**Technische Universität München**

Lehrstuhl für Biomolekulare NMR-Spektroskopie

Bayerisches NMR-Zentrum

Chemie Department

# NMR studies of dynamic double-stranded RNA recognition by RNA binding proteins

**Jan-Niklas Tants**

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

| | |
|---|---|
| Vorsitzende: | Prof. Dr. Kathrin Lang |
| Prüfer der Dissertation: | 1. Prof. Dr. Michael Sattler |
| | 2. Prof. Dr. Klaus Förstemann |

Die Dissertation wurde am 26.06.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 20.12.2017 angenommen.

„To see the world,
things dangerous to come to,
to see behind walls, draw closer,
to find each other, and to feel.
That is the purpose of life.“

(James Thurber, The Secret Life of Walter Mitty)

# Abstract

Ribonucleoproteins (RNP) comprising a protein and a nucleic acid component are diverse complexes involved in various regulatory processes within a cell. Protein-RNA interaction usually is key for proper function and the recognition of RNA can range from complete unspecific binding to high sequence or shape specificity. For this purpose distinct RNA recognition motifs have evolved and are either an integral part of a larger protein or protein complex or a solely RNA binding protein (RBP). Many RNA binding proteins comprise multiple RNA binding domains (RBD) and considerable dynamics within the protein and RNPs are often an intrinsic and important feature.

In this thesis a major theme is double-stranded RNA (dsRNA) recognition by different proteins. For this work NMR is used as a powerful tool to obtain information especially on dynamic systems and is combined with other biophysical techniques.

RNA interference (RNAi) is a pivotal tool for cells to regulate protein synthesis posttranscriptionally either via mRNA cleavage or inhibition of translation. A double-stranded RNA is processed by tightly regulated protein complexes into a single-stranded effector short interfering (siRNA) or microRNA (miRNA). The RNA induced silencing complex (RISC) comprising an Argonaute protein (Ago) and the single-stranded RNA (ssRNA) binds sequence-specifically to RNA targets and initiates subsequent degradation, inhibits translation or serves in viral defense. Important factors in processing of dsRNA are double-stranded RNA binding proteins (dsRBP). These multidomain proteins exhibit different functions and show a dynamic and usually sequence-independent binding. In the main project of this thesis I show that Loquacious-PD (Loqs-PD), a *Drosophila* dsRBP, could serve as a sensor for siRNA asymmetry which is a crucial step when converting dsRNA into single-stranded siRNA. Furthermore, I confirm the dynamic binding of Loqs-PD to its RNA substrates and provide a semi-quantitative NMR-based analysis of sliding.

Another possibility to control translation is directed messenger RNA (mRNA) localization. Using motor proteins and complex networks of microtubules and filaments as cellular highways mRNAs are transported to designated cellular compartments which allows spatial control of protein expression. One protein involved in mRNA transport within the

nervous system is the dsRBP Staufen-2 of *M. musculus*. It is crucial for memory formation and recognizes stem loop structures within the RNAs as localization signals. In this thesis I could confirm that double-stranded RNA binding domain 1 (dsRBD1) and 2 of Staufen-2 are independent of each other and contribute to dsRNA binding. Additionally, binding seems to be accompanied by dynamics as shown before for Loqs.

Besides RNA interference and RNA localization the degradation of RNA is another option for translational control. Distinct RNA degradation pathways exist that strictly control RNA turnover but are also crucial for processing of functional precursor RNAs: The *S. cerevisiae* exosome contributes to ribosomal precursor tRNA maturation. As an unspecific degradation machinery the exosome requires tight control and regulation by various co-factors. One of these, the helicase Mtr4, enables the exosome to degrade and process structured RNAs after unwinding of the substrates. Mtr4 comprises an usual insertion which contributes to RNA binding. In this work I prove that this so called KOW domain binds structured RNAs and tRNAs. Binding to Nop53, an exosome recruiter that directs the complex to the ribosome, occurs via a distinct interface and allows simultaneous binding of both ligands to the domain.

# Zusammenfassung

Ribonukleoproteine (RNP) bestehen aus einem Protein und einer Nukleinsäure-Komponente und bilden vielfälftige Komplexe die in unterschiedlichen regulatorischen Prozessen der Zelle involviert sind. Üblicherweise sind Protein-RNA-Interaktionen entscheidend für die Funktion der Komplexe und die RNA-Erkennung variiert von unspezifischer Bindung bis hin zu hoher Sequenz- oder Struktur-Spezifizität. Daher entwickelten sich differenzierte RNA-Erkennungs-Motive, die entweder ein integraler Bestandteil eines größeren Proteins bzw. Proteinkomplexes sind oder einzig ein RNA-bindendes Protein (RBP) darstellen. Viele RNA-bindende Proteine beinhalten mehrere RNA-bindende Domänen (RBD) und dynamische Veränderungen innerhalb der Proteine und Ribonukleoproteine sind wichtige intrinsische Eigenschaften.

