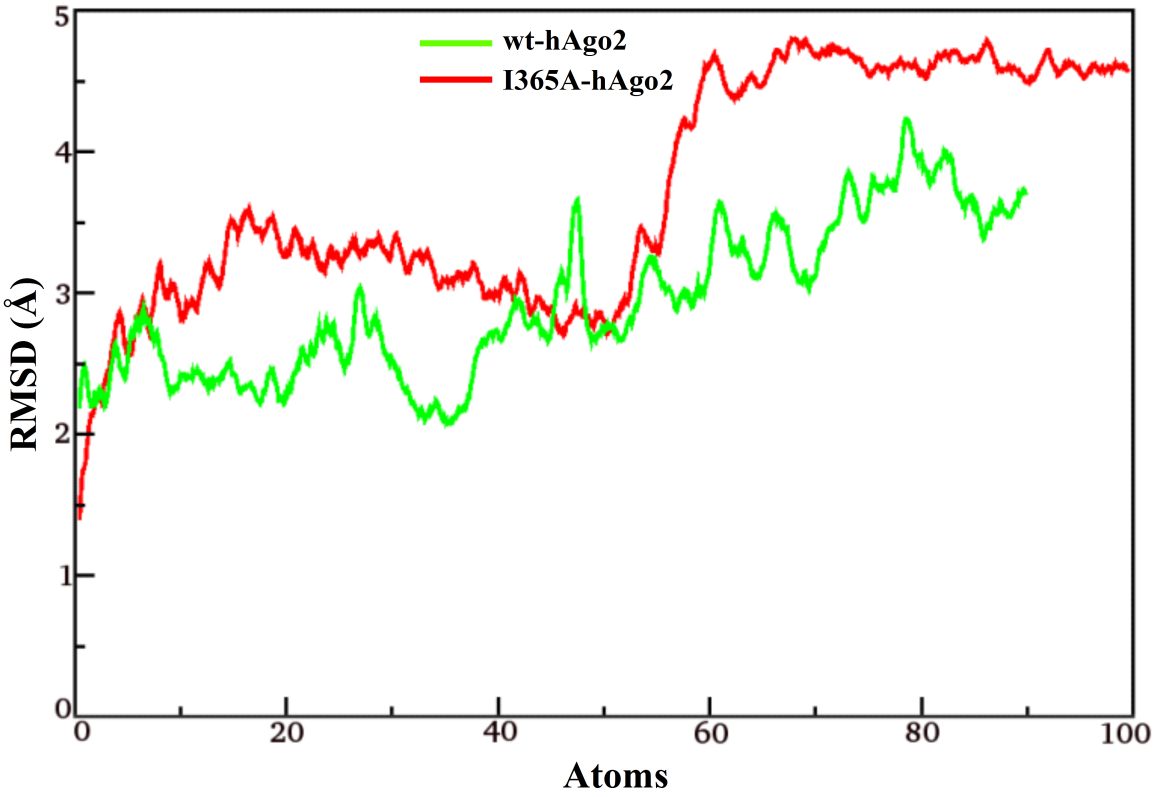

Figure 3-18: RMSD of guide RNA bound to the wt-hAgo2 (green) and I365A-hAgo2 (red). The RMSD is calculated by fitting all the structures during the simulations to the PIWI domain of the structure at the beginning of the simulations.

3.4.3. Influence of the I365A mutation on hAgo2 flexibility

Interestingly, the I365A mutation does not seem to affect the overall flexibility of the protein itself. The RMSF of the protein backbone of the wt-hAgo2 and I365A-hAgo2 suggests that the flexibility of individual domains and the linker domains remain unaffected by the I365A mutation. As observed in the Figure 3-19 the RMS fluctuations of the protein backbone of I365A-hAgo2 (red), is quite similar to that of the wt-hAgo2 (green). Therefore, this suggests that the I365A mutation exclusively affects the guide RNA bound to hAgo2, however does not affect the protein itself.

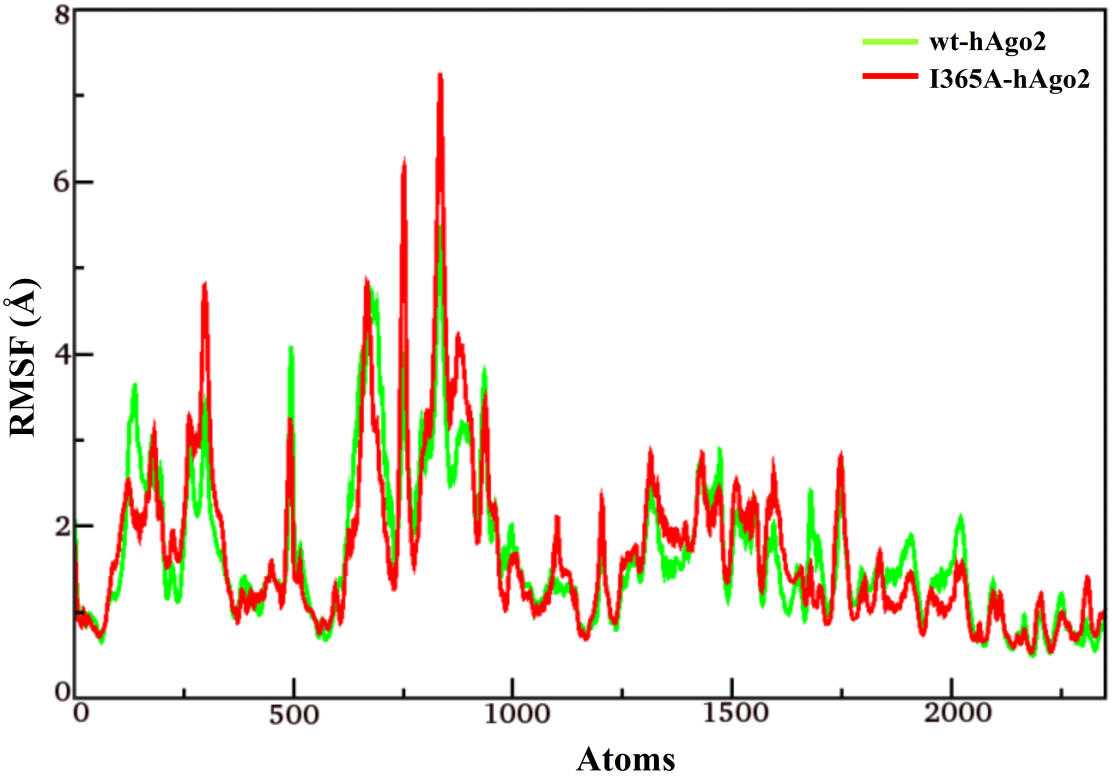

Figure 3-19: RMSF comparison between the backbone of the wt-hAgo2 (green) and I365A-hAgo2 (red).

3.5. The role of D356 in hAgo1

3.5.1. Overall dynamics of hAgo1

The recently reported crystal structures of hAgo1 provided crucial insights into its structural organization. The sequence identity between hAgo1 and hAgo2 is 88%; moreover, the L2 linker domain is highly conserved. Following the lead of previous observations made on hAgo2, the role of the L2 linker domain in hAgo1 was investigated. Long timescale (~100ns) MD simulations of a hAgo1-guide RNA complex (4KXT.pdb) were performed with the Gromacs simulation package.

As observed during the MD simulations of hAgo2, hAgo1 in complex with a bound guide RNA is also flexible. The overall domain motion of hAgo1 is comparable to that of hAgo2. The PAZ domain fluctuates a lot about its initial position and moves away from the Mid domain towards

the N domain. The N domain also moves outwards from the PIWI domain and towards the PAZ domain. This synchronous motion of the PAZ and N domains towards each other widens the nucleic acid binding channel. This domain motion is also observed through the RMSD analysis of the MD simulations. The overall fold of hAgo1 is retained; however, the individual domains seem to undergo domain motion changes. Figure 3-20 (a) reveals that the PAZ domain deviates up to ~7.5 Å from its initial position. The RMSD of the N domain deviates ~6.5 Å from its initial position. It was observed that the RMSD of Mid and PIWI remain almost constant at ~2.5 Å.

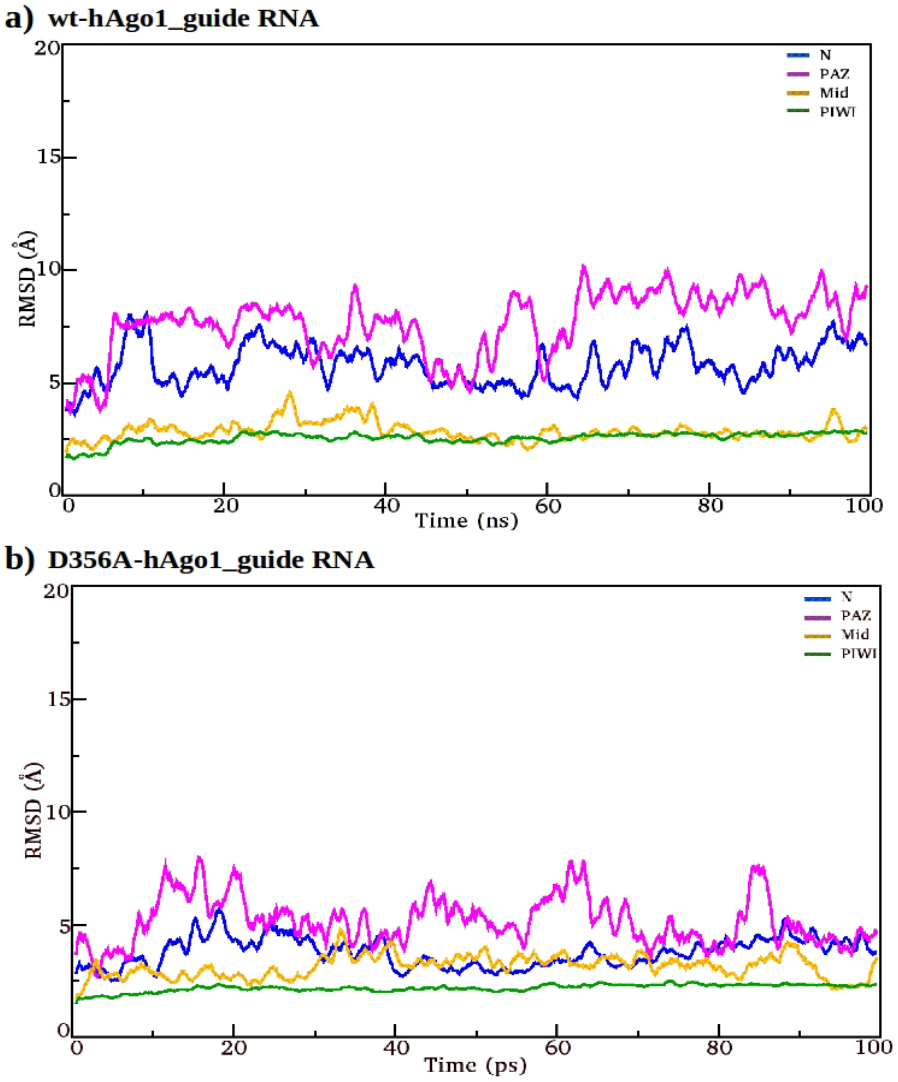

Figure 3-20: RMSD comparison of wt-hAgo1 and D356A-hAgo1 protein domain backbones. The individual domains are color-coded; N (blue), PAZ (magenta), Mid (yellow), PIWI (green). a) RMSD plot of the wt-hAgo1 individual protein domains b) RMSD plot of D356A-hAgo1 individual protein domains.

This observation is further corroborated by PCA of the trajectory. PC1 of hAgo1 (Figure 3-21 (a)), which represents 27% of the variance, shows a synchronous domain motion of the PAZ and Mid domains away from each other. In addition, it can also be observed that the vectors from the PAZ and N domains move towards each other, which is representative of a movement of the domain motion of the PAZ and N domains towards each other. The PC2 obtained from the

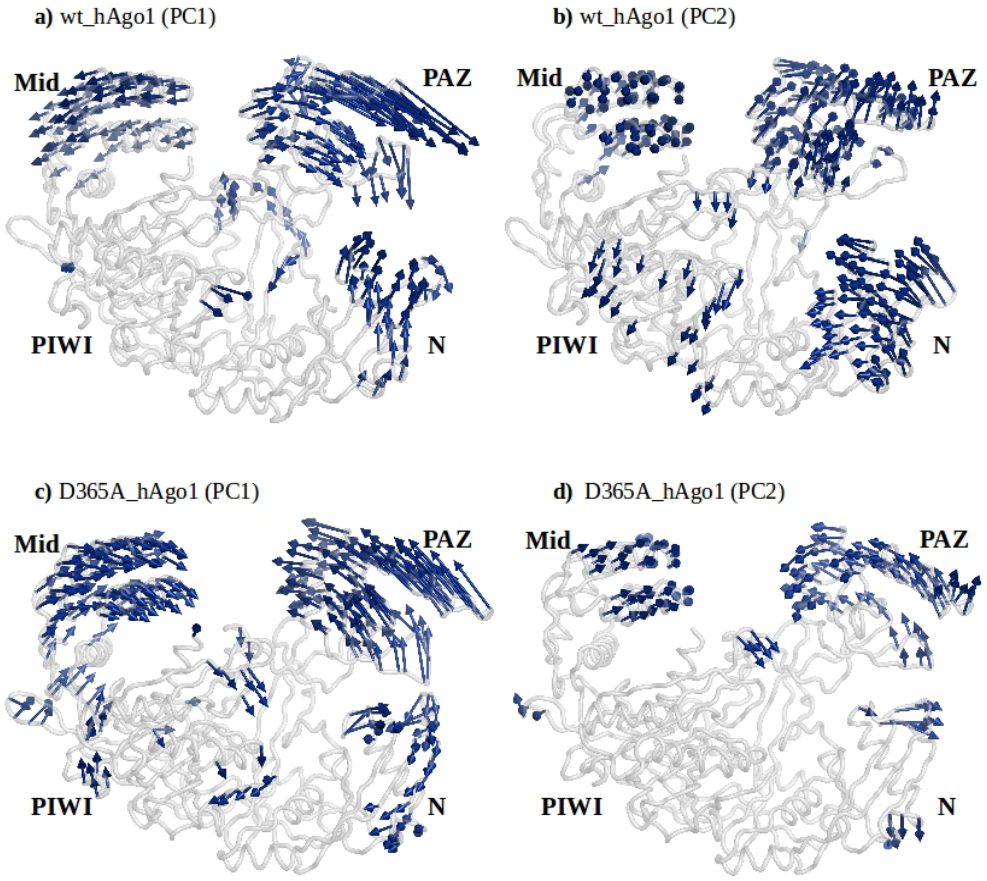

Figure 3-21: PCA of the trajectories obtained from MD simulations of wt-hAgo1 and D356A-hAgo1. Cartoons represent the protein backbone and blue arrows, pointing in the direction and magnitude of motion, represent the domain motion. a) PC1 of wt-hAgo1. The arrows present in PAZ and Mid domains point away from each other illustrating the motion of the PAZ and Mid domains away from each other b) PC2 of wt-hAgo1. The arrows present in PAZ and Mid domains point in opposite directions representing the domain motion of PAZ and Mid away from each other. c) PC1 of D356A-hAgo2. The arrows present in the PAZ and Mid domains point towards each other, illustrating the domain motion of PAZ and Mid towards each other. d) PC2 of D356A-hAgo1. The arrows present in PAZ and Mid in this case also point towards each other further corroborating that PAZ and Mid domains move towards each other due to the D356A mutation.

simulation of hAgo1 represents 21% of the variance. It demonstrates that the PAZ and Mid domains tend to move away from each other, while the PAZ and N domains tends to move towards each other. Collectively, these data point towards a flexibility of PAZ and N domains.