Das Hauptthema dieser Thesis ist die Erkennung doppelsträngiger RNA (dsRNA) durch verschiedene Proteine. Für diese Arbeit wurde NMR als ideales Werkzeug für dynamische Systeme eingesetzt um Informationen über die Komplexe zu erhalten und wurde durch weitere biophysikalische Methoden ergänzt.

RNA-Interferenz (RNAi) ist ein wichtiges Werkzeug für Zellen um die Proteinsynthese posttranskriptionell zu regulieren. Dies geschieht entweder durch den Abbau von mRNA oder Inhibition der Translation. Eine doppelsträngige RNA wird durch streng regulierte Proteinkomplexe in eine einzelsträngige, funktionsfähige, kurze interferierende (siRNA) oder microRNA (miRNA) umgewandelt. Der RNA-induzierte Silencing Komplex (RISC) besteht aus einem Argonautenprotein (Ago) und dieser einzelsträngigen RNA (ssRNA) und bindet sequenzspezifisch an Ziel-RNAs. Dadurch initiiert der RISC den RNA-Abbau, inhibiert die Translation oder fungiert als Abwehr gegen Viren. Wichtige Faktoren für die Prozessierung der dsRNA sind doppelsträngige-RNA-bindende Proteine (dsRBP). Diese Multidomänen-Proteine haben vielfältige Funktionen und binden üblicherweise sequenzunabhängig und dynamisch an RNAs. In dem Hauptprojekt dieser Arbeit zeige ich, dass Loquacious-PD (Loqs-PD), ein dsRBP aus der Fruchtfliege *Drosophila*, als Sensor für siRNA-Asymmetrie dienen kann. Dies ist ein entscheidender Schritt bei der Umwandlung von doppelsträngiger zu einzelsträngiger siRNA. Des Weiteren konnte ich das dynamische Binden von Loqs-PD an RNA-

Substraten belegen und stelle eine semi-quantitative NMR-basierte Analyse des sogenannten Slidings vor.

Eine weitere Möglichkeit die Translation zu kontrollieren stellt gerichtete Lokalisierung von messenger RNA (mRNA) dar. Mit Hilfe von Motorproteinen und komplexen Netzwerken aus Mikrotubuli und Filamenten als einer Art zellulärer Schnellstraβe werden mRNAs in bestimmte zelluläre Bereiche transportiert, was die räumliche Kontrolle der Proteinexpression erlaubt. Ein Protein im mRNA-Transport im Nervensystem ist das dsRBP Staufen-2 aus der Hausmaus *M. musculus*, welches wichtig ist für die Gedächtnisbildung. Staufen-2 erkennt Stamm-Schlaufen-Strukturen innerhalb der RNAs als Lokalisierungssignal. In dieser Thesis konnte ich bestätigen, dass die doppelsträngige-RNA-bindende Domänen 1 (dsRBD1) und 2 von Staufen-2 unabhängig voneinander sind und beide dsRNA binden. Zusätzlich scheint die RNA-Bindung, wie zuvor für Loqs beschrieben, dynamisch zu sein.

Neben RNA-Interferenz und der Lokalisierung von RNA bildet der RNA-Abbau eine weitere Möglichkeit der translationalen Kontrolle. Es existieren verschiedene RNA-Abbau-Wege die den RNA-Umsatz kontrollieren und an der Prozessierung von funktionalen RNA-Vorläufern beteiligt sind: Das Exosom der Bäckerhefe *S. cerevisiae* trägt zur Reifung ribosomaler Vorläufer-tRNAs bei. Als eine unspezifische Degradations-Maschinerie muss das Exosom streng durch zahlreiche Kofaktoren kontrolliert werden. Einer dieser Faktoren, die Helikase Mtr4, ermöglicht dem Exosom nach Entfaltung der Substrate den Abbau strukturierter RNAs. Mtr4 beinhaltet eine ungewöhnliche Einfügung die ebenfalls zur RNA-Bindung beiträgt. In dieser Arbeit belege ich, dass diese sogenannte KOW-Domäne strukturierte RNAs und tRNAs bindet. Die Bindung an Nop53, welches das Exosom rekrutiert und zum Ribosom führt, erfolgt über eine andere Stelle an KOW und ermöglicht dadurch das gleichzeitige Binden beider Liganden an KOW.