3.5.2. L2-Mid interaction in hAgo1

The MD simulations of hAgo1 in complex with guide RNA (4KXT.pdb) revealed a salt bridge interaction between the L2 and Mid domains. This salt bridge is formed between the D356 present in α helix7 of L2 and K523 present in the NS loop of the Mid domain (Figure 3-22). Interestingly this salt bridge occurs at the same position as in hAgo2. The salt bridge formation occurs almost instantaneously from the start of the MD production within the first nanosecond of the simulation. Once the salt bridge formation occurs, it is retained for almost ~54% of the time

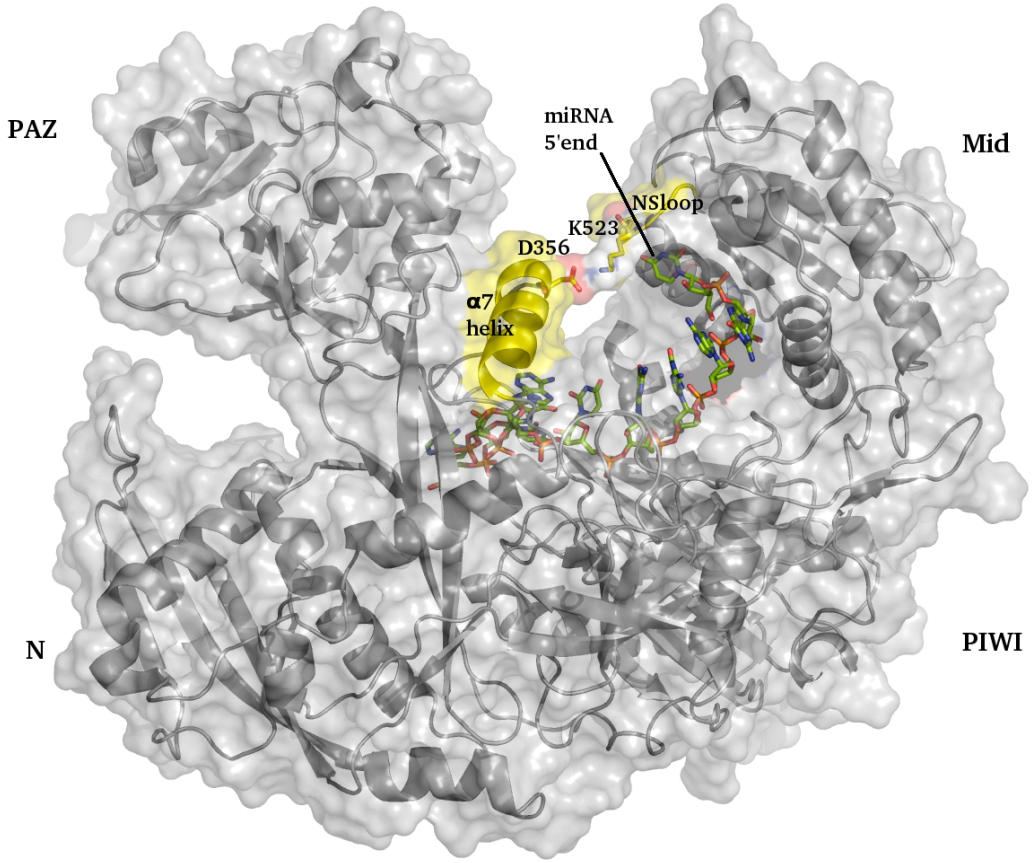

Figure 3-22: A snapshot from the MD simulation of hAgo1 (grey cartoon) in complex with guide RNA (green sticks) illustrating the formation of the L2-Mid salt bridge interaction at 1 ns. Highlighted in yellow is the salt bridge interaction between helix 7 present in the L2 linker domain and the NS loop present in the Mid domain. D356 and K523 protein residues are represented by yellow sticks.

of the simulation. It clearly suggests that this is a strong interaction.

The location of this salt bridge interaction is illustrated in Figure 3-22, helix 7 is colored in yellow and the D356 sidechain is represented in sticks, a close up is shown in Figure 3-23. It is however important to note that, this salt bridge interaction was not observed in the X-ray crystal structure (4KXT.pdb) of hAgo1. The L2-Mid salt bridge interaction was identified through the aid of MD simulations. However, there seems to be a propensity of the salt bridge interaction, as the D356 and K523 sidechains point towards each other. Further examination reveals that the electron density of the K525 sidechain is very sparse. In addition, K525 sidechain has a very high B factor. These factors might explain the absence of a L2-Mid salt bridge interaction in the

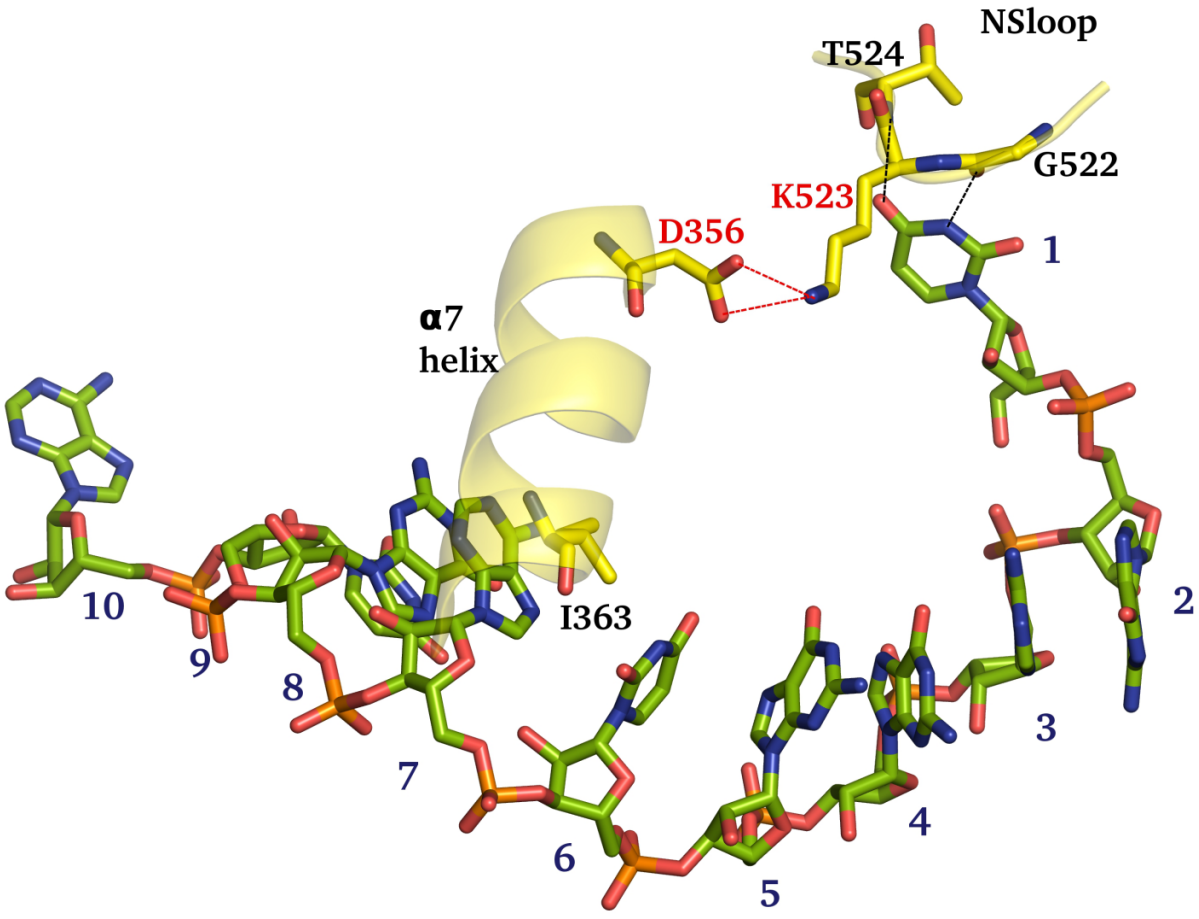

Figure 3-23: Closer view of the L2-Mid salt bridge interaction in hAgo1. Yellow cartoons represent helix 7 and NS loop. The salt bridge interaction between D356 and K523 (sticks) is represented by red dotted lines. The hydrogen bond interaction between nucleotide at position 1 (5'-end) of the guide RNA (green sticks) and NS loop represented by black dotted lines.

crystal structure (4KXT.pdb).

In Chapter 3.3 of this thesis, the biological role of the D358A mutation in hAgo2 was established. It was determined that the D358A mutation abolishes the cleavage activity in hAgo2. Following the lead for these experiments in hAgo2, the next step was to establish a potential biological role of the D356 residue in hAgo1. Although hAgo1 does not possess a catalytic function, it would be interesting to establish a potential biological significance of the D356 residue in hAgo1.

3.5.3. The role of D356 residue in hAgo1

The MD simulations and experimental studies performed on the D358 residue in hAgo2 demonstrated the significance of L2 linker domain as illustrated in Chapter 4.2 of this thesis. Due to the striking structural similarity between hAgo1 and hAgo2, it would be plausible to hypothesize that the D356 residue could play an important role in hAgo1 as well. To investigate the role of the D356 residue, MD simulations of a D356-hAgo1 were performed in which the D356 residue was mutated to alanine.

The most significant effect of the D356A mutation was observed on the domain motion of the D356A-hAgo1. A PCA of the D356A-hAgo1 simulation reveals that the domain motion seems to be reversed in comparison to wt-hAgo1. The PC1 (Figure 3-21(c)) obtained from D356A-hAgo1 simulation represents 26% of the variance. It illustrated that the vectors from the PAZ and Mid domains point towards each other, suggesting that the PAZ and Mid domains move towards each other. This is in contrast to the wt-hAgo1 simulations, were it was observed that the PAZ and Mid domains tend to move in opposite directions. Moreover, it was observed that the N domain motion is less pronounced in D356A-hAgo1 in comparison to wt-hAgo1. The PC2 obtained from the simulations of D356A-hAgo1 represents 14% of the variance (Figure 3-21(d)). The PC2 further illustrated a domain motion similar to the PC1. It can be observed that the vectors from PAZ and Mid domains point towards each other. The N domain motion in PC2 seems to be diminished in comparison to the PC1.

The effect of the D356A mutation on domain motions is also corroborated from the RMSD analysis of wt-hAgo1 and D356A-hAgo1. As observed in Figure 3-24, the RMSD of hAgo1 is considerably lowered due to the D356A mutation. The N domain seems to be most affected as its

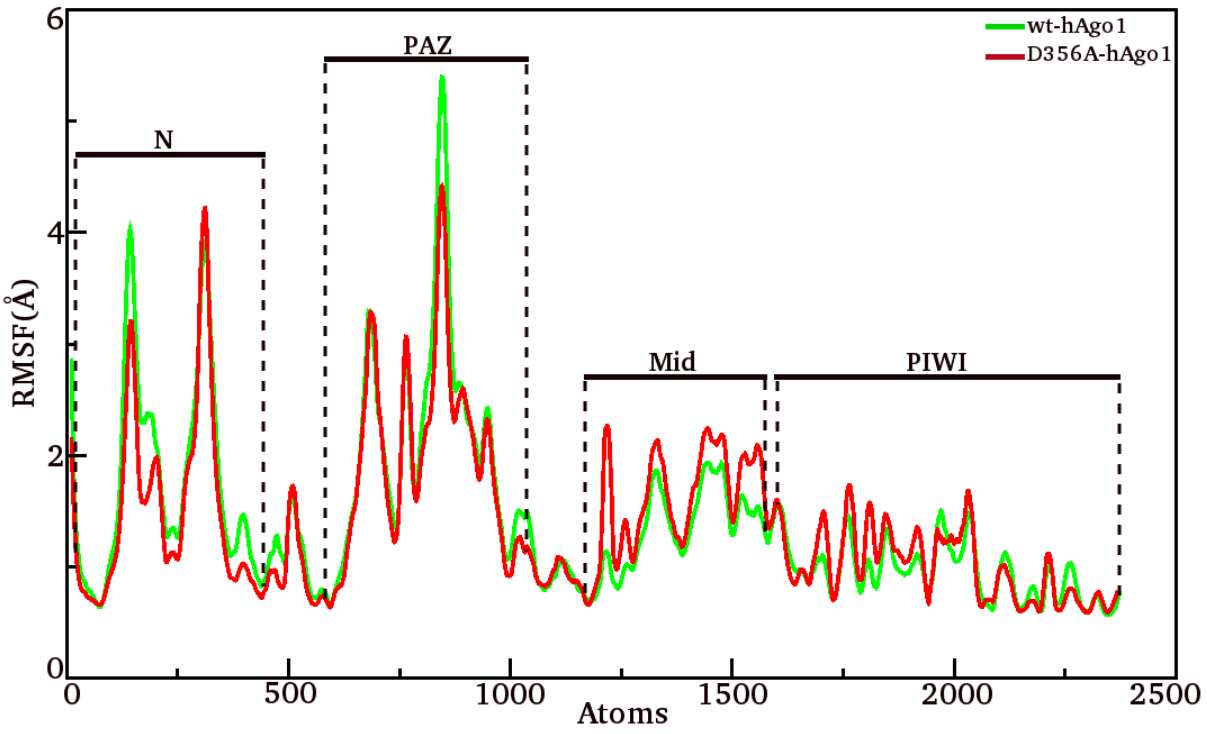

Figure 3-24: RMSF comparison of protein backbone of wt-hAgo1 (green) and D356A-hAgo1 (red). It exemplifies that the RMSF of D356A-hAgo1 is comparable with wt-hAgo1; a slight difference of ~0.5Å can be noticed for the Mid domain in case of D356A-hAgo1.

RMSD decreases from ~ 6 Å to ~ 3 Å. A significant difference can also be observed in the PAZ domain. The RMSD of the PAZ domain decreases from ~ 7.5 Å to ~ 5 Å. The RMSD of the Mid domain in D356A-hAgo1 is ~2.5 Å, which is similar to the RMSD of the Mid domain in wt-hAgo1. The RSMD of the PIWI domain in D356A-hAgo1 is also comparable to the wt-hAgo1. This indicates that major impact of the D356A mutation occurs on the domain motion of the PAZ and N domains of hAgo1.

The RMSF of the wt-hAgo1 and D356A-hAgo1 is very different to that of the wt-hAgo2 and D358A-hAgo2. In contrast to the D358A-hAgo2, no huge impact of the D356A mutation was observed on the RMS fluctuations in hAgo1 (Figure 3-24). A comparison of the RMSF between the wt-hAgo1 and D356A-hAgo1 backbones reveals that the D356A mutation does not seem to have a huge effect on the flexibility of the hAgo1. Only a slight increase in the flexibility of the Mid domain was observed, however the difference was not considerable.

3.6. The role of the L2 domain in other Argonautes

To further investigate the role of the L2 linker domain in Agos in general, a series of MD simulations were performed in a range of Agos, for which the full-length crystal structures were available.

3.6.1. KpAgo

The recently reported crystal structure of KpAgo illustrates the structural organization of individual domains (111), which is comparable to both its eukaryotic and prokaryotic counterparts. Apart from the hAgo structures, KpAgo is the only full-length structure of a eukaryotic Ago. It has four major domains: N, PAZ, Mid and PIWI, interconnected by two linker domains; L1 and L2. Similar to other Agos the 5'-end of the guide RNA binds to the Mid domain and follows its course inside a nucleic acid binding channel. The goal of this study was

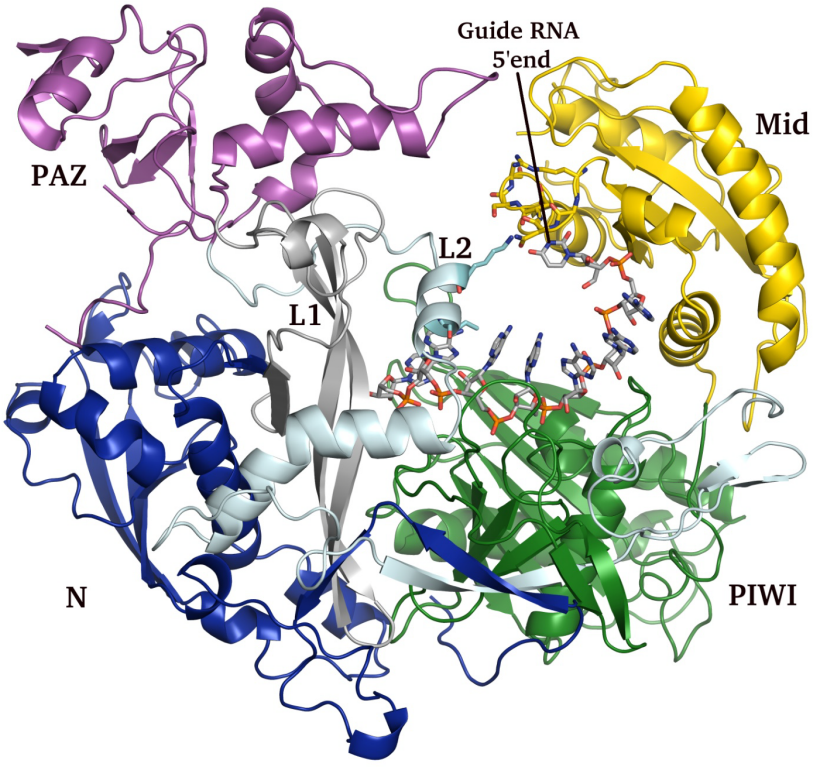

Figure 3-25: Structural organization of the KpAgo illustrating L2-Mid interaction. KpAgo is represented in cartoons, individual domains are color-coded; N (blue), L1 (silver), PAZ (magenta), L2 (cyan), Mid (yellow), PIWI (green), the L2-Mid interaction is represented in sticks. The guide RNA is represented by sticks (grey).

to investigate a possible L2-Mid interaction, similar to the L2-Mid interaction (see Chapters 3.3 and 3.5) observed for the human Agos. To investigate such a L2-Mid interaction, MD simulations of the KpAgo crystal structure (4F1N.pdb) were performed. A comparative analysis of KpAgo and hAgo sequences revealed a conserved Aspartate (D677) residue present within the helix 7 of the L2 linker domain of KpAgo. Thus, initially it was thought that the conserved D677 could potentially interact with the NS loop. However, during the MD it turned out K679 rather than D677 makes interactions with the Mid domain.

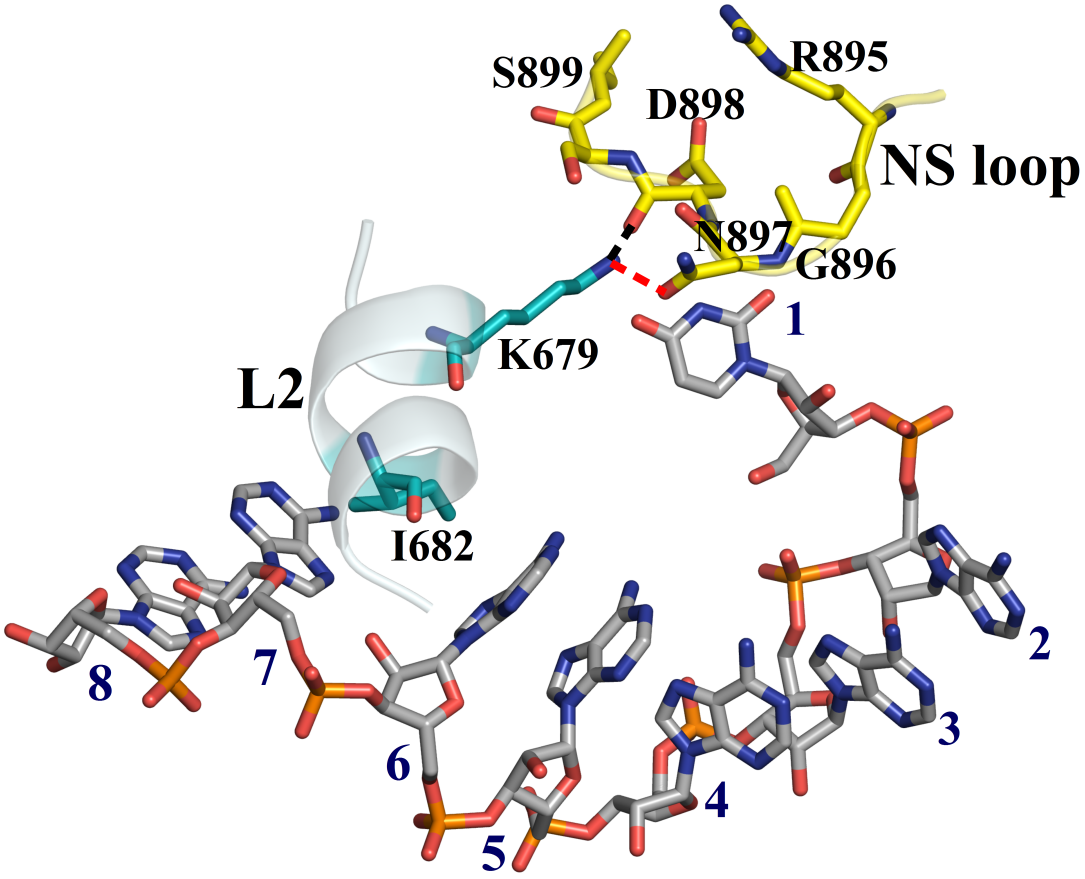

Figure 3-26: Close view of the L2-Mid interaction in KpAgo. L2 helix and NS loop are represented in pale cyan and yellow cartoons respectively. The guide RNA and protein residues are represented in sticks. Black dotted lines represent the hydrogen bond interaction between K679 sidechain and D898 backbone. The hydrogen bond interaction between K679 and N897 sidechains are represented by red dotted lines. I682 is represented by blue sticks introducing a kink between the nucleotides at position 6 and 7.

In fact, the MD simulations revealed two important findings. First, a L2-Mid interaction does occur in KpAgo, second the I682 sidechain retains a kink between nucleotides at position 6 and 7. Figure 3-25, illustrates the L2-Mid interaction and the structural organization of the KpAgo. Interestingly, a hydrogen bond interaction occurs between the K679 sidechain and the D898 backbone (Figure 3-26), which is retained for 89% of the time of the 100ns simulations performed. In addition to this interaction, another major interaction was observed between the K679 and N897 sidechains (Figure 3-26), which was retained for ~66% of the time of the simulations performed. These hydrogen bond interactions occur at different times and are retained for different timescales during the 100ns simulation performed. This study therefore illustrated that the L2-Mid interaction occurs in KpAgo, similar to that observed in the human Agos.

3.6.2. TtAgo

A series of crystal structures of TtAgo in different combinations of guide DNAs and target RNAs have been reported (109, 110). A noteworthy fact about TtAgo and prokaryotic Agos is that they bind a 'guide DNA'. For a long time, these structures have been the only ones providing precious insights into the structural organization of Agos. However, the recently reported human and yeast Ago structures have provided additional insights into the working of Agos. As reported in the previous chapters, a 'L2-Mid interaction' has been observed in the human (Chapter 4.2 – 4.4) and yeast (Chapter 4.5.1) Ago structures, which was shown to play an important role in the catalytic function of hAgo2. The goal of this study was to investigate if there is a similar L2-Mid interaction in the TtAgo structure. For this purpose, a 100ns MD simulation was performed on the TtAgo crystal structure with a bound guide DNA (3DLH.pdb). Although other crystal structures of ternary complexes were available, this structure was selected to maintain uniformity with the previous simulations, which were performed in the presence of a bound guide strand only.

The structural organization of the TtAgo is shown in the , in which the individual domains are color-coded and the important residues in the L2 and Mid domains are highlighted in sticks. Figure 3-28 illustrates the protein residues present in the L2 helix linker facing the NS loop present in the Mid domain. Three positively charged glutamic acid residues; E272, E273 and E274 are shown. Due to their key location, they were expected to interact with the protein

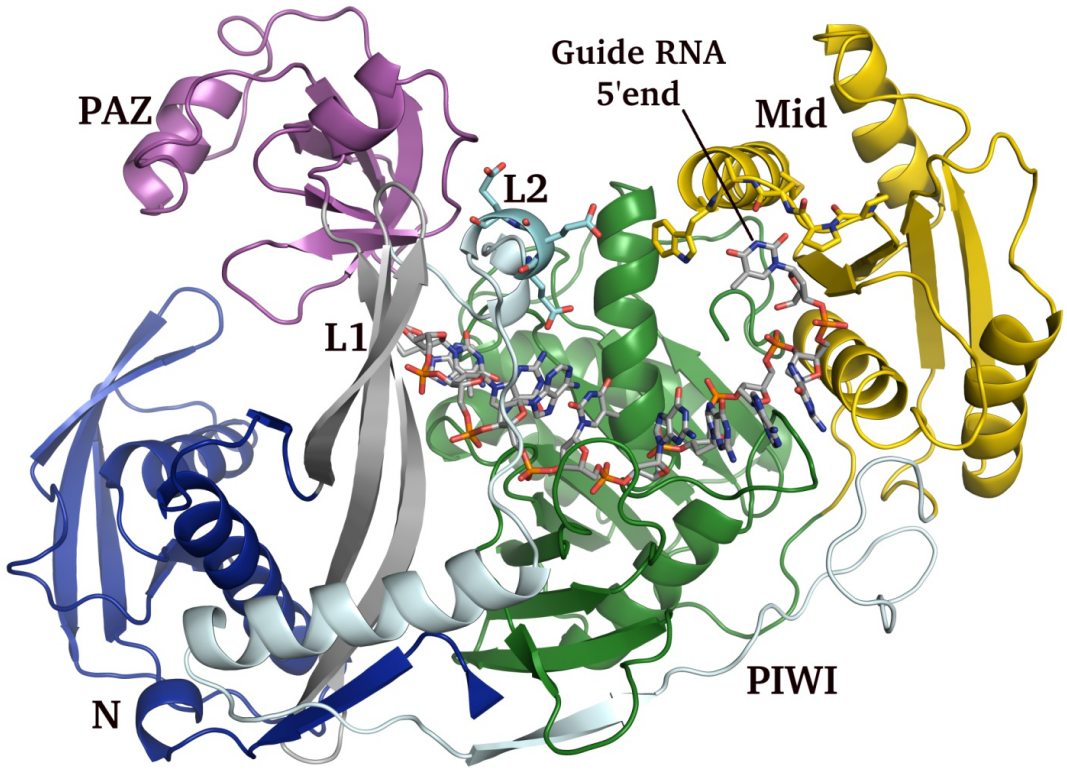

Figure 3-27: Structural organization of TtAgo in complex with a guide strand (DLH.pdb). Individual domains are color-coded; N (blue), L1 (silver), PAZ (magenta), L2 (cyan), Mid (yellow), PIWI (green). The guide RNA is represented in grey sticks. The residues in L2 and NS loop are represented in pale cyan and yellow sticks respectively.

residues present in the NS loop of the Mid domain. However, during the course of simulations performed for 100ns no interactions were observed between protein residues in the L2 helix and the NS loop.

Interestingly, it was observed that a kink or destacking occurs in the guide strand bound to TtAgo between the nucleotides at position 5 and 6. The location of the kink in TtAgo is disparate to that of the human and yeast Agos. In the guide strand bound to human and yeast Agos, the kink was observed between nucleotides at position 6 and 7. The kink in human and yeast Agos is due to a protruding sidechain of an isoleucine sticking out from the L2 linker domain. However, in case of the TtAgo this kink occurs due to the aromatic stacking between F610 and nucleotide at position 5 (Figure 3-28). Intriguingly this F610 residue is part of the PIWI domain and not the L2 domain as observed in the case of human and yeast Agos.

One of the most intriguing features of the kink observed in the guide strand bound to TtAgo is its unique location. Contrary to the human and yeast Agos in which the kink is caused by L2 linker domain, in case of TtAgo the kink is caused by the PIWI domain. Comparative analysis of the electron density of F610 residue in all the available crystal structures of TtAgo provides further insights into the formation of the kink in TtAgo. A complete electron density of F610 residue was only observed in two TtAgo structures; 3DLB.pdb (short guide DNA attached to the PAZ domain) and 3HO1.pdb (ternary structure). In the 3DLH.pdb crystal structure, which was used to perform MD simulations, the electron density was only observed for F610 backbone. In the remaining TtAgo structures, no electron density was observed for F610 residue (Table 3-8).

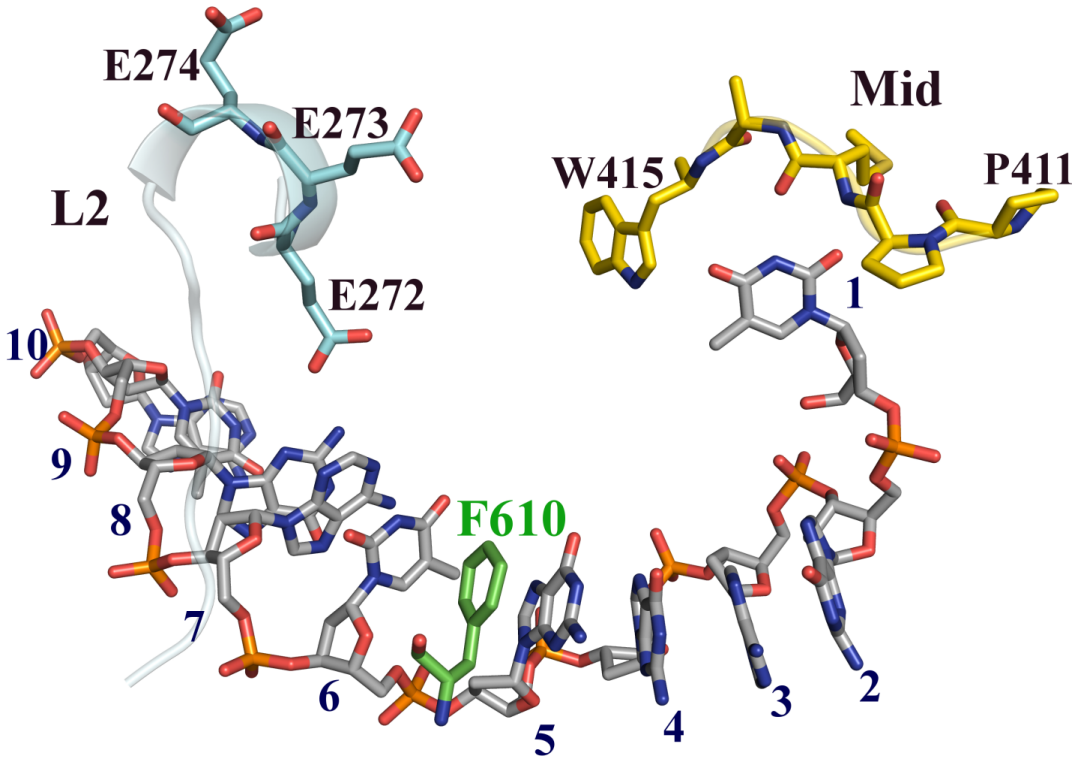

Figure 3-28: Close view of a section of the L2 domain and NS loop illustrating that no interaction occurs between the L2 and Mid domains of TtAgo. The L2 and Mid domains are represented by pale cyan and yellow sticks and cartoons, respectively. The guide strand is represented by grey sticks. The F610 protein residue (green sticks) is situated between nucleotides at position 5 and 6 of the guide strand, introducing a kink between these nucleotides.

Table 3-8: List of TtAgo crystal structures in different combinations of guide and target strands illustrating the electron density of F610 residue.

TtAgo (PDB ID)	Binary/Ternary Complex	F610 Electron density
3DLB	Apoenzyme	Complete
3DLH	Binary	Backbone only
3HO1	Ternary	Complete
3F73	Ternary	None
3HVR	Ternary	None
3HM9	Ternary	None
3HK2	Ternary	None
3HJF	Ternary	None

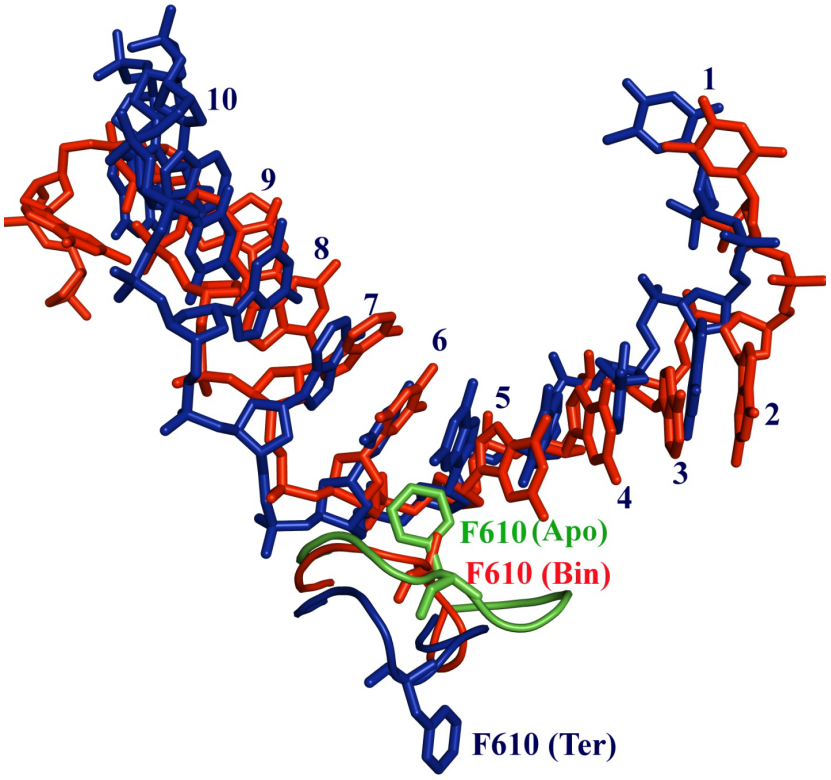

Figure 3-29: A section of the superposition of the TtAgo apoenzyme (3DLB.pdb), binary complex (3DLH.pdb) and the ternary complex (3HO1.pdb). The guide RNA bound to the TtAgo binary and ternary complexes is represented by red and blue sticks, respectively. The loop hosting the F610 residue in the TtAgo apoenzyme, binary and ternary complexes is represented by green, red and blue cartoons, respectively.

Structural superposition of the three structures, which had electron density for F610 residues, gives interesting insights into the movement of the loop withholding F610. On aligning the TtAgo apoenzyme (3DLB.pdb) and the binary complex (3DLH.pdb), it was observed that the position of F610 remains unchanged (Figure 3-29). However, alignment of the binary (3DLH.pdb) and ternary complex (3HO1.pdb) reveals a shift in the loop carrying the F610 residue. It can be observed in the ternary complex (3HO1.pdb) the F610 loop moves further away from the guide strand. In addition to the movement of the loop carrying the F610 residue, the orientation of the F610 residue is also completely opposite to that observed in the binary complex (3DLH.pdb). This movement of the F610 loop on target binding abolishes the kink between the nucleotides at position 5 and 6 in the guide strand in the TtAgo ternary complex. In contrast, the kink can be clearly observed between the nucleotides at positions 5 and 6 in the binary complex (3DLH.pdb).

3.6.3. PfAgo

Pyrococcus furiosus (*P. furiosus*) belongs to Archae kingdom. Structurally PfAgo is relatively closer to its eukaryotic then its prokaryotic counterparts. The L2 domain specifically is similar to the eukaryotic Agos. The L2 domain in PfAgo has a helix facing the NS loop, which is similar to helix7 of the eukaryotic Agos (Figure 3-31). In comparison, the prokaryotic Agos such as TtAgo and AaAgo, the section of L2 domain facing the NS loop does not have a well-defined helix.

To investigate whether a L2-Mid interaction occurs in PfAgo, MD simulations of PfAgo (1U04.pdb) were performed for 100 ns using the Gromacs simulation package. During the simulations, it was observed that the L2 domain does interact with the NS loop of the Mid domain. However, this interaction does not occur through the helix facing the NS loop, it occurs through the D280 residue which is present before the helix. Hence, the L2-Mid interaction in PfAgo occurs between protein residues D280 and N481 (Figure 3-30Figure 3-28).

A hydrogen bond formation occurs between the D280 and N481 backbones. The N481 protein residue is part of the NS loop of the Mid domain. Once this interaction was formed, it was retained for almost ~50% of the time of the 100ns simulations performed. Hence, it seems to be a significant interaction.