# Table of Contents

Abstract ..... I
Zusammenfassung ..... III
Table of Contents ..... V
List of Figures ..... IX
List of Tables ..... XI
Abbreviations ..... XII
1 Introduction ..... 1
1.1 RNA interference ..... 2
1.1.1 Formation of RISC and RNA interference ..... 2
1.1.2 Protein key players in RNA interference ..... 5
1.1.3 Double-stranded RNA binding proteins – features and functions ..... 6
1.1.4 The isoform Loquacious-PD is specific for endogenous siRNA maturation in *Drosophila* ..... 7
1.2 RNA localization ..... 10
1.2.1 Mechanisms of RNA localization ..... 10
1.2.2 Disease related RNA binding proteins in RNA localization ..... 11
1.2.3 Staufen – a dsRNA recognizing protein crucial for brain development ..... 11
1.3 RNA degradation ..... 13
1.3.1 RNA degradation pathways and the importance of the exosome for RNA homeostasis ..... 13
1.3.2 The exosome is highly regulated by *trans*-acting factors ..... 15
1.3.3 The positive exosome regulator Mtr4 unwinds RNA substrates ..... 16
1.4 NMR spectroscopy ..... 19
1.4.1 A tool for structural biology ..... 19
1.4.2 Theory of NMR spectroscopy ..... 19
1.4.3 Relaxation processes ..... 23
1.4.4 Benefits and mechanisms of magnetization transfer ..... 25
1.4.5 Water suppression techniques ..... 27
1.4.6 2D homonuclear NMR spectroscopy ..... 28
1.4.7 2D heteronuclear NMR spectroscopy ..... 29
1.4.8 3D NMR – protein backbone assignment experiments ..... 29
1.4.9 Paramagnetic Relaxation Enhancement ..... 31
2 Scope of the Thesis ..... 33
3 Materials and Methods ..... 34
3.1 Chemicals ..... 34

3.2 Consumables ..... 35
3.3 Devices ..... 35
3.4 NMR Spectrometers ..... 36
3.5 Softwares and webservers ..... 37
3.5.1 Software ..... 37
3.5.2 Webservers ..... 37
3.6 Buffers, solutions, templates and constructs ..... 37
3.6.1 Vectors ..... 37
3.6.2 *E. coli* strains ..... 38
3.6.3 Media and plates ..... 38
3.6.3 Primer for linker mutant cloning ..... 39
3.6.4 Protein constucts used ..... 40
3.6.5 Buffers and solutions for protein purification ..... 41
3.6.6 In vitro transcription templates ..... 42
3.6.7 Buffers and solutions for RNA transcription and purification ..... 43
3.6.8 Buffers for NMR ..... 43
3.7 Molecular biology ..... 44
3.7.1 Cell culture for cloning and plasmid isolation ..... 44
3.7.2 Minipreparation of plasmids ..... 44
3.7.3 Cloning of Loqs-PD$_{\Delta NC}$ linker mutants ..... 44
3.8 Protein expression and purification of Loqs ..... 45
3.8.1 Cell culture for protein expression ..... 45
3.8.2 Cell lysis ..... 45
3.8.3 Protein purification ..... 45
3.8.4 SDS-PAGE ..... 46
3.8.5 Protein quantification ..... 46
3.9 RNA transcription and purification ..... 46
3.9.1 RNA in vitro transcription optimization ..... 46
3.9.2 RNA large scale in vitro transcription ..... 47
3.9.3 RNA purification via denaturing HPLC ..... 47
3.9.4 RNA purification via denaturing polyacrylamide gel ..... 47
3.9.5 Electroelution of RNA gels ..... 47
3.9.6 Dialysis and lyophylization of RNAs ..... 48
3.9.7 RNA quantification ..... 48
3.10 Biophysical experiments ..... 48
3.10.1 Static light scattering ..... 48

3.10.2 Small angle X-ray scattering ..... 49
3.10.3 Circular dichroism spectroscopy ..... 49
3.11 NMR spectroscopy ..... 50
3.11.1 NMR sample preparation ..... 50
3.11.2 RNA assignment ..... 50
3.11.3 Protein backbone assignment ..... 50
3.11.4 Titrations and calculation of chemical shift perturbations ..... 51
3.11.5 Paramagnetic relaxation enhancement experiments and analysis ..... 51
3.11.6 Line broadening experiments and analysis ..... 52
3.11.7 Heteronulcear NOE experiments and data analysis ..... 52
4 Results ..... 54
4.1 SiRNA binding of Loqs-PD ..... 54
4.1.1 Summary of previous results ..... 54
4.1.2 In vitro transcription and analysis of 21 bp siRNA ..... 54
4.1.3 Loqs-PD$_{\Delta NC}$ binds dsRNA in a 1:1 complex ..... 58
4.1.4 Both dsRBDs bind RNA independently and compete for binding ..... 59
4.1.5 The linker does not affect Loqs' domain structure ..... 61
4.1.6 Linker mutants bind RNA via conserved binding interface and interact with Dicer ..... 62
4.1.7 Loqs stabilizes the RNA duplex and has different effects on both strands ..... 64
4.1.8 Intermolecular PREs confirm dynamics and suggest a weak asymmetric binding ..... 69
4.1.9 The C-terminus mediates dimerization of Loqs and interaction with Dicer to form an asymmetric complex ..... 72
4.2 Dynamic dsRNA binding by Loqs-PD ..... 75
4.2.1 Chemical exchange as a tool to study dynamics ..... 75
4.2.2 Transcription and analysis of palindromic RNAs ..... 75
4.2.3 Single and tandem dsRBDs slide along dsRNA ..... 77
4.2.4 Sliding is independent of RNA sequence composition ..... 80
4.2.5 The linker does not affect sliding ..... 81
4.2.6 Elevated temperatures enhance sliding ..... 82
4.2.7 Viscosity limits sliding ..... 85
4.3 Staufen-2 dsRBD1-2 RNA binding ..... 88
4.3.1 Both domains are independent in the free form ..... 88
4.3.2 Structure homology models indicate conserved fold and extended loop in dsRBD2 ..... 88
4.3.3 The two dsRBDs bind dsRNA dynamically ..... 90
4.3.4 Simultaenous binding of both dsRBDs is independent and dynamic ..... 92
4.4 Mtr4 KOW ligand binding ..... 94

4.4.1 Backbone assignment of KOW ..... 94
4.4.2 Solution secondary structure elements confirm crystal structure ..... 95
4.4.3 Structured RNAs and Nop53 bind to distinct interfaces ..... 95
4.4.4 KOW does not bind single-stranded RNA ..... 99
4.4.5 The arch domain interacts with KOW ..... 100
5 Discussion ..... 102
5.1 Loqs-PD contribution to RISC formation and asymmetry sensing ..... 102
5.2 A semi-quantitative look at dsRBD sliding on dsRNA ..... 106
5.3 Staufen-2 dsRBD1 and 2 bind RNA localization signal ..... 109
5.4 The Mtr4 KOW domain contributes to tRNA binding and recruitment to the ribosome ..... 111
6 Appendix ..... 114
6.1 Loqs siRNA binding ..... 114
6.2 Loqs sliding ..... 116
6.3 Staufen 2 ..... 117
6.4 Mtr4 KOW ..... 122
7 References ..... 123
8 Publications and talks ..... 135
8.1 Publications ..... 135
8.2 Talks ..... 135
Acknowledgements ..... 136