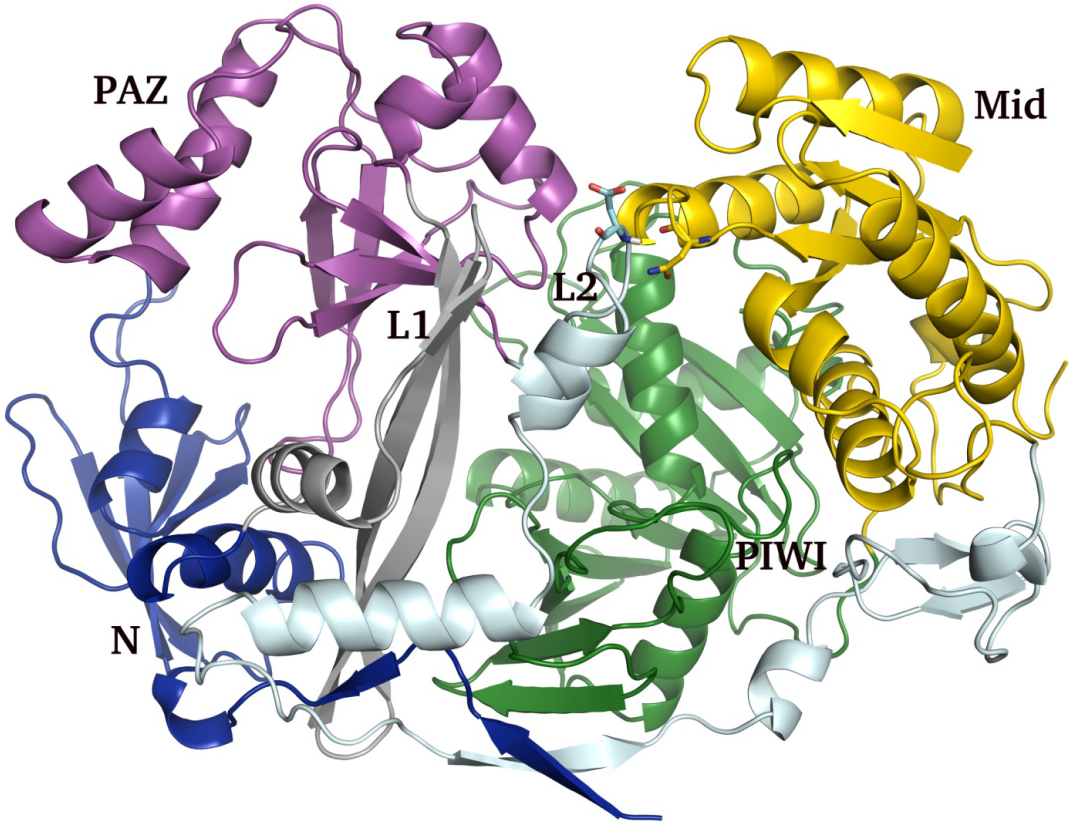

Figure 3-31: Structural organization of the PfAgo protein (1U04.pdb). Individual domains are color-coded; N (blue), L1 (grey), PAZ (magenta), L2 (pale cyan), Mid (yellow) and PIWI (green).

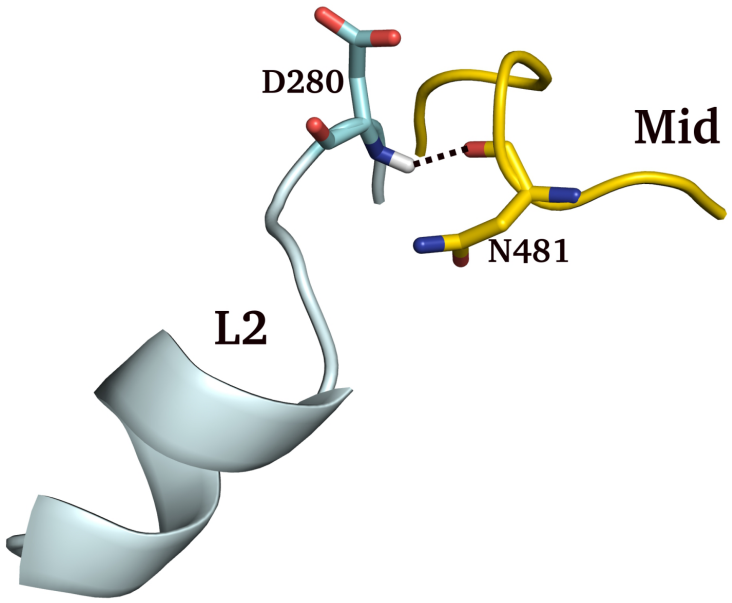

Figure 3-30: A snapshot of the L2 –Mid interaction in PfAgo at 20ns of the MD simulations. L2 and Mid are represented by pale cyan and yellow cartoons. D280 and N481 protein residues are represented in pale cyan and yellow sticks respectively. The black dotted line represents the hydrogen bond interaction between the D280 and N481 protein residues.

3.6.4. AaAgo

AaAgo is a prokaryotic Ago like TtAgo, the structural organization as seen in Figure 3-32 is similar to its eukaryotic counterparts. However, a major difference occurs in the structural organization of the L2 linker domain. The part of the L2 linker domain facing the NS loop forms a helix called the helix7 in all the eukaryotic Agos. However, in AaAgo there is no well-defined helix in the segment of the L2 domain facing the NS loop. To investigate whether a L2-Mid interaction occurs in AaAgo, MD simulations of the AaAgo (2NUB.pdb) were performed. The most striking observation during the simulations was the domain motions affecting the entire protein. The domains move towards each other forming a very compact structure.

During the simulations, multiple interactions were observed between the L2 and Mid domains. The first interaction occurs between the E271 residue with residues K429 and V430. A hydrogen bond formation occurs between the sidechains of E271 and K429. In addition to this interaction E271 interacts alternately with the V430 backbone (Figure 3-33). However, these interactions are

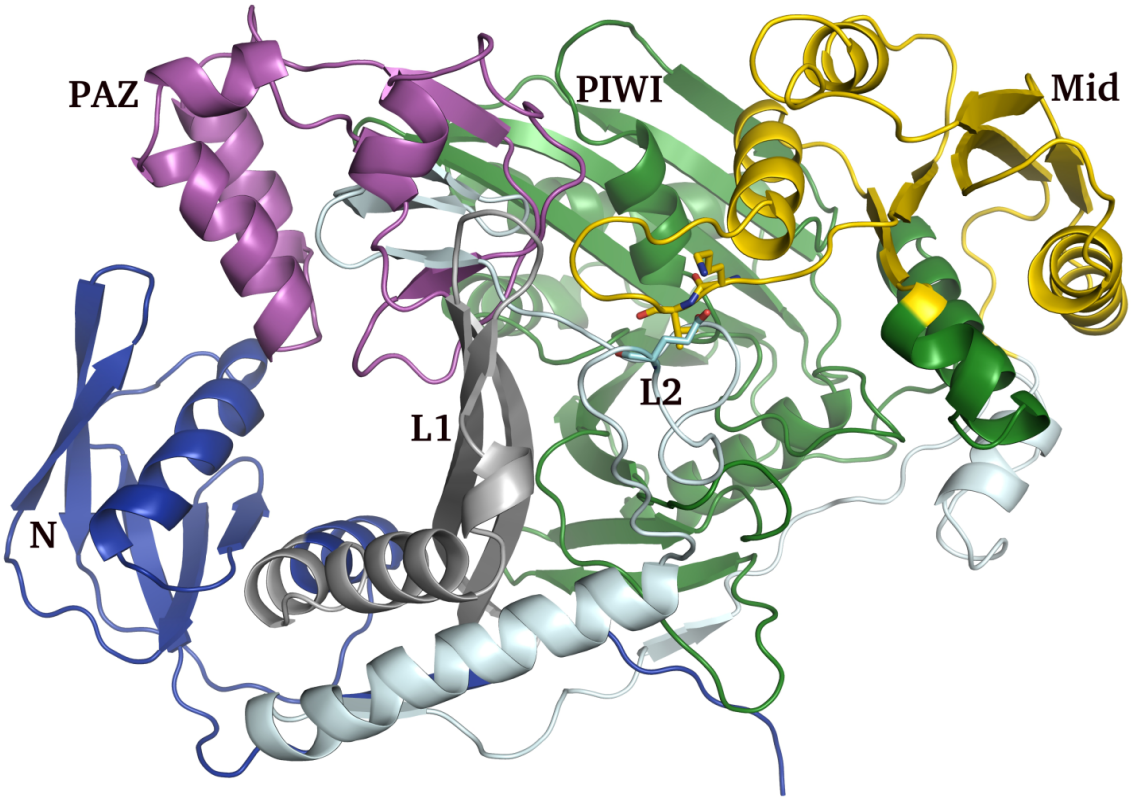

Figure 3-32: Structural organization of the AaAgo (2NUB.pdb). Individual domains are color-coded; N (blue), L1 (silver), PAZ (magenta), L2 (pale cyan), Mid (yellow), PIWI (green).

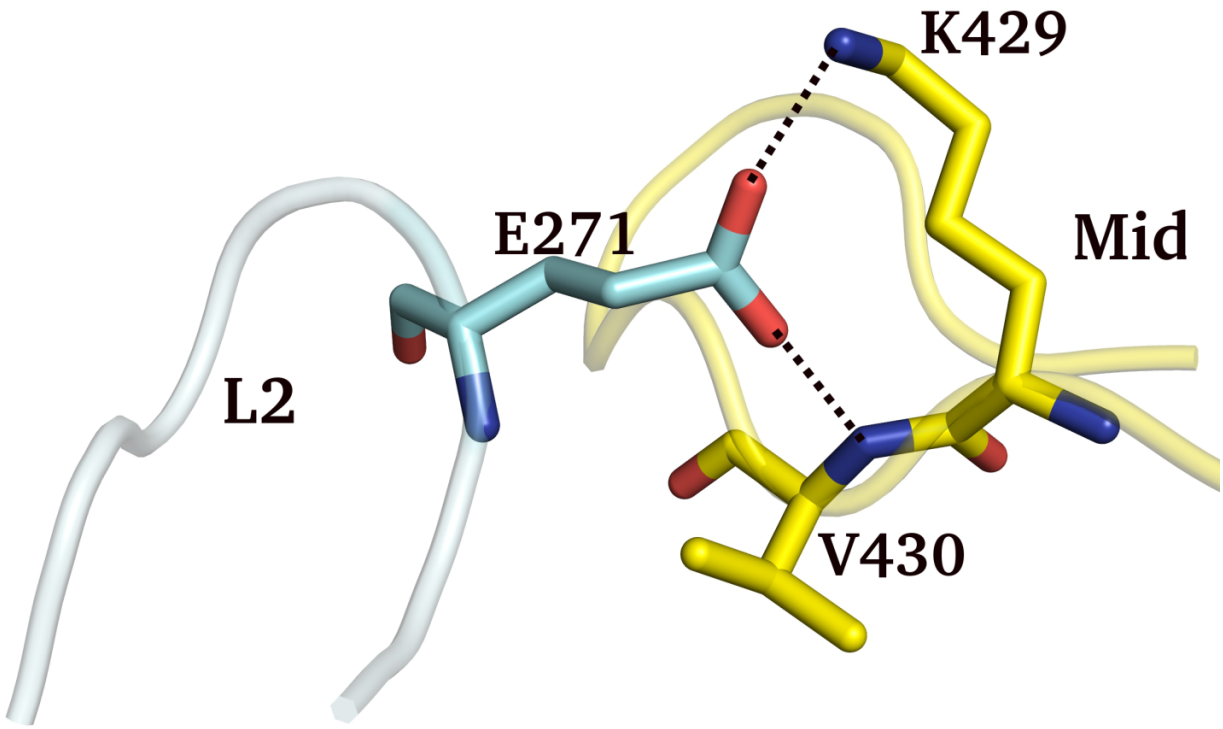

Figure 3-33: A snapshot from the MD simulations of AaAgo at 10 ns illustrating the L2-Mid interaction. The L2 and Mid are represented in pale cyan and yellow cartoons, respectively. E271, K429 and V430 residues are represented by blue and yellow sticks, respectively. The hydrogen bond interaction between E271 and K429 or V430 sidechains (yellow sticks) is represented by black dotted lines.

short lived and are only retained for ~30% and ~2% of the time of the 100 ns simulations performed.

In addition to the interaction of the E271 sidechain with the residues from the NS loop, the backbone of E271 protein residue forms a hydrogen bond interaction with the V430 backbone (Figure 3-34). Once this interaction is formed it is retained for almost ~56% of the time of the 100ns simulations performed, therefore, it seems to be a significant interaction.

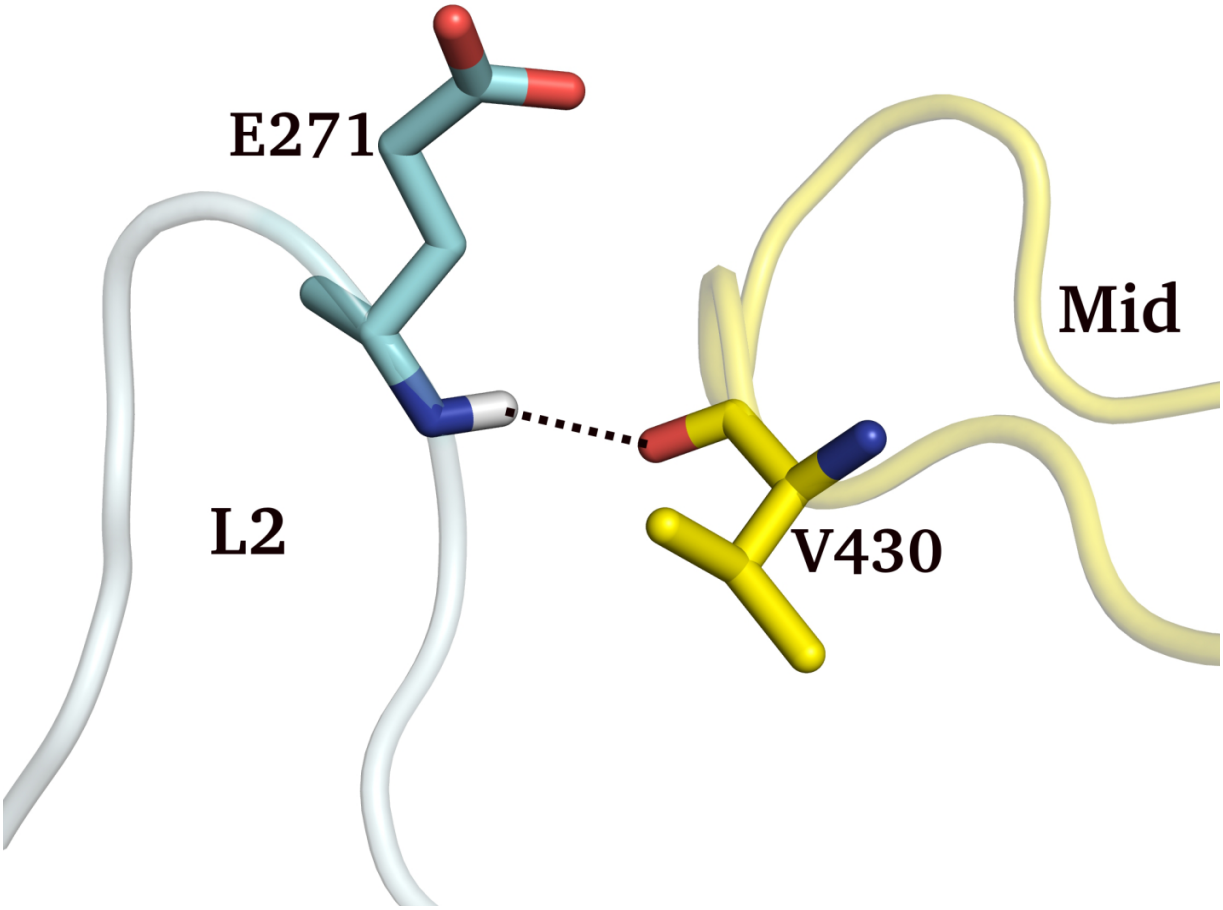

Figure 3-34: A snapshot from the MD simulations of AaAgo at 30 ns illustrating the L2-Mid interaction. The L2 and Mid domains are represented in pale cyan and yellow cartoons, respectively. The E271 and V430 are represented in blue and yellow sticks, respectively. Black dotted lines represent the hydrogen bond interaction between E271 and V430 backbones.

Chapter 4. DISCUSSION

4.1. Homology modelling of hAgo2

The crystal structures of hAgo2 in complex with bound guide RNA were recently reported in the year 2012 (113, 123). However, before these crystal structures were reported there was no structural information available for the full-length hAgo2. The only structural information available for hAgo2 was the crystal structures of isolated Mid domain (126). Therefore, in this present study the homology modelling of full-length hAgo2 was performed.

However, the task of predicting a reliable 3D structure for hAgo2 was extremely challenging primarily due to very low sequence identity with other protein sequences. The sequence identity between hAgo2 and the template sequences varied between 8% - 13%, which is extremely low to predict an accurate homology model (197). In order to overcome this challenge, multiple template sequences were selected, which covered different sections of the hAgo2 sequence. All together, hundred models of hAgo2 were predicted with Modeller9v10 software (169) and the final model was selected based on the lowest DOPE score (171).

Further assessment of the quality of the final model was performed with MolProbity, which has been shown to be a good method to quantify the quality of protein and nucleic acid structures (185). A low MolProbity score reflects a high quality structure. The MolProbity score of the hAgo2 homology model was 4.11, while the MolProbity score of hAgo2 crystal structure was 1.42 (174, 175). Although there is a large difference between the MolProbity scores of the hAgo2 homology model and the crystal structure, the score is still reasonable for a homology model predicted based on such low sequence identity. It has been demonstrated in the literature that a homology model of inhibitor κB kinase-β (IKK-β) based on sequence identity ranging between 29% to 31% yields a MolProbity score of 3.08 (198).