# List of Figures

Figure 1.1: RNPs control translation, a key step in protein production ..... 1
Figure 1.2: Schematic pathway of RNA interference ..... 4
Figure 1.3: Pathways of siRNA maturation in Drosophila melanogaster ..... 8
Figure 1.4: Principle of active mRNA transport and local protein expression ..... 11
Figure 1.5: Different eukaryotic RNA degradation pathways ..... 14
Figure 1.6: Scheme of the TRAMP complex ..... 17
Figure 1.7: Schematic principle of NMR spectroscopy ..... 22
Figure 1.8: T1 and T2 relaxation are exponential processes ..... 25
Figure 1.9: Schematic pulse sequence of INEPT magnetization transfer ..... 26
Figure 1.10: 3D experiments for protein backbone assignment ..... 30
Figure 1.11: Principle of paramagnetic relaxation enhancement ..... 32
Figure 4.1: Transcription and purification of 21bp siRNA ..... 55
Figure 4.2: Duplex 21bp siRNA assignment based on homonuclear NOESY spectrum ..... 57
Figure 4.3: Biophysical characterization of Loqs-PD$_{\Delta NC}$ 21bp siRNA complex ..... 59
Figure 4.4: Loqs dsRBDs are independent and compete for RNA binding ..... 60
Figure 4.5: Chemical shift perturbations of Loqs-PD$_{\Delta NC}$ linker mutants ..... 62
Figure 4.6: RNA titrations of Loqs-PD$_{\Delta NC}$ linker mutants ..... 63
Figure 4.7: Co-immunoprecipitation of Loqs-PD wildtype and Loqs-PD$_{\Delta 41}$ with Dicer-2 ..... 64
Figure 4.8: Crosslinking of Loqs-PD constructs to siRNA reveals terminal binding preference ..... 65
Figure 4.9: Protein bound RNA shows severe line broadening ..... 67
Figure 4.10: HNN-COSY quantifies hydrogen bonds within the RNA duplex. ..... 68
Figure 4.11: Intermolecular PRE measurements of Loqs-PD$_{\Delta NC}$ complexed with spin labelled 21bp siRNA ..... 71
Figure 4.12: Effect of N-terminus on domain structures and RNA binding ..... 72
Figure 4.13: Crosslinking of Dicer-2-Loqs-PD complex reveals asymmetric complex formation ..... 74
Figure 4.14: Palindromic RNAs used for line broadening analysis ..... 76
Figure 4.15: Line broadening analysis of single dsRBD RNA titrations ..... 78
Figure 4.16: Line broadening analysis of tandem dsRBD RNA titrations ..... 79
Figure 4.17: Different siRNAs used for line broadening analysis ..... 80
Figure 4.18: Line broadening analysis of Loqs-PD$_{\Delta NC}$ siRNA titrations ..... 81
Figure 4.19: Line broadening analysis of Loqs-PD$_{\Delta NC}$ linker mutant RNA titrations ..... 82
Figure 4.20: Line broadening analysis of Loqs-PD$_{\Delta NC}$ RNA titration at different temperatures ..... 83
Figure 4.21: Inverse line broadening increases at higher temperatures ..... 85
Figure 4.22: Line broadening analysis of Loqs-PD$_{\Delta NC}$ RNA titrations at different glycerol concentrations ..... 86
Figure 4.23: Fingerprint spectra of Staufen-2 constructs ..... 88
Figure 4.24: Structure homology models of Staufen-2 dsRBD1 and 2 ..... 89
Figure 4.25: RNA titrations of Staufen-2 single dsRBD1 and 2 ..... 91
Figure 4.26: Binding site 2 in Staufen-2 dsRBD1 ..... 92
Figure 4.27: Staufen-2$_{\Delta NC}$ RNA titration ..... 93
Figure 4.28: Example walk for KOW backbone assignment ..... 94
Figure 4.29: Secondary structure elements and flexibility of KOW ..... 95
Figure 4.30: KOW titrations with different ligands ..... 97

Figure 4.31: Analysis of simultanoues titration of dsRNA and Nop53 peptide ..... 98
Figure 4.32: KOW titration with single-stranded RNA ..... 99
Figure 4.33: Comparison of KOW with and without the extended stalk ..... 101
Figure 5.1: Binding properties of Loqs-PD and its interaction partners ..... 102
Figure 5.2: Possible function of Loqs-PD during RISC formation ..... 104
Figure 5.3: Effects on sliding of Loqs-PD tandem dsRBDs ..... 107
Figure 5.4: Staufen-2 tandem dsRBDs 1-2 are independent and bind RNA dynamically ..... 109
Figure 5.5: Interaction modes of Mtr4 KOW domain with different ligands ..... 112
Figure 5.6: A model for ribosome precursor-RNA processing via the TRAMP-Exosome complex ..... 113
Supplementary Figure 6.1: $^{1}H,^{15}N$-HSQC overlay of single dsRBDs and mixed ..... 115
Supplementary Figure 6.2: Methyl groups confirm successful reduction of TEMPO spin label ..... 115
Supplementary Figure 6.3: $^{1}H,^{1}H$-NOESYs of palindromic RNAs ..... 116
Supplementary Figure 6.4: $^{1}H,^{1}H$-NOESYs and CD spectra of different siRNAs ..... 116
Supplementary Figure 6.5: Extended data for structure models of Staufen-2 dsRBD1 and 2 ..... 121
Supplementary Figure 6.6: Sequence of Rgs4 SRS2 RNA and $^{1}H,^{1}H$-NOESY ..... 121
Supplementary Figure 6.7: Titrations of KOW domain with different ligands ..... 122
Supplementary Figure 6.8: hetNOE data for KOW in complex with ligands ..... 122

# List of Tables

Table 1.1: Nuclei commonly used in biomolecular NMR spectroscopy and their properties ..... 20
Table 3.1: Vectors used in this thesis ..... 37
Table 3.2: E. coli strains used in this thesis ..... 38
Table 3.3: Media and plates and their applications ..... 38
Table 3.4: Primers used to clone Loqs-PD linker mutants ..... 39
Table 3.5: Overview of all protein constructs used ..... 40
Table 3.6: Buffer composition and solutions for protein purification and analysis ..... 41
Table 3.7: The DNA templates used for RNA in vitro transcription ..... 42
Table 3.8: Composition and application of buffers and solutions used for RNA production and purification ..... 43
Table 3.9: NMR buffer compositions and their applications ..... 43
Table 4.1: SAXS concentration series prove dimerization of Loqs-PD$_{\Delta N}$ ..... 73
Table 4.2: Average line broadening for dsRBD1 and 2 over selected residues at different temperatues ..... 84
Table 4.3: Average line broadening for dsRBD1 and 2 over selected residues at different glycerol concentrations ..... 87
Supplementary Table 6.1: Dali alignments for dsRBD1 and dsRBD2. ..... 114
Supplementary Table 6.2: Sequence alignment of Staufen-2 dsRBD1 and 2 ..... 117

# Abbreviations

| | |
|---|---|
| Å | Ångström |
| ADP | Adenosinediphosphate |
| Ago | Argonaute |
| AIM | Arch interaction motif |
| ALS | Amyotrophe lateral sclerosis |
| Alu | *Arthrobacter luteus* |
| ARE | AU-rich element |
| aRLC | alternative RISC loading complex |
| ATP | Adenosintriphosphate |
| bp | base pair |
| BSA | Bovine serum albumin |
| *C. elegans* | *Caenorhabditis elegans* |
| CCR4-NOT | C-C chemokine receptor type 4 |
| Co-IP | Co-immunoprecipitation |
| COSY | Correlation spectroscopy |
| CUT | cryptic unstable transcript |
| *D. melanogaster* | *Drosophila melanogaster* |
| Dcp 1/2 | *Drosophila* caspase 1/2 |
| DcpS | Decapping scavenger |
| Dcr | Dicer |
| DGCR8 | DiGeorge syndrome critical region gene 8 |
| DHX36 | DEAH-box helicase 36 |
| DNA | desoxyribonucleic acid |
| DSB | double-strand brake |
| dsDNA | double-stranded DNA |
| dsRBD | double-stranded RNA binding domain |
| dsRBP | double-stranded RNA binding protein |
| dsRNA | double-stranded RNA |
| endo-siRNA | endogenous siRNA |
| eRNA | enhancerRNA |
| ESF | exosome specificity factor |
| exo-siRNA | exogenous siRNA |
| FID | Free induction decay |
| FMRP | Fragile X mental retardation protein |
| HMQC | Heteronuclear multiple quantum correlation spectroscopy |
| HPLC | High performance liquid chromatography |
| HSQC | Heteronuclear single quantum correlation spetrocopy |
| HuR | Human antigen R |
| iMet | Initiator methionine |
| INEPT | Insensitive nuclei enhanced by polarization transfer |
| kDa | kilodalton |
| Khd1 | KH-domain protein 1 |
| KOW | Kyrpides-Ouzounis-Woese |
| LB | Lysogeny broth |
| Loqs-PD | Loquacious-PD |
| LSm | Like Sm |
| m | methyl |
| *M. musculus* | *Mus musculus* |
| mA | milliampere |
| miRISC | micro-RISC |
| miRNA | microRNA |