A comparison of the homology model of hAgo2 with the X-ray crystal structure reveals a RMSD of ~24 Å. The overall fold in the major domains seems to be comparable, however a closer inspection reveals multiple differences. The most profound distinction was noted outside the folded domains in the L1 and L2 linker domains. In addition, a large difference between the two structures was also observed in the N and PAZ domains. The greatest level of structural

similarity between the hAgo2 homology model and the crystal structure was observed in the Mid domain. It is however not very surprising as the sequence identity in the region of Mid domain was identical due to the already available Mid domain crystal structure (126).

However, in the light of the recently reported hAgo2 crystal structures the purpose of a homology based 3D structure is undermined. It was a rather challenging undertaking to predict a homology model at such a low sequence identity, which is illustrated by the low accuracy of the predicted homology model hAgo2. Nevertheless, the homology model of hAgo2 provided a better understanding of its structural organization before its crystal structures were reported. Interestingly, it also provided a unique opportunity of comparing the accuracy of a model predicted based on the homology with an actual crystal structure. Moreover, for all the subsequent studies the crystal structure of hAgo2 was utilized (43).

4.2. The effect of different guide RNA 5'-bases on the dynamic behaviour of the hAgo2

hAgo2 is a keyplayer of RNA interference (RNAi). The Mid domain closely interacts with the phosphorylated 5'-end of the guide RNA through multiple interactions. hAgo2 differentiates between the bases at the 5'-position of the guide RNA with a strong preference for U or A. hAgo2 has a 30 fold higher binding affinity for 5'-U and 5'-A over the 5'-C and 5'-G nucleoside monophosphates (126). This suggests that the hAgo family has gone into great lengths to discriminate different 5'-bases. It can be speculated that hAgo proteins evolved in a manner to accommodate guide RNAs with 5'-U over other 5'-bases based on the sheer abundance of the human miRNAs with 5'-U. However, there is little understanding of the after effects of the different 5'-bases, once the guide RNA is bound to the hAgo2.

In this study, the effect of different bases present at the 5'-end of the guide RNA on the dynamic behaviour of the hAgo2 was investigated. Previous studies have primarily focused on the impact of an isolated nucleotide on the Mid domain. However, the understanding of entire hAgo2 in presence of guide RNA is still limited. Therefore, MD simulations were performed on the recently reported crystal structure (PDB ID code: 4F3T) of the hAgo2 in complex with a bound truncated (1-10 nt) guide RNA. This study illustrates the effect of a specific 5'-base on the

flexibility of the entire hAgo2. To investigate the influence of a 5'-base on the hAgo2, the 5'-U present in the hAgo2-guide RNA was replaced by a 5'-A, 5'-C and 5'-G, respectively.

MD simulations of the hAgo2-guide RNA complex reveal that the nature of the 5'-base of guide RNA influences the flexibility of entire hAgo2. The 5'-G induces domain motions in the Mid domain, which triggers novel interactions between the 5'-G and the helix 7 of the L2 linker domain. This interaction is an interesting finding as it is observed only, in the presence of a 5'-G base of the guide RNA. It is important to note that in the hAgo2 Mid domain co-crystal structure with 5'-GMP, residual electron density was observed only for the phosphate and ribose of the nucleotide, whereas the electron density of the base was notably missing (126). It implies that a lack of stabilizing interactions between 5'-G and helix 7 of L2 linker, results in dynamic disorder in the crystal.

In addition, these interactions occur after ~20 ns into the simulation, when hAgo2 undergoes additional domain movements and conformational changes in order to accommodate 5'-G in the most stable orientation. One of the prerequisites for this interaction is the inward movement of the Mid domain towards the helix7 of L2 region. Since the 5'-end of guide RNA is tightly bound to the Mid binding pocket, the guide RNA could be pushed deeper into the nucleic acid binding channel. This novel hydrogen bond and stacking interactions suggest that although hAgo2 has been described to show the lowest binding affinity of all four nucleoside monophosphates (126) it could be most stable in the context of oligonucleotide once these interactions are formed. These novel interactions close to the 5'-end could potentially affect the positioning of guide RNA in its nucleotide binding channel.

It is interesting to note that the flexibility of the PAZ domain has been documented previously (106, 155), while the flexibility of the N and Mid domains has not been much discussed so far. It is an intriguing finding that 5'-G tremendously affect the flexibility of the Mid domain, to which the 5'-end of guide RNA binds. This study alludes to the fact that the hAgo2 can conceive a myriad of conformational changes to adapt the different 5'-bases. This suggests that although the hAgo2 might show an initial discrimination against different guide 5'- bases due to the NS loop geometry, once the guide RNA is bound the hAgo2 it becomes complacent. This could explain how hAgo2 can bind a diverse range of the miRNAs. These MD studies are also in agreement with the biochemical studies performed Sarah Willkomm in the Restle group with different bases

at the 5’-end of the guide RNA. There were no apparent differences observed in the binding of different 5'-guide substrates during binary complex formation or dissociation. Moreover, it was observed that the effect of different 5’-bases on the cleavage efficiency of the target RNAs is comparable (Kalia, Willkomm, et al. ms in preparation).

These data provide crucial insights into the dynamic behaviour of hAgo2 in the presence of different 5’-bases. However, one of the major limitation of this study is that the 5’-phosphate was excluded from the MD simulations, due to the lack of parameters. It is known that the 5’-phosphate forms multiple interactions within the Mid binding pocket which has been deemed important for the effective positioning of the guide RNA within the nucleic acid binding channel (43, 126, 195). This could somehow interfere with the observations made from the MD studies. To overcome this limitation all the simulations were repeated, especially in case of the 5’-G the simulation was performed three times and the novel interaction between the 5’-G and the L2 linker domain was observed each time. Moreover, binding and cleavage experiments performed by Sarah Willkomm at the Restle group also corroborated that guide strands, regardless if the 5'-end is phosphorylated or not (with the exception of 5’-U and to some extent 5’-G), do not affect binding affinity. Moreover, neither the hAgo2-mediated target RNA cleavage activity nor the observed cleavage position is changed by any of the four different nucleotides.

Taken together, through this study some crucial findings have surfaced with the aid of MD simulations, which have shed some light into the role different 5’-bases play in the dynamic behaviour of the hAgo2.

4.3. Role of the D358 residue in the catalytic function of hAgo2

Agos play an important role in the functioning of the RNAi pathway. hAgo2 possesses a slicer/endonucleolytic activity through which it cleaves and silences a diverse range of target mRNAs. The recently reported structures of hAgo2 suggest that different domains within the protein play important roles in guide RNA binding and in its slicer activity (113, 123). The 5’-end of the guide RNA binds to the Mid domain, the 3’-end binds to the PAZ domain and the PIWI domain contains the catalytic tetrad that cleaves the target mRNA. The N domain might

facilitate unwinding of the small RNA duplex and release of the passenger RNA strand (199). However, to date no functional role has been ascribed to either the L1 or L2 linker domains.

It has been suggested that the conformational flexibility is crucial for the catalytic function of Agos (152, 153). Recently, Deerberg *et al.* (19) demonstrated that the release of the guide RNA 3'-end from the PAZ domain is obligatory for the catalytic function of the recombinant hAgo2. Comparison of some of the earliest full-length structures of the AaAgo, pointed towards a PAZ domain flexibility (106, 107). The PAZ domain flexibility was proposed as a potential regulator of the catalytic function of Agos (106). Moreover, structures of TtAgo bound to guide DNA and various lengths targets are associated with significant domain motions (109, 154). In addition, single molecule FRET studies further suggest that the interaction between the PAZ domain and the 3'-end of the guide RNA is dynamic (19, 152).

The conformational flexibility of the PAZ domain has been further highlighted by several MD studies performed on various TtAgo structures (155-157). One of these studies showed that disruptive mutations (G→C) in the seed region of the guide DNA is associated with significant conformational changes in the L1/L2 linker region that results in opening of the nucleic acid binding channel (157). It is important to note that these studies were performed only on prokaryotic Agos, and it is difficult to discern how these conformational changes are related to the actual function of hAgo2.

To further clarify the role of the conformational flexibility in hAgo2 binding to guide RNA, all atom simulations of the hAgo2-guide RNA complex were performed. The data suggests that mobility of the PAZ domain is modulated by the L2 linker domain. Furthermore, a residue was identified in the L2 linker domain that may play a role in guide RNA binding. The D358 residue closely interacts with the Mid domain through a salt bridge formation with its NS loop. Moreover, the salt bridge allows the NS loop, which hydrogen bonds to the 5'-end of the guide RNA, to be optimally oriented such that the guide RNA can be correctly positioned within the hAgo2 central cleft. Secondly, this salt bridge may play a role in regulating the flexibility of the PAZ domain. Simulations of the D358A-hAgo2 suggest that the PAZ domain is considerably more flexible when the salt bridge is absent.

Biochemical analysis of a recombinant D358A-hAgo2 revealed that binary complex formation occurs which consists of D358A-hAgo2 and guide RNA. There are three phases of guide RNA binding to hAgo2 observable in the case of wt-hAgo2 (19). These phases can be assigned to the formation of a collision complex, guide 5'-end binding to Mid domain, and the attachment of the guide 3'-end to the PAZ domain. In the case of D358A-hAgo2, binary complex formation is merely a two-step process; i.e. binding of the 3'-end to PAZ is not observable.

The simulations of the D358A- hAgo2-guide RNA complex help explain the results of the *in vitro* studies. The D358A mutation increases the mobility of NS loop. The 5'-end of guide RNA forms hydrogen bonds with the NS loop, the increased mobility of NS loop also affects the relative positioning of guide RNA in the nucleic acid binding channel. An overlay of the guide RNA bound to the wt-hAgo2 and D358A-hAgo2 reveals that the relative positioning of the guide RNA is affected due to the D358A mutation. This could further alter the positioning of guide RNA such that the formation of a functional RISC is negatively affected, which is supported by the fact that we observe neither formation of ternary complexes nor D358A-hAgo2-mediated cleavage of target or passenger RNAs.

While these data are encouraging, it is important to realize that MD simulations have limitations. The crystal structure of wt-hAgo2 bound to a guide RNA (4F3T) has six missing nucleotides. As the precise structure of these nucleotides is not known, the 3'-end of the guide RNA (which binds the PAZ domain) was excluded from these calculations and a truncated version of the guide RNA (1-10 nucleotides) was employed. Consequently, it is difficult to make conclusions about the how the structure of the guide at the 3'-end is affected by introducing a D358A mutation. Nevertheless, it was observed that the relative motion of the PAZ domain is increased in the D358A-hAgo2. These data are consistent with the experimental observation that binding of the 3'-end of the guide RNA to the PAZ domain is impaired in the D358A-hAgo2.

Overall, these data argue that residue D358 in the L2 linker domain plays an important role in the functioning of hAgo2. Moreover, the simulations suggest that the importance of this residue stems from its interaction with an oppositely charged K525 residue in the NS loop.

4.4. Role of the I365 residue in hAgo2

One of the universal features of the guide RNAs bound to the Agos is the presence of kinks or destacking at distinctive locations, usually caused due to protruding sidechains of the neighbouring protein residues. In all the eukaryotic Agos for which a full-length structure with bound guide RNA was available, a distinctive kink was observed between nucleotides at position 6 and 7. The kink is caused by a protruding sidechain of an isoleucine residue emanating from the L2 linker domain (104, 111, 113, 115, 123).

In hAgo2, the first prominent kink is caused by the I365 residue, its sidechain forms a stacking interaction with the nucleotide at position 7 of the guide RNA (113, 123). So far, the significance of this kink or destacking is not clearly understood. However, there are two noteworthy facts related to the I365 residue first, it is important to note that the I365 residue is situated at the bottom of the helix 7, which hosts the D358 residue at its pinnacle. The significance of the D358 residue in the catalytic function of hAgo2 has been established in Chapter 3.3 of this thesis. Second, it directly interacts with the guide RNA, which makes it tempting to speculate that it might have some crucial role in the functioning of hAgo2.

To further investigate the significance of the kink long timescale MD simulations of the wt-hAgo2 were performed. Interestingly, a closer inspection of the kink during the MD simulations of the wt-hAgo2 revealed that this kink persists throughout the length of the simulations performed. In wt-hAgo2, guide RNA binding occurs in three phases, which correspond to collision complex formation, guide 5'-end-Mid binding and guide 3'-end-PAZ binding (19). With a recombinant I365A-hAgo2 mutant it was observed that during binary complex formation, the second and third phases of association are slightly slowed down. Furthermore, dissociation of binary complexes is slowed down in case of the second phase by a factor of > 10. Moreover, the cleavage efficiency of this mutant is significantly reduced. The reason for this observation is currently unclear and could in part have to do with a higher tendency of this particular mutant for aggregation.

In order to explain the biochemical data, the I365 residue was mutated to an alanine and lengthy MD simulations were performed. It was observed in the I365A-hAgo2 simulation that the kink between the nucleotides at position 6 and 7 diminishes with time. This is in contrast to the wt-

hAgo2 where the kink was retained during the time of the simulations performed. Towards the end of the I365A-hAgo2 simulation, destacking between nucleotides 6 and 7 was abolished and a perfect re-stacking of the nucleotides took place. In the absence of the I365 residue, which anchors nucleotide 7 into the nucleic acid binding channel, the flexibility of the guide RNA is considerably increased. The hypothesis is further corroborated by the increased mobility of the nucleotides at position 6 and 7. The I365A mutation also increased the RMSD of the guide RNA bound to the protein immensely. Interestingly, the I365A mutation did not seem to increase the RMSF of the protein backbone, hinting that the I365A mutation might exclusively affect the guide RNA and not the protein.

Although these results are exciting, the present study has a major limitation. The MD simulations of the wt-hAgo2 and I356A-hAgo2 were performed in the presence of a truncated guide RNA. The 5'-end of the guide RNA and the subsequent 10 nucleotides were used in the simulations. The succeeding six nucleotides (11-16) were not present in either of the crystal structures (4F3T, 4OLA). Therefore, these six nucleotides and the segment of the guide RNA bound to the PAZ domain were excluded from the simulations. The fact that the guide RNA was only bound at its 5'-end during the simulations might have contributed to the increased flexibility. The guide RNA bound to hAgo2 has an additional kink between the nucleotides at position 9 and 10, which is caused by the protruding sidechains of R635 and R710 residues. These residues might also play an important role in anchoring the guide RNA inside the nucleic acid binding channel, however at present it is only a speculation.

This study clearly illustrated the significance of the I365 protein residue; moreover, it strengthens the crucial role the L2 linker domain plays in the overall functioning of hAgo2. The data suggests that the L2 linker domain does two things in hAgo2; first, it acts as a bridge between the PAZ and Mid domains through the L2-Mid interaction, thereby regulating the PAZ domain flexibility. Second, it pins down the guide RNA effectively in its nucleic acid binding channel while awaiting the advent of the target RNA.

4.5. The role of D356 residue in hAgo1

hAgo1 is one of the four members of the human Ago family, it closely associates with small RNAs, however it does not possess a catalytic slicer activity (63, 64, 104). In addition, it has been demonstrated that in humans only hAgo1 and hAgo2 can dissociate miRNA duplexes termed as the 'strand-dissociation activity' (200). Although Ago1 does not have a catalytic function in the humans, it has been demonstrated that Ago1 possesses a catalytic slicer activity in *Arabdopsis* (201). In addition, it has been recently shown that hAgo1 plays an important role in bodily functions in humans under stress conditions such as hypoxia, which induces tumor angiogenesis. The angiogenesis induced due to hypoxia is decreased due to the over expression of hAgo1 (202).

Recently two crystal structures of the hAgo1 in presence of a bound guide RNA were reported (104, 115). These structures revealed than hAgo1 is structurally strikingly similar to hAgo2. The sequence identity between hAgo1 and hAgo2 is ~88%. The major differences between the hAgo1 and hAgo2 sequences occur in the PIWI domain, which embodies the catalytic site in hAgo2. In addition, hAgo1 has an arginine instead of a histidine in comparison to the DEDH catalytic tetrad in hAgo2. Surprisingly, the restoration of the catalytic tetrad in hAgo1 did not activate the slicer activity in hAgo1. It was observed that an additional point mutation in a loop adjacent to the catalytic site was required to turn on the catalytic activity in hAgo1 (104).

In the Chapter 3.3 of this thesis, a salt bridge interaction crucial for the catalytic function of hAgo2 was identified between the L2 and Mid domains. In the light of the structural and sequence similarities between hAgo1 and hAgo2, it was envisaged that a similar interaction might occur which could be important for the functioning of hAgo1. A MD simulation of the wt-hAgo1 revealed that a L2-Mid salt bridge interaction occurs in hAgo1. The interaction occurs between the K523 residue present in the NS loop of the Mid domain and D356 residues present in the helix 7 of L2 linker domain, in the exact same manner and location as hAgo2. In addition, to the D356 residue, hAgo1 has the I363 residue, which emanates from the conserved helix 7 of the L2 linker domain and introduces a kink between nucleotides at position 6 and 7 of the guide RNA. The kink between nucleotide 6 and 7 was preserved throughout the length of the simulations. It was observed that the domain motions in wt-hAgo1 were comparable to that of wt-hAgo2.

To investigate the role of the D356 residue in hAgo1, MD simulations of the hAgo1 were performed in which the D356 residue was mutated to an alanine. It was observed that the motions of the PAZ and N domains were increased. Interestingly, the domain motions of D356A-hAgo1 were quite contrasting to that of D358A-hAgo2. In D356A-hAgo1 the domain motion of the PAZ and N domains was considerably reduced, whilst in D358A-hAgo2 corresponding domain motions were increased. Furthermore, in D356-hAgo1 the direction of domain motions was also reversed, the PAZ and Mid domains moved towards each other. Whereas, in D358A-hAgo2 the PAZ and Mid domains move away from each other. In D356A-hAgo1 the flexibility of the PAZ and N domains seems to be alleviated, relative to D358A-hAgo2 were the flexibility of the PAZ and N domains increased manifold.

The biological role of the D356 in the functioning of hAgo1 is not substantiated forthwith. However, it can be speculated that it might have a similar role as the D358 had in the hAgo2, as demonstrated in the Chapter 3.3 of this thesis where the PAZ binding of the guide RNA 3'-end was affected due to the D358A mutation. Collectively, these data suggest that the D356 residue might be crucial for the functioning of hAgo1.

4.6. L2-Mid interaction in all the Argonautes

This part of the study had dual aims; first to investigate the presence of a L2-Mid interaction in all Agos, similar to the one observed in hAgos as shown in the Chapter 3.3 and Chapter 3.5 of this thesis. The second goal was to investigate the role of a kink in the guide RNA bound to the Agos as reported in the Chapter 3.4 of this thesis. This study was performed on all the Agos for which full-length crystal structures were available. A L2-Mid interaction was observed in the KpAgo, which was the only other eukaryotic Ago for which a full-length crystal structure was available (111). In addition, a distinguished kink occurs in the KpAgo between nucleotides at position 6 and 7, identical to the hAgos. Moreover, the kink is caused by the protruding sidechain of an isoleucine residue similar to that of hAgos. This suggests that the L2 domain in eukaryotes might have a universal function of regulating the PAZ domain flexibility and keeping the guide RNA effectively pinned down in its nucleic acid binding channel.

This L2-Mid interaction was also observed in PfAgo, which is part of the archae kingdom (105). Since the crystal structure did not have any bound guide RNA, the role of kinks in the guide

RNA could not be investigated. A visual inspection of the PfAgo structure shows that it is structurally more similar to the eukaryotic Agos than its prokaryotic counterparts. The biggest difference occurs in the L2 domain, a helix similar to helix7 of eukaryotic Agos is present in the PfAgo, and such a helix does not occur in the L2 domain of the prokaryotic structures.

A number of recent studies have suggested that prokaryotic Agos protects the host from the invasive foreign DNA (100, 203, 204). In the light of these recent studies, it is interesting to speculate that a potential L2-Mid interaction could play a crucial role in the functioning of the prokaryotic Agos. The MD simulations of the two prokaryotic Agos; TtAgo and AaAgo (106, 110), revealed that the L2- Mid interaction occurs in AaAgo. No L2-Mid interaction was observed in TtAgo, even though protein residues, which have a propensity of forming salt bridge and hydrogen bond interactions, were present in the NS loop and the part of L2 facing the Mid domain. One plausible explanation could be that in the part of the L2 domain facing the Mid domain a three residues (273-275) were missing in the crystal structure. Therefore, these missing residues were modelled in before performing the MD simulations. However, there is very little possibility that these residues could affect the conformational changes of the entire L2 domain.

Although no L2-Mid interaction was observed in TtAgo, a kink was noticed in the guide DNA. This kink however is unique and different in two aspects from the kink observed in the guide RNA observed in eukaryotic Agos. First, the kink is formed between the nucleotides at position 5 and 6 in comparison to the eukaryotic Agos were the kink occurs at position 6 and 7. However, a fact that is even more fascinating is that this kink is formed by the aromatic ring of the F610 protein residue, which is part of the PIWI domain, as compared to the eukaryotic counterparts were the L2 domain causes the link.

Collectively, these data suggest that the L2-Mid interaction occurs in almost all Agos, which belong to different kingdoms of life. Therefore, it is tempting to contemplate that this L2-Mid interaction could be an evolutionary conserved regulatory mechanism, which is imperative for the functioning of Agos. In addition, the range of TtAgo structures with both guide and target strands have provided crucial insights into the importance of the presence of kinks in the guide strand. A comparative analysis of three TtAgo structures; apoenzyme, binary complex and ternary complex, shows that the loop containing the F610 residue sits right next to the nucleotides 5 and 6. When the target RNA binds this loop shifts away from the guide DNA.

Therefore, this kink seems to be a mechanism of effectively positioning the guide strand within the nucleic acid binding channel.

Altogether, this study has provided key insights into the L2-Mid interaction in Agos as a potential evolutionary conserved mechanism. It has also shed some light into the importance of the kinks present in the guide strands bound to Agos.

Chapter 5. CONCLUSIONS

The findings of this thesis provide crucial insights into the working of the Agos on a dynamic and microscopic level. Several key aspects related to the functioning of Agos were investigated under the scope of this thesis. First, the effect of different 5'-bases of the guide RNA on the dynamic motions of the hAgo2 was studied. It was observed that a G at the 5'-end of the guide RNA substantially elevates the flexibility of the Mid domain. In addition, a novel interaction between 5'-G and the helix 7 within the L2 linker domain was observed. Such an interaction between a nucleotide at the 5'-end of the guide RNA and the L2 linker domain has never been observed previously. These findings suggest that the hAgo2 is highly flexible and it can adopt various conformations to accommodate different 5'-bases.

Most importantly, through the MD simulations of the hAgo2-guide RNA complex, two novel residues within the L2 linker domain, crucial for the catalytic function of hAgo2 were identified. Interestingly, the two are located within the helix 7; the D358 residue is situated at the pinnacle, whilst the I365 residue is located at the bottom of the helix 7. The D358 residue forms a salt bridge interaction with the NS loop of the Mid domain, while I365 forms a stacking interaction with the aromatic ring of the nucleotide 7.

In vitro studies illustrated that the single point mutation of the D358 residue to an alanine abolishes the catalytic function of the hAgo2 by affecting binding of the guide RNA to hAgo2. The MD simulations of hAgo2 with a D358A mutation revealed that the PAZ domain flexibility increases tremendously due to this mutation and the relative positioning of the guide RNA is affected. The *in vitro* studies illustrate that the I365A mutation slows down the rates of association and dissociation of guide RNA to hAgo2 and decrease the cleavage efficiency. MD simulations revealed that the kink was abolished due to an I365A mutation in hAgo2. Moreover, it illustrated that the guide RNA becomes increasingly flexible in the absence of this kink.

The L2-Mid interaction, which was initially identified in hAgo2, was further investigated in all prokaryotic and eukaryotic Agos for which a full-length crystal structure was available. It was observed that the L2-Mid interaction is universally present in all Agos with the exception of TtAgo. Therefore, the L2-Mid interaction within Agos can be proposed as an evolutionary conserved regulatory mechanism.

Chapter 6. REFERENCES

1. Fire A*, et al.* (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391(6669):806-811.

2. Napoli C, Lemieux C, & Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The Plant Cell Online* 2(4):279-289.

3. Romano N & Macino G (1992) Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular Microbiology* 6(22):3343-3353.

4. de Carvalho F*, et al.* (1992) Suppression of beta-1, 3-glucanase transgene expression in homozygous plants. *The EMBO Journal* 11(7):2595.

5. Lee RC, Feinbaum RL, & Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75(5):843-854.

6. Kloosterman WP & Plasterk RH (2006) The diverse functions of microRNAs in animal development and disease. *Developmental Cell* 11(4):441-450.

7. James D & Mary C F G (2014) The regulatory role of small RNAs in plant innate immunity. *International Journal of Life Sciences Biotechnology and Pharma Research* 3(1):72 - 82.

8. Johnson C & Sundaresan V (2007) Regulatory small RNAs in plants. *Plant Systems Biology*, (Springer), pp 99-113.

9. Guo S & Kemphues KJ (1995) par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed.

Cell 81(4):611-620.

10. Izant JG & Weintraub H (1984) Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. *Cell* 36(4):1007-1015.

11. Nellen W & Lichtenstein C (1993) What makes an mRNA anti-sensi-itive? *Trends in Biochemical Sciences* 18(11):419-423.

12. Hammond SM, Bernstein E, Beach D, & Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* 404(6775):293-296.

13. Hammond SM, Caudy AA, & Hannon GJ (2001) Post-transcriptional gene silencing by double-stranded RNA. *Nature Reviews Genetics* 2(2):110-119.

14. Bernstein E, Caudy AA, Hammond SM, & Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409(6818):363-366.

15. Zamore PD, Tuschl T, Sharp PA, & Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101(1):25-33.

16. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, & Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293(5532):1146-1150.

17. Chendrimada TP*, et al.* (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436(7051):740-744.

18. Rivas FV*, et al.* (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Structural & Molecular Biology* 12(4):340-349.

19. Deerberg A, Willkomm S, & Restle T (2013) Minimal mechanistic model of siRNA-

dependent target RNA slicing by recombinant human Argonaute 2 protein. *Proc. Natl. Acad. Sci. USA* 110(44):17850-17855.

20. MacFarlane L-A & Murphy PR (2010) MicroRNA: biogenesis, function and role in cancer. *Current Genomics* 11(7):537.

21. Jinek M & Doudna JA (2008) A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457(7228):405-412.

22. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136(2):215-233.

23. Rigoutsos I (2009) New tricks for animal microRNAS: targeting of amino acid coding regions at conserved and nonconserved sites. *Cancer Research* 69(8):3245-3248.

24. Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, & Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* 34(suppl 1):D140-D144.

25. Wilson RC & Doudna JA (2013) Molecular Mechanisms of RNA Interference. *Annual Review of Biophysics* 42:217-239.

26. Lagos-Quintana M, Rauhut R, Lendeckel W, & Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294(5543):853-858.

27. Lau NC, Lim LP, Weinstein EG, & Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* 294(5543):858-862.

28. Lee RC & Ambros V (2001) An extensive class of small RNAs in Caenorhabditis elegans. *Science* 294(5543):862-864.

29. Mourelatos *Z, et al.* (2002) miRNPs: a novel class of ribonucleoproteins containing

numerous microRNAs. *Genes & Development* 16(6):720-728.

30. Kim VN, Han J, & Siomi MC (2009) Biogenesis of small RNAs in animals. *Nature Reviews Molecular Cell Biology* 10(2):126-139.

31. Winter J, Jung S, Keller S, Gregory RI, & Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology* 11(3):228-234.

32. Bracht J, Hunter S, Eachus R, Weeks P, & Pasquinelli AE (2004) Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *RNA* 10(10):1586-1594.

33. Cai X, Hagedorn CH, & Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10(12):1957-1966.

34. Lee Y*, et al.* (2004) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal* 23(20):4051-4060.

35. Saini HK, Enright AJ, & Griffiths-Jones S (2008) Annotation of mammalian primary microRNAs. *BMC Genomics* 9(1):564.

36. Hutvágner G & Zamore PD (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297(5589):2056-2060.

37. Khvorova A, Reynolds A, & Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115(2):209-216.

38. Denli AM, Tops BB, Plasterk RH, Ketting RF, & Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* 432(7014):231-235.

39. Gregory RI*, et al.* (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432(7014):235-240.

40. Gregory RI, Chendrimada TP, & Shiekhattar R (2006) MicroRNA biogenesis: isolation and characterization of the microprocessor complex. (Methods Molecular Biology), Vol 342, pp 33-47.

41. Shiohama A, Sasaki T, Noda S, Minoshima S, & Shimizu N (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochemical and Biophysical Research Communications* 304(1):184-190.

42. Lindsay EA (2001) Chromosomal microdeletions: dissecting del22q11 syndrome. *Nature Reviews Genetics* 2(11):858-868.

43. Hutvágner G*, et al.* (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293(5531):834-838.

44. Grishok A*, et al.* (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* 106(1):23-34.

45. Yi R, Qin Y, Macara IG, & Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* 17(24):3011-3016.

46. Bohnsack MT, Czaplinski K, & Görlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10(2):185-191.

47. Ketting RF*, et al.* (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & Development* 15(20):2654-2659.

48. Knight SW & Bass BL (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science*

293(5538):2269-2271.

49. Elbashir SM, Lendeckel W, & Tuschl T (2001) RNA interference is mediated by 21-and 22-nucleotide RNAs. *Genes & Development* 15(2):188-200.

50. Han J*, et al.* (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18(24):3016-3027.

51. Krol J*, et al.* (2004) Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *Journal of Biological Chemistry* 279(40):42230-42239.

52. Schwarz DS*, et al.* (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115(2):199-208.

53. Okamura K*, et al.* (2008) The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. *Nature Structural & Molecular Biology* 15(4):354-363.

54. Cougot N, Babajko S, & Séraphin B (2004) Cytoplasmic foci are sites of mRNA decay in human cells. *The Journal of Cell Biology* 165(1):31-40.

55. Brengues M, Teixeira D, & Parker R (2005) Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science* 310(5747):486-489.

56. Eystathioy T*, et al.* (2003) A panel of monoclonal antibodies to cytoplasmic GW bodies and the mRNA binding protein GW182. *Hybridoma and Hybridomics* 22(2):79-86.

57. Eystathioy T*, et al.* (2003) The GW182 protein colocalizes with mRNA degradation associated proteins hDcp1 and hLSm4 in cytoplasmic GW bodies. *RNA* 9(10):1171-1173.

58. Wu L, Fan J, & Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA.

Proc. Natl. Acad. Sci. USA 103(11):4034-4039.

59. van Dijk E*, et al.* (2002) Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *The EMBO Journal* 21(24):6915-6924.

60. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2):281-297.

61. Perron MP & Provost P (2008) Protein interactions and complexes in human microRNA biogenesis and function. *Frontiers in Bioscience: a Journal and Virtual Library* 13:2537.

62. Yekta S, Shih I-h, & Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304(5670):594-596.

63. Meister G*, et al.* (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell* 15(2):185-197.

64. Liu J*, et al.* (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305(5689):1437-1441.

65. Valencia-Sanchez MA, Liu J, Hannon GJ, & Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development* 20(5):515-524.

66. Mallory AC*, et al.* (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. *The EMBO Journal* 23(16):3356-3364.

67. Zeng Y & Cullen BR (2003) Sequence requirements for micro RNA processing and function in human cells. *RNA* 9(1):112-123.

68. Doench JG & Sharp PA (2004) Specificity of microRNA target selection in translational repression. *Genes & Development* 18(5):504-511.

69. Pillai RS (2005) MicroRNA function: multiple mechanisms for a tiny RNA? *RNA*

11(12):1753-1761.

70. Petersen CP, Bordeleau M-E, Pelletier J, & Sharp PA (2006) Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell* 21(4):533-542.

71. Kong YW*, et al.* (2008) The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc. Natl. Acad. Sci. USA* 105(26):8866-8871.

72. Chekulaeva M, Filipowicz W, & Parker R (2009) Multiple independent domains of dGW182 function in miRNA-mediated repression in Drosophila. *RNA* 15(5):794-803.

73. Lazzaretti D, Tournier I, & Izaurralde E (2009) The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. *RNA* 15(6):1059-1066.

74. Tritschler F, Huntzinger E, & Izaurralde E (2010) Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of deja vu. *Nature Reviews Molecular Cell Biology* 11(5):379-384.

75. Braun JE, Huntzinger E, Fauser M, & Izaurralde E (2011) GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Molecular Cell* 44(1):120-133.

76. Till S*, et al.* (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nature Structural & Molecular Biology* 14(10):897-903.

77. Eulalio A, Huntzinger E, & Izaurralde E (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nature Structural & Molecular Biology* 15(4):346-353.

78. Braun JE, Huntzinger E, & Izaurralde E (2013) The role of GW182 proteins in miRNA-mediated gene silencing. *Ten Years of Progress in GW/P Body Research*, (Springer, New York, NY, USA), pp 147-163.

79. Paddison PJ, Caudy AA, & Hannon GJ (2002) Stable suppression of gene expression by RNAi in mammalian cells. *Proc. Natl. Acad. Sci. USA* 99(3):1443-1448.

80. Ruby JG, *et al.* (2006) Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans. *Cell* 127(6):1193-1207.

81. Pak J & Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* 315(5809):241-244.

82. Song R, *et al.* (2011) Male germ cells express abundant endogenous siRNAs. *Proc. Natl. Acad. Sci. USA* 108(32):13159-13164.

83. Watanabe T, *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453(7194):539-543.

84. Burnett JC, Rossi JJ, & Tiemann K (2011) Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnology Journal* 6(9):1130-1146.

85. Okamura K, *et al.* (2008) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453(7196):803-806.

86. Czech B, *et al.* (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature* 453(7196):798-802.

87. Sijen T, Steiner FA, Thijssen KL, & Plasterk RH (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315(5809):244-247.

88. Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and

regulations. *Genes & Development* 20(7):759-771.

89. Martienssen RA, Zaratiegui M, & Goto DB (2005) RNA interference and heterochromatin in the fission yeast Schizosaccharomyces pombe. *Trends in Genetics* 21(8):450-456.

90. Kawamura Y*, et al.* (2008) Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453(7196):793-797.

91. Mello CC & Conte D (2004) Revealing the world of RNA interference. *Nature* 431(7006):338-342.

92. Meister G & Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* 431(7006):343-349.

93. Tomari Y & Zamore PD (2005) Perspective: machines for RNAi. *Genes & Development* 19(5):517-529.

94. Cernilogar FM*, et al.* (2011) Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. *Nature* 480(7377):391-395.

95. Dalzell JJ*, et al.* (2011) RNAi effector diversity in nematodes. *PLoS Neglected Tropical Diseases* 5(6):e1176.

96. Kanellopoulou C*, et al.* (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & Development* 19(4):489-501.

97. Murchison EP, Partridge JF, Tam OH, Cheloufi S, & Hannon GJ (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl Acad. Sci. USA* 102(34):12135-12140.

98. Bohmert K*, et al.* (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf

development. *The EMBO Journal* 17(1):170-180.