| | |
|---|---|
| Mpp6 | Membrane palmitoylated protein 6 |
| mRNA | messenger RNA |
| Nab | NGFI-A binding protein |
| NEXT | Nuclear exosome targeting |
| NFIB | Nuclear factor 1 B-type |
| nM | nanomolar |
| NMR | Nuclear magnetic resonance |
| NOESY | Nuclear Overhauser enhancement spectroscopy |
| Nop53 | Nucleolar protein 53 |
| Nrd1 | N-arginine dibasic convertase |
| p53 | Phosphoprotein 53 |
| PAGE | polyacrylamide gel electrophoresis |
| PAN | Proteasome-activating nucleotidase |
| PARN | Poly(A)-specific ribonuclease |
| Pasha | Partner of Drosha |
| PAZ | Piwi-Argonaute-Zwille |
| PFG | pulsed field gradient |
| PICT1 | Protein interacting with carboxyl terminus 1 |
| piRNA | Piwi-interacting RNA |
| PIWI | P-element induced wimpy testis |
| ppm | parts per million |
| PRE | paramagnetic relaxation enhancement |
| pre-miRNA | Precursor miRNA |
| pri-miRNA | Primary miRNA |
| PROMPT | promotor upstream transcripts |
| PTM | posttranslational modification |
| Pub1 | Poly(U)-binding protein 1 |
| Puf6p | Pumilio/fem-3 mRNA binding factor protein 6 |
| r | distance |
| Rat1p | Ribonucleic acid-trafficking protein 1 |
| RBD | RNA binding domain |
| RBP | RNA binding protein |
| RecA | Recombinase A |
| rf | Radiofrequency |
| RHAU | RNA helicase associated with AU-rich elements |
| RISC | RNA induced silencing complex |
| RLC | RISC loading complex |
| RNA | ribonucleic acid |
| RNAi | RNA interference |
| RNase | ribonuclease |
| RNP | ribonucleoprotein |
| ROESY | Rotating-frame nuclear overhauser effect spectroscopy |
| rpm | revolutions per minute |
| rRNA | ribosomal RNA |
| Rrp | Ribosomal RNA-processing protein |
| S | Svedberg |
| *S. cerevisiae* | *Saccharomyces cerevisiae* |
| SANS | Small angle neutron scattering |
| SAXS | small angle X-ray scattering |
| She2p | SWI5-dependent HO expression protein 2 |
| siRISC | short interfering RNA-RISC |
| siRNA | short interfering RNA |
| SLS | static light scattering |
| SPR | surface plasmon resonance |
| ssRNA | single-stranded RNA |
| TAF11 | TATA-box binding protein associated factor |

| | |
|---|---|
| TEV | tobacco etch virus |
| TOCSY | Total correlation spectroscopy |
| TRAMP | Trf4/5-Air1/2-Mtr4 polyadenylation complex |
| TRBP | HIV-1 TAR RNA binding protein |
| tRNA | transfer RNA |
| Utp18 | U3 small nucleolar RNA-associated protein 18 homolog |
| UTR | untranslated region |
| v/v | volume/volume |
| w/v | weight/volume |
| watergate | WATER suppression by GrAdient Tailored Excitation |
| Xrn1 | 5’-3’ exoribonuclease 1 |

# 1 Introduction

Every organism interacts permanently with its environment and is exposed to varying stimuli. Its' well being is depending on adaptation to this dynamic environment which requires the organism to be flexible and integrate both internal and external signals. This capability to react is based on properties of the smallest building block: the cell and its components. Each cell is composed of the main biomolecules DNA, RNA, proteins, lipids and sugars and a complex interaction network confers features to the cell and the organism. Proteins for example occur as enzymes, antibodies, co-factors, binding factors, molecular sensors, receptors and motors. They pursue diverse tasks, be it e.g. in metabolism, cell division or viral defense. Protein production, modification and degradation thus needs to be carefully regulated and a pivotal regulation tool is to control translation (Figure 1.1).

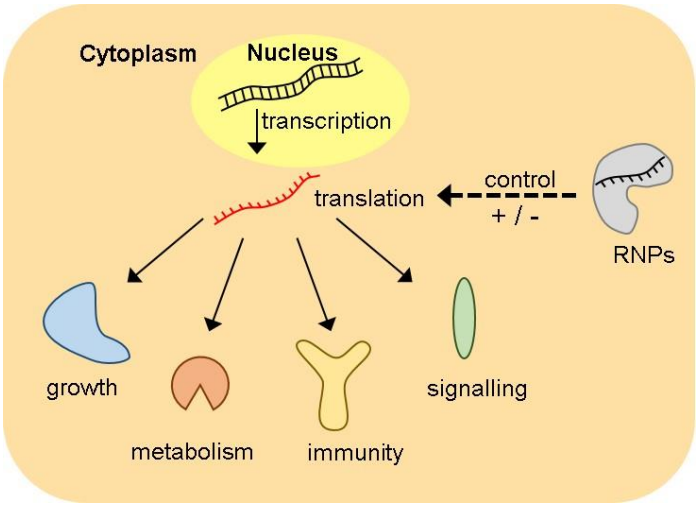



Figure 1.1: RNPs control translation, a key step in protein production. The dogma of molecular biology comprises transcription of DNA to RNA in the nucleus, processing of the mRNA and translation into proteins in the cytoplasm. All processes are strictly regulated and one possibility to do so is translational control by ribonucleoproteins. Different RNPs can enhance or limit translation and exhibit different functions apart from translation control.