99. Sheng G*, et al.* (2014) Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. USA* 111(2):652-657.

100. Swarts DC*, et al.* (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507:258-261.

101. Makarova KS, Wolf YI, & Koonin EV (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research*:gkt157.

102. Meister G (2013) Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics* 14(7):447-459.

103. Ender C & Meister G (2010) Argonaute proteins at a glance. *Journal of Cell Science* 123(11):1819-1823.

104. Faehnle CR, Elkayam E, Haase AD, Hannon GJ, & Joshua-Tor L (2013) The making of a slicer: activation of human argonaute-1. *Cell Reports* 3(6):1901-1909.

105. Song J-J, Smith SK, Hannon GJ, & Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305(5689):1434-1437.

106. Rashid UJ*, et al.* (2007) Structure of Aquifex aeolicus argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. *Journal of Biological Chemistry* 282(18):13824-13832.

107. Yuan Y-R, Pei Y, Chen H-Y, Tuschl T, & Patel DJ (2006) A Potential Protein-RNA Recognition Event along the RISC-Loading Pathway from the Structure of A. aeolicus Argonaute with Externally Bound siRNA. *Structure* 14(10):1557-1565.

108. Wang Y*, et al.* (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456(7224):921-926.

109. Wang Y*, et al.* (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461(7265):754-761.

110. Wang Y, Sheng G, Juranek S, Tuschl T, & Patel DJ (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature* 456(7219):209-213.

111. Nakanishi K, Weinberg DE, Bartel DP, & Patel DJ (2012) Structure of yeast Argonaute with guide RNA. *Nature* 486(7403):368-374.

112. Elkayam E*, et al.* (2012) The structure of human argonaute-2 in complex with miR-20a. *Cell* 150(1):100-110.

113. Schirle NT & MacRae IJ (2012) The crystal structure of human Argonaute2. *Science* 336(6084):1037-1040.

114. Faehnle CR, Elkayam E, Haase AD, Hannon GJ, & Joshua-Tor L (2013) The making of a slicer: activation of human Argonaute-1. *Cell reports* 3(6):1901-1909.

115. Nakanishi K*, et al.* (2013) Eukaryote-specific insertion elements control human Argonaute slicer activity. *Cell Reports* 3(6):1893-1900.

116. Liu J*, et al.* (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science Signaling* 305(5689):1437.

117. Song J-J, Smith SK, Hannon GJ, & Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science Signaling* 305(5689):1434.

118. Höck J & Meister G (2008) The Argonaute protein family. *Genome Biology* 9(2):210.

119. Alisch RS, Jin P, Epstein M, Caspary T, & Warren ST (2007) Argonaute2 is essential for

mammalian gastrulation and proper mesoderm formation. *PLoS Genetics* 3(12):e227.

120. O'Carroll D*, et al.* (2007) A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. *Genes & Development* 21(16):1999-2004.

121. Iosue I*, et al.* (2013) Argonaute 2 sustains the gene expression program driving human monocytic differentiation of acute myeloid leukemia cells. *Cell Death & Disease* 4(11):e926.

122. Bronevetsky Y*, et al.* (2013) T cell activation induces proteasomal degradation of Argonaute and rapid remodeling of the microRNA repertoire. *The Journal of Experimental Medicine* 210(2):417-432.

123. Elkayam E*, et al.* (2012) The structure of human Argonaute-2 in complex with miR-20a. *Cell* 150(1):100-110.

124. Kuhn C-D & Joshua-Tor L (2013) Eukaryotic Argonautes come into focus. *Trends in Biochemical Sciences* 38(5):263-271.

125. Boland A, Huntzinger E, Schmidt S, Izaurralde E, & Weichenrieder O (2011) Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc Natl Acad Sci USA* 108(26):10466-10471.

126. Frank F, Sonenberg N, & Nagar B (2010) Structural basis for 5 [prime]-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465(7299):818-822.

127. Ghildiyal M*, et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* 320(5879):1077-1081.

128. Lau NC, Lim LP, Weinstein EG, & Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science Signaling* 294(5543):858.

129. Ghildiyal M, Xu J, Seitz H, Weng Z, & Zamore PD (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* 16(1):43-56.

130. Hu HY*, et al.* (2009) Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* 10(1):413.

131. Mi S*, et al.* (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* 133(1):116-127.

132. Beilhartz GL & Götte M (2010) HIV-1 ribonuclease H: structure, catalytic mechanism and inhibitors. *Viruses* 2(4):900-926.

133. Nowotny M, Gaidamakov SA, Crouch RJ, & Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* 121(7):1005-1016.

134. Dror RO, Dirks RM, Grossman J, Xu H, & Shaw DE (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics* 41:429-452.

135. Alder B & Wainwright T (1957) Phase transition for a hard sphere system. *The Journal of Chemical Physics* 27(5):1208.

136. Rahman A (1964) Correlations in the motion of atoms in liquid argon. *Physical Review* 136(2A):A405.

137. Rahman A & Stillinger FH (1971) Molecular dynamics study of liquid water. *The Journal of Chemical Physics* 55(7):3336-3359.

138. McCammon J, Gelin B, & Karplus M (1977) Dynamics of folded proteins. *Nature* 267(5612):585.

139. Duan Y & Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282(5389):740-744.

140. Shaw DE*, et al.* (2009) Millisecond-scale molecular dynamics simulations on Anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, (IEEE), pp 1-11.

141. Tang X-l, Wang Y, Li D-l, Luo J, & Liu M-y (2012) Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets. *Acta Pharmacologica Sinica* 33(3):363-371.

142. Cherezov V*, et al.* (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein–coupled receptor. *Science* 318(5854):1258-1265.

143. Rasmussen SG*, et al.* (2011) Structure of a nanobody-stabilized active state of the β(2) adrenoceptor. *Nature* 469(7329):175-180.

144. Rasmussen SG*, et al.* (2011) Crystal structure of the [bgr] 2 adrenergic receptor-Gs protein complex. *Nature* 477(7366):549-555.

145. Dror RO*, et al.* (2009) Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. USA* 106(12):4689-4694.

146. Lyman E*, et al.* (2009) A Role for a Specific Cholesterol Interaction in Stabilizing the Apo Configuration of the Human A(2A) Adenosine Receptor. *Structure* 17(12):1660-1668.

147. Niesen MJ, Bhattacharya S, & Vaidehi N (2011) The role of conformational ensembles in ligand recognition in G-protein coupled receptors. *Journal of the American Chemical Society* 133(33):13197-13204.

148. Romo TD, Grossfield A, & Pitman MC (2010) Concerted Interconversion between Ionic Lock Substates of the beta(2) Adrenergic Receptor Revealed by Microsecond Timescale Molecular Dynamics. *Biophysical Journal* 98(1):76-84.

149. Vanni S, Neri M, Tavernelli I, & Rothlisberger U (2009) Observation of "ionic lock" formation in molecular dynamics simulations of wild-type β1 and β2 adrenergic receptors. *Biochemistry* 48(22):4789-4797.

150. Schames JR*, et al.* (2004) Discovery of a novel binding trench in HIV integrase. *Journal of Medicinal Chemistry* 47(8):1879-1881.

151. Hazuda DJ*, et al.* (2004) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc. Natl. Acad. Sci. USA* 101(31):11233-11238.

152. Jung S-R*, et al.* (2013) Dynamic Anchoring of the 3 -End of the Guide Strand Controls the Target Dissociation of Argonaute–Guide Complex. *Journal of the American Chemical Society* 135(45):16865-16871.

153. Zander A, Holzmeister P, Klose D, Tinnefeld P, & Grohmann D (2013) Single-molecule FRET supports the two-state model of Argonaute action. *RNA Biology* 11(1):0-11.

154. Wang Y, Li Y, Ma Z, Yang W, & Ai C (2010) Mechanism of microRNA-target interaction: molecular dynamics simulations and thermodynamics analysis. *PLoS Computational Biology* 6(7):e1000866.

155. Xia Z, Huynh T, Ren P, & Zhou R (2013) Large Domain Motions in Ago Protein Controlled by the Guide DNA-Strand Seed Region Determine the Ago-DNA-mRNA Complex Recognition Process. *PloS One* 8(1):e54620.

156. Joseph TT & Osman R (2012) Convergent Transmission of RNAi Guide-Target Mismatch Information across Argonaute Internal Allosteric Network. *PLoS*

Computational Biology 8(9):e1002693.

157. Xia Z*, et al.* (2012) Molecular dynamics simulations of Ago silencing complexes reveal a large repertoire of admissible 'seed-less' targets. *Scientific Reports* 2(569).

158. Magrane M (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database: The Journal of Biological Databases & Curation* 2011(9).

159. Jain E*, et al.* (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* 10(1):136.

160. Apweiler R, Bairoch A, & Wu CH (2004) Protein sequence databases. *Current Opinion in Chemical Biology* 8(1):76-80.

161. Mount DW (2007) Using a FASTA sequence database similarity search. *Cold Spring Harbor Protocols* 2007(7):pdb. top16.

162. Söding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* 33(suppl 2):W244-W248.

163. Guex N & Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 18(15):2714-2723.

164. Schwede T, Kopp J, Guex N, & Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31(13):3381-3385.

165. Arnold K, Bordoli L, Kopp J, & Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22(2):195-201.

166. Madhusudhan M, Webb BM, Marti-Renom MA, Eswar N, & Sali A (2009) Alignment of

multiple protein structures based on sequence and structure features. *Protein Engineering Design and Selection* 22(9):569-574.

167. Braberg H*, et al.* (2012) SALIGN: a web server for alignment of multiple protein sequences and structures. *Bioinformatics* 28(15):2072-2073.

168. Martí-Renom MA*, et al.* (2000) Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure* 29(1):291-325.

169. Šali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234(3):779-815.

170. Fiser A, Do RKG, & Šali A (2000) Modeling of loops in protein structures. *Protein Science* 9(9):1753-1773.

171. Eswar N*, et al.* (2006) Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics,* Chapter 5, Unit 5.6.

172. Brooks BR*, et al.* (2009) CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry* 30(10):1545-1614.

173. Hess B, Kutzner C, Van Der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4(3):435-447.

174. Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* 66(1):12-21.

175. Davis IW*, et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35(suppl 2):W375-W383.

176. Hayward S, Kitao A, & Berendsen HJ (1997) Model-free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins: Structure, Function, and Bioinformatics* 27(3):425-437.

177. Hayward S & Berendsen HJ (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T 4 lysozyme. *Proteins Structure Function and Genetics* 30(2):144-154.

178. Hartigan JA (1975) *Clustering Algorithms* (John Wiley & Sons, Inc., New York, NY, USA).

179. Goldstein H (1980) *Classical Mechanics, 2nd Edition* (Addison-Wesley Publishing Co., Reading, MA, USA).

180. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *Journal of Molecular Graphics* 14(1):33-38.

181. Schrodinger L (2010) The PyMOL molecular graphics system, version 1.3 r1, Schrödinger, LLC.

182. Emsley P, Lohkamp B, Scott W, & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 66(4):486-501.

183. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60(12):2126-2132.

184. Shen M-y & Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Science* 15(11):2507-2524.

185. Keedy DA*, et al.* (2009) The other 90% of the protein: Assessment beyond the Cαs for CASP8 template-based and high-accuracy models. *Proteins: Structure, Function, and*

Bioinformatics 77(S9):29-49.

186. Berman HM*, et al.* (2000) The protein data bank. *Nucleic Acids Research* 28(1):235-242.

187. Pérez A*, et al.* (2007) Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha/gamma conformers. *Biophysical Journal* 92(11):3817-3829.

188. Mechelke M & Habeck M (2010) Robust probabilistic superposition and comparison of protein structures. *BMC Bioinformatics* 11(1):363.

189. Maiorov VN & Crippen GM (1995) Size-independent comparison of protein three-dimensional structures. *Proteins: Structure, Function, and Bioinformatics* 22(3):273-283.

190. Skjaerven L, Martinez A, & Reuter N (2011) Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins: Structure, Function, and Bioinformatics* 79(1):232-243.

191. Amadei A, Linssen A, & Berendsen HJ (1993) Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics* 17(4):412-425.

192. Hubbard RE & Kamran Haider M (2001) Hydrogen bonds in proteins: role and strength. *Encyclopedia of Life Science*:1–6.

193. Durrant JD & McCammon JA (2011) Molecular dynamics simulations and drug discovery. *BMC Biology* 9(1):71.

194. Donald JE, Kulp DW, & DeGrado WF (2011) Salt bridges: Geometrically specific, designable interactions. *Proteins: Structure, Function, and Bioinformatics* 79(3):898-915.

195. Rüdel S*, et al.* (2010) Phosphorylation of human Argonaute proteins affects small RNA

binding. *Nucleic Acids Research* 39:2330-2343.

196. Hayward S, Kitao A, & Berendsen HJ (1997) Model-free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins: Structure, Function, and Bioinformatics* 27(3):425-437.

197. Tramontano A (1998) Homology modeling with low sequence identity. *Methods* 14(3):293-300.

198. Kalia M & Kukol A (2011) Structure and dynamics of the kinase IKK-β–A key regulator of the NF-kappa B transcription factor. *Journal of Structural Biology* 176(2):133-142.

199. Kwak PB & Tomari Y (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature Structural & Molecular Biology* 19(2):145-151.

200. Wang B*, et al.* (2009) Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. *Nature Structural & Molecular Biology* 16(12):1259-1266.

201. Dunoyer P*, et al.* (2010) Small RNA duplexes function as mobile silencing signals between plant cells. *Science* 328(5980):912-916.

202. Chen Z*, et al.* (2013) Hypoxia-responsive miRNAs target argonaute 1 to promote angiogenesis. *The Journal of Clinical Investigation* 123(3):1057.

203. Olovnikov I, Chan K, Sachidanandam R, Newman DK, & Aravin AA (2013) Bacterial argonaute samples the transcriptome to identify foreign DNA. *Molecular Cell* 51(5):594-605.

204. Hur JK, Olovnikov I, & Aravin AA (2014) Prokaryotic Argonautes defend genomes against invasive DNA. *Trends in Biochemical Sciences* 39(6):257-259.

Chapter 7. APPENDIX

7.1. Python script to align multiple sequences

```
from modeller import *
log.verbose()
env = environ()
env.io.atom_files_directory = '.'
aln = alignment(env)
for (code, chain) in (('4NCB','A'),('1U04','A'),('2YHA','A'),('2YHB','A'),('3QIR','A')):
    mdl = model(env, file=code, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(mdl, atom_files=code, align_codes=code+chain)
for (weights, write_fit, whole) in (((1., 0., 0., 0., 1., 0.), False, True),
                                    ((1., 0.5, 1., 1., 1., 0.), False, True),
                                    ((1., 1., 1., 1., 1., 0.), True, False)):
    aln.salign(rms_cutoff=3.5, normalize_pp_scores=False,
               rr_file='$(LIB)/as1.sim.mat', overhang=30,
               gap_penalties_1d=(-450, -50),
               gap_penalties_3d=(0, 3), gap_gap_score=0, gap_residue_score=0,
               dendrogram_file='ALIGNS.tree',
               alignment_type='tree',
               feature_weights=weights, # For a multiple sequence alignment only
               improve_alignment=True, fit=True, write_fit=write_fit,
               write_whole_pdb=whole, output='ALIGNMENT QUALITY')
aln.write(file='aligns.pap', alignment_format='PAP')
```

```
aln.write(file='aligns.ali', alignment_format='PIR')
aln.salign(rms_cutoff=1.0, normalize_pp_scores=False,
        rr_file='$(LIB)/as1.sim.mat', overhang=30,
        gap_penalties_1d=(-450, -50), gap_penalties_3d=(0, 3),
        gap_gap_score=0, gap_residue_score=0, dendrogram_file='1is3A.tree',
        alignment_type='progressive', feature_weights=[0]*6,
        improve_alignment=False, fit=False, write_fit=True,
        write_whole_pdb=False, output='QUALITY')
```

7.2. Python script to align the query and template sequences

```
from modeller import *

log.verbose()

env = environ()

env.libs.topology.read(file='$(LIB)/top_heav.lib')

# Read aligned structure(s):

aln = alignment(env)

aln.append(file='fm00495.ali', align_codes='all')

aln_block = len(aln)

# Read aligned sequence(s):

aln.append(file='hAgo2.ali', align_codes='hAgo2')

# Structure sensitive variable gap penalty sequence-sequence alignment:

aln.salign(output='', max_gap_length=20,

        gap_function=True,   # to use structure-dependent gap penalty

        alignment_type='PAIRWISE', align_block=aln_block,

        feature_weights=(1., 0., 0., 0., 0., 0.), overhang=0,

        gap_penalties_1d=(-450, 0),

        gap_penalties_2d=(0.35, 1.2, 0.9, 1.2, 0.6, 8.6, 1.2, 0., 0.),

        similarity_flag=True)

aln.write(file='hAgo2-mult.ali', alignment_format='PIR')

aln.write(file='hAgo2-mult.pap', alignment_format='PAP')
```

7.3. Pthong script to predict the homology model using multiple templates

```
from modeller import *
from modeller.automodel import *


log.verbose()
env = environ()
env.io.atom_files_directory = ['.']
env.io.hetatm = False                          # if you want to model a cofactor
a = automodel(env,
        alnfile  = 'hAgo2-mult.ali',           # alignment filename
        knowns     =('4NCBA','1U04A','2YHAA','2YHB A','3LUCA',),
                  sequence = 'hAgo2',          # code of the target
                       assess_methods = assess.normalized_dope)


a.starting_model= 1                            # index of the first model
a.ending_model  = 100                          # index of the last model
a.library_schedule = autosched.slow
a.max_var_iterations = 300
# Thorough MD optimization:
a.md_level = refine.slow
a.make()         # do the actual homology modeling
```

7.4. Python script to model the missing protein residues

```
from modeller import *
from modeller.automodel import *    # Load the automodel class

    log.verbose()
    env = environ()
    env.io.atom_files_directory = ['.']

   class MyModel(automodel):
    def select_atoms(self):
        return selection(self.residue_range('120', '126'),
                         self.residue_range('186', '188'))

   a = MyModel(env, alnfile = 'alignment.ali',
            knowns = '4f3t', sequence = '4f3t_fill')
a.starting_model= 1
a.ending_model  = 1


a.make()
```

7.5. Parameters to energy minimize MD systems in vacuum

Define can be used to control processes

```
define          = -DFLEXIBLE
```

Parameters describing what to do, when to stop and what to save

```
integrator      = steep         ; Algorithm (steepest descent minimization)
emtol           = 10.0          ; Stop min when the max force < 1.0 kJ/mol
nsteps          = 50000         ; Maximum number of (min) steps to perform
nstenergy       = 1             ; Write energies to disk every nstenergy steps
energygrps      = System        ; Which energy group to write to disk
```

Parameters describing how to find the neighbors of each atom and how to calculate the interactions

```
nstlist         = 1             ; Frequency to update the neighbor list
ns_type         = grid          ; Method to determine neighbor list (grid)
coulombtype     = PME           ; long range electrostatic interactions
rlist           = 1.0           ; Cut-off for making neighbor list
rcoulomb        = 1.0           ; long range electrostatic cut-off
rvdw            = 1.0           ; long range Van der Waals cut-off
constraints     = none          ; Bond types to replace by constraints
pbc             = xyz           ; Periodic Boundary Conditions (yes)
```

7.6. Parameters for the energy minimization of the solvated MD systems

```
; Define can be used to control processes
define          = -DFLEXIBLE

; Parameters describing what to do, when to stop and what to save
integrator      = steep         ; Algorithm (steep = steepest descent minimization)
emtol           = 1.0           ; Stop minimization when the maximum force < 1.0 kJ/mol
nsteps          = 5000          ; Maximum number of (minimization) steps to perform
nstenergy       = 1             ; Write energies to disk every nstenergy steps
energygrps      = System        ; Which energy group(s) to write to disk

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist         = 1             ; Frequency to update the neighbor list
ns_type         = grid          ; Method to determine neighbor list (grid)
coulombtype     = PME           ; Treatment of long range electrostatic interactions
rlist           = 1.0           ; short-range neighborlist cutoff (in nm)
rcoulomb        = 1.0           ; long range electrostatic cut-off
rvdw            = 1.0           ; long range Van der Waals cut-off
constraints     = none          ; Bond types to replace by constraints
pbc             = xyz           ; Periodic Boundary Conditions (yes)
```

7.7. Parameters for position restrained equilibration of the MD systems

```
; PREPROCESSING OPTIONS
define                  = -DPOSRES

; RUN CONTROL PARAMETERS
integrator              = md
dt                      = 0.002         ; time step (in ps)
nsteps                  = 50,000; number of steps

; OUTPUT CONTROL OPTIONS
nstxout                 = 500   ; save coordinates every ps
nstvout                 = 500   ; save velocities every ps
nstenergy               = 500   ; save energies every ps
nstlog                  = 500   ; update log file every ps
energygrps              = Protein Non-Protein

; NEIGHBORSEARCHING PARAMETERS
nstlist                 = 5
ns_type                 = grid
pbc                     = xyz
rlist                   = 1.0

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype             = PME   ; Particle Mesh Ewald for long-range electrostatics
pme_order               = 4     ; cubic interpolation
fourierspacing          = 0.16  ; grid spacing for FFT
```

```
rcoulomb                = 1.0
vdw-type                = Cut-off
rvdw                    = 1.0

; TEMPERATURE COUPLING
tcoupl                  = v-rescale             ;Berendsen method
tc-grps                 = Protein  Non-Protein  ; separate heat baths
tau_t                   = 0.1      0.1          ;Coupling time constant,
ref_t                   = 310      310          ; Temperature of heat bath
pcoupl                  = no                    ; no pressure coupling in NVT

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel                 = yes   ; Assign velocities to particles by taking them
                                randomly from a Maxwell distribution
gen_temp                = 310   ; Temperature to generate corresponding
                                Maxwell distribution
gen_seed                = -1    ; Seed for (semi) random number generation.

; OPTIONS FOR BONDS
constraints             = all-bonds     ; All bonds treated as constraints (fixed length)
continuation            = no            ; first dynamics run
constraint_algorithm    = lincs         ; holonomic constraints
lincs_iter              = 1             ; accuracy of LINCS
lincs_order             = 4             ; also related to accuracy
```

7.8. Parameters for equilibration of the MD systems

```
; PREPROCESSING OPTIONS
define                  = -DPOSRES

; RUN CONTROL PARAMETERS
integrator              = md
dt                      = 0.002
nsteps                  = 50000

; OUTPUT CONTROL OPTIONS
nstxout                 = 500      ; save coordinates every ps
nstvout                 = 500      ; save velocities every ps
nstfout                 = 500      ; save forces every ps
nstenergy               = 500      ; save energies every ps
nstlog                  = 500      ; update log file every ps
energygrps              = Protein Non-Protein

; NEIGHBORSEARCHING PARAMETERS
nstlist         = 5
ns-type         = Grid
pbc             = xyz
rlist           = 1.0

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype     = PME
pme_order       = 4             ; cubic interpolation
fourierspacing  = 0.16          ; grid spacing for FFT
```

```
rcoulomb            = 1.0
vdw-type            = Cut-off
rvdw                = 1.0

; TEMPERATURE COUPLING
Tcoupl              = v-rescale
tc-grps             = Protein  Non-Protein
tau_t               = 0.1      0.1
ref_t               = 310      310

; PRESSURE COUPLING
Pcoupl              = Berendsen
Pcoupltype          = Isotropic
tau_p               = 1.0
compressibility     = 4.5e-5
ref_p               = 1.0
refcoord_scaling    = COM

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel             = no
; OPTIONS FOR BONDS
constraints             = all-bonds
continuation            = yes        ; continuation from NVT
constraint_algorithm    = lincs      ; holonomic constraints
lincs_iter              = 1          ; accuracy of LINCS
lincs_order             = 4          ; also related to accuracy
```

7.9. Parameters for unrestrained equilibration

```
;RUN CONTROL PARAMETERS
integrator              = md
dt                      = 0.002
nsteps                  = 10000

; OUTPUT CONTROL OPTIONS
nstxout                 = 500     ; save coordinates every ps
nstvout                 = 500     ; save velocities every ps
nstfout                 = 500     ; save forces every ps
nstenergy               = 500     ; save energies every ps
nstlog                  = 500     ; update log file every ps
energygrps              = Protein Non-Protein

; NEIGHBORSEARCHING PARAMETERS
nstlist                 = 5
ns-type                 = Grid
pbc                     = xyz
rlist                   = 1.0

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype       = PME
pme_order         = 4             ; cubic interpolation
fourierspacing    = 0.16          ; grid spacing for FFT
rcoulomb          = 1.0
vdw-type          = Cut-off
```

```
rvdw                    = 1.0

; TEMPERATURE COUPLING
Tcoupl                  = v-rescale
tc-grps                 = Protein  Non-Protein
tau_t                   = 0.1      0.1
ref_t                   = 310      310

; PRESSURE COUPLING
Pcoupl              = Berendsen
Pcoupltype          = Isotropic
tau_p               = 1.0
compressibility     = 4.5e-5
ref_p               = 1.0

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel             = no

; OPTIONS FOR BONDS
constraints             = all-bonds
continuation            = yes               ; continuation from NPT PR10
constraint_algorithm    = lincs             ; holonomic constraints
lincs_iter                      = 1         ; accuracy of LINCS
lincs_order                     = 4         ; also related to accuracy
```

7.10. Parameters for the final production run in MD

```
; RUN CONTROL PARAMETERS
integrator          = md
tinit               = 0             ; Starting time
dt                  = 0.002         ; 2 femtosecond time step for integration
nsteps              = 500,000       ; 100 ns

; OUTPUT CONTROL OPTIONS
nstxout             = 250000        ; Writing full precision coordinates every 0.5 ns
nstvout             = 250000        ; Writing velocities every 0.5 ns
nstlog              = 5000          ; Writing to the log file every 10ps
nstenergy           = 5000          ; Writing out energy information every 10ps
nstxtcout           = 5000          ; Writing coordinates every 10ps
energygrps          = Protein Non-Protein
; NEIGHBORSEARCHING PARAMETERS
nstlist             = 5
ns-type             = Grid
pbc                 = xyz
rlist               = 1.0
; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype         = PME
pme_order       = 4             ; cubic interpolation
fourierspacing = 0.16           ; grid spacing for FFT
rcoulomb            = 1.0
vdw-type            = Cut-off
```

```
rvdw                    = 1.0
; TEMPERATURE COUPLING
Tcoupl                  = v-rescale
tc-grps                 = Protein  Non-Protein
tau_t                   = 0.1      0.1
ref_t                   = 310      310
; PRESSURE COUPLING
Pcoupl                  = Berendsen
Pcoupltype              = Isotropic
tau_p                   = 1.0
compressibility         = 4.5e-5
ref_p                   = 1.0
; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel                 = no
; OPTIONS FOR BONDS
constraints             = all-bonds
constraint-algorithm    = lincs
continuation            = yes                   ; Restarting after NPT without position restraints
lincs-order             = 4
lincs-iter              = 1
lincs-warnangle         = 30
```

7.11. Script used to run MD simulations at the HPC centers

```
#!/bin/ksh
#PBS -j oe
#PBS -N gromacs
#PBS -l feature=ice1
#PBS -M kalia@imm.uni-luebeck.de
#PBS -q mediumq
#PBS -l naccesspolicy=singlejob
#PBS -l nodes=16:ppn=8
#PBS -l walltime=12:00:00
#### preliminaries ####
  eval `/sw/swdist/bin/modulesinit`
  export MPI_GROUP_MAX=128
  module load gromacs/4.6.3
  module load mpt
cd $PBS_O_WORKDIR
#### start parallel GROMACS Job ####
mpiexec mdrun_mpi -s ago_md.tpr -o ago_md.trr -x ago_md.xtc -c ago_md.gro -cpi
ago_20.cpt
-cpo ago_36.cpt -e ago.edr -g md.log
```

Acknowledgements

I first whole-heartedly thank Prof. Tobias Restle for supervising and guiding me during my Ph.D. I have thoroughly enjoyed working and learning from you, this experience would certainly leave a lasting impression on my scientific career. I am highly indebted for your support and understanding during past three years. I would also like to acknowledge Prof. Georg Sczakiel and Dr. Jens-Christian Claussen for their guidance and support during the course of my Ph.D.

I am also very grateful to Prof. Piotr Sliz at Harvard Medical School for granting me an opportunity to work with him. I would also like to express my profound gratitude towards Prof. Alexandre M.J.J. Bonvin, at Utrecht University for a very fruitful 'HPC-Europa2' grant application and hosting me at his laboratory. I had a great opportunity to gain additional computational skills, which have been crucial for my research work during my Ph.D. I am also greatly indebted to Prof Collin M. Stultz at MIT for collaborating with us and hosting me at his laboratory. I have learnt a lot from you, most of all I have learned to be critical of my own work.

I would not have been able to enjoy and succeed during my Ph.D. without the support of my colleagues, most of which ended up being my good friends. First, I would like to thank Sarah for performing all the biochemical studies, which have been mentioned in my thesis. You have been a great colleague, but most of all a genuine friend. I would also like to acknowledge Simon and Maria for their insightful scientific discussions, moral support and enabling to have a memorable time in Lübeck. I would also like to thank Patrick, Juliana, Felix and Yang for their feedback on my results and suggestions. I would also like to thank Alessandra and Rosel for their feedback, discussions and overall warm heartedness. I am also very thankful to Petra for her help with the paper work, German translation and lovely small talks.

I am also very thankful to all my colleagues at the Prof. Bonvin's lab; I had the opportunity of making new friends. I am thankful to Marc for answering my non-stop questions, Adrien for his in-depth knowledge in Gromacs and giving me crucial insights into the French culture. I would really like to thank João for helping me out with the modelling of missing nucleotides in the RNA. I am also very grateful to Fariha for enlightening me about NMR, but most of all for bringing home closer to Utrecht and the constant moral and emotional support.

I would also like to thank all the members of Prof. Stultz lab, I learned a lot from everyone. I am thankful to Thomas, Orly and Yun with helping me to set up everything on my arrival in the lab. I am also thankful to all of them for their critical feedback and detailed discussions about my work. I am also thankful to Linder for helping me out with Charmm. I would also like to thank Virginia for her feedback on my research and reviving coffee breaks, which were imperative during the thesis, write up.

I am indebted to Christian Tuma from the HLRN supercomputing center for helping me out in the dire straits and providing me the computing time to run my simulations. I am also grateful to the Graduate school for computing in medicine and life sciences for the financial support and paper work.

In the end, I am extremely thankful to my friends and family for their constant support and unconditional love.

Curriculum Vitae

EDUCATION

2011 – 2014	Institute of Molecular Medicine, *University of Luebeck, Germany*
Ph.D.	PI: Prof. Dr. Tobias Restle
2010 – 2011	School of Life Sciences, *University of Hertfordshire, UK*
M.Sc. (Research)	PI: Dr. Andreas Kukol
2006 – 2009 B.Sc	*Punjab University, Chandigarh, India*

RESEARCH EXPERIENCE

2013 – 2014	Computational Biophysics Laboratory, *MIT, USA* Prof. Collin M. Stultz Performed implicit and explicit molecular dynamics simulations of range of argonaute proteins
2011 – 2014	Institute of Molecular Medicine, *University of Luebeck, Germany* PI: Prof. Dr. Tobias Restle Thesis: Insights into the catalytic function of human Argonaute2 protein through molecular dynamics simulations
2012 – 2013	Bijvoet center, *Utrecht University, The Netherlands* Prof. Alexandre M.J.J.Bonvin Applied for the HPC-Europa2 European grant and performed molecular dynamics simulations of human argonaute2 – miRNA

complex at the SARA (Supercomputing center), Amsterdam

2011 – 2012 Structural Biology & Computing Lab, *Harvard Medical School, MA, USA*, Prof. Piotr Sliz
Performed long timescale (microseconds), molecular dynamics simulations of lin28-let7 (protein-miRNA) complex at TACC supercomputing center

2009 – 2011 School of Life Sciences, *University of Hertfordshire, UK*
Dr. Andreas Kukol
Predicted a comparative model of human kinase IKK-β protein, identified a cryptic binding site through molecular dynamics & performed virtual screening of novel IKK-β inhibitors

2008 *Indian Agriculture Research Institute, Shimla, India*
Dr. D. K. Kishore
Performed tissue culture of temperate fruit crops

GRANTS & AWARDS

2012 HPC-Europa2 (European Translational access grant) in collaboration with Prof. Alexandre M.J.J.Bonvin Utrecht University, The Netherlands

2012 CECAM travel award, EPFL, Lausanne, Switzerland

2010 Travel award, **'**Celebrating Computational Biology Conference: A tribute to Frank Blaney', Oxford University, UK

TEACHING EXPERIENCE

2011 Tutorial 'Introduction to COPASI' at 'Systems Biology Workshop', *University of Luebeck, Germany*

Sep2010 - April2011	School of Life Sciences, *University of Hertfordshire, UK* Teaching Assistant for the Bioinformatics & Molecular Biology in the laboratory classes

ATTENDED WORKSHOPS

2013	'MOLSIM', *University of Amsterdam, The Netherlands*
2012	'Tutorial for the AMBER set of modelling tools', *CECAM- HQ-EPFL,* *Lausanne, Switzerland*
2011	'Systems Biology Workshop', *University of Luebeck, Germany*
2010	'MicroRNAs and their targets: promises and pitfalls', *London, UK*

List of Publications

1. **Kalia M.**, Willkomm S., Claussen J., Bonvin A.M.J.J., Stultz C.M., Restle T., 'A single point mutation in L2 linker domain abolishes the RNAase activity in human argonaute-2 protein' (*In preparation*)
2. **Kalia M.,** Willkomm S., Claussen J., Bonvin A.M.J.J., Restle T., 'Insights into the 5'-nucleotide bias of human Argonaute 2 protein through molecular dynamics simulations' (*In preparation*)
3. **Kalia M.**, Restle T., Stultz C.M., 'Molecular mechanics of the Argonaute proteins' (*In preparation*)
4. **Kalia M**., Kukol A., 2011, 'Structure and dynamics of the kinase IKK-β – A key regulator of the NF-kappa B transcription factor' *Journal of Structural Biology*, vol 176, no. 2, pp. 133-142.

TALKS

Kalia M., Willkomm S., Claussen J., Bonvin A.M.J.J., Stultz C.M., Restle T., 'A single inter-domain salt bridge within the human Argonaute 2 protein crucially affects protein folding and consequently enzymatic activity', *58th Biophysical Annual Meeting*, 2014, San Francisco, USA

Kalia M., Kukol A., 'Homology Modeling of IKK-β and prediction of its novel inhibitors', *Life Science Research Day*, 2010, University of Hertfordshire, UK

POSTERS

Kalia M., Willkomm S., Claussen J., Restle T., 'Comparative structure and molecular dynamics of TAR RNA-binding protein (TRBP); A keyplayer in RNA interference pathway', GLBIO conference, 2012, University of Michigan, Ann Arbor, USA

Kalia M., Kukol A., 'Host cell proteins as novel drug targets against influenza A virus', Celebrating Computational Biology Conference: A tribute to Frank Blaney, 2010, Oxford University, UK