Translation is the conversion of mRNA into functional proteins and different mechanisms are available and commonly used by cells to control it. Translational control can occur via RNA interference (section 1.1), RNA localization (section 1.2) and RNA degradation (section 1.3) and in all pathways complexes of proteins and RNAs are involved. Most complexes exhibit functions beyond translational control like viral defense (section 1.1) or

maturation of particles (section 1.3). In this thesis three different ribonucleoproteins (RNP) are characterized, their RNA binding is analyzed and potential implications for biology are deducted.

## 1.1 RNA interference

### 1.1.1 Formation of RISC and RNA interference

Three different classes of RNA are involved in RNA interference (RNAi): microRNAs (miRNAs)[1] were first described in *C. elegans* to exhibit regulatory functions via complementary binding[2] while short interfering RNAs (siRNA)[1] were analyzed five years afterwards[3]. Here, the great regulatory potential of dsRNA compared to single-stranded RNA was analyzed. The third group of RNA are Piwi-interacting RNAs (piRNA)[4,5] that have a distinct interference mechanism than the other two. For mi- and siRNA the fundamental mechanism is the same: they both interfere with RNAs and require a certain degree of sequence complementarity[1]. For these two RNA classes most of the processing pathway is very similar: Both mi- and siRNA derive from longer dsRNA precursors[1,6] that get processed to shorter fragments. During processing one strand is aborted and only a single strand is loaded onto the so called RNA-induced silencing complex (RISC)[6]. The RISC comprises an endonuclease[7] that exhibits the – mostly – inhibitory function of RNA interference. RNA interference thus is a specific type of RNA degradation and shares several features with other RNA degrading pathways (see section 1.3).

Despite important similarities all pathways differ significantly: miRNAs are usually transcribed by the RNA polymerase II[8] or III[9] (Figure 1.2 A). These products are called primary miRNAs (pri-miRNA)[6] and are very often co-transcribed with the coding regions nearby or derive from mono- or polycistronic clusters[10,11]. In fact they often are indeed introns[6]. These precursors are further processed by the so called microprocessor comprising Drosha[12], an RNase III enzyme[13], Pasha[12], a double-stranded RNA binding protein, or in humans DGCR8 (DiGeorge syndrome critical region 8)[14]. Besides their function in RNAi Drosha is involved in neurogenesis via suppression of the transcription factor NFIB[15] and DGCR8 can also act independently and recruit the exosome to degrade dsRNAs[16]. While DGCR8/Pasha[17] is crucial

for accurate positioning of the RNA within the complex[14,18], Drosha harbors the catalytic unit that cleaves the RNA[13]. The product forms stem loop RNAs (pre-miRNA) which are exported via exportin 5[19] from the nucleus to the cytoplasm where another endoribonuclease Dicer (Dcr) cleaves off the loop to leave a dsRNA[20] (Figure 1.2 B). One of the RNA strands (called miRNA) is loaded onto an Argonaute protein (Ago) that forms the functional and mature RISC[6,20] (Figure 1.2 D). This transient complex is called the RISC loading complex (RLC) and comprises a dsRBP, Dicer, Ago and the respective RNA[21,22] (Figure 1.2 C). Potentially certain factors like TAF11 support tetramerization of RLC components which might enhance loading activity[23]. The other strand (miRNA*) is usually degraded. However, several cases are known where the miRNA* remains intact and exhibits a specific function, e.g. miR-9*[24] and miR-21*[25]. Though most miRNAs are processed as described above, some have different origins, processing and function: Specific miRNAs derive from the spliceosome where an intron forms a miRNA called mirtron[26,27]. In these cases no processing with Drosha is required, instead the RNA can directly be cleaved by Dcr[26] and is loaded onto RISC. Similarly to siRNAs miRNAs may also derive from viruses[28].

MiRISC binds sequence-dependently to mRNAs and inhibits protein translation[6,29] (Figure 1.2 E). MiRNAs act post-transcriptionally on both translation intiation by blocking proper ribosome assembly[30] or other mechamisms[31] and translation elongation[32]. Therefore, usually only a partial sequence complementarity[1,33] within the seed region (2-8 nt) and the 3'-UTR of the RNA is necessary allowing miRISC to silence multiple RNAs at once[34] which often derive from one gene cluster. In case of perfect matching miRNA miRISC acts like siRISC and leads to mRNA decay[35]. MiRISC thus does not only inhibit specific functions, it can also enhance degradation of mRNAs.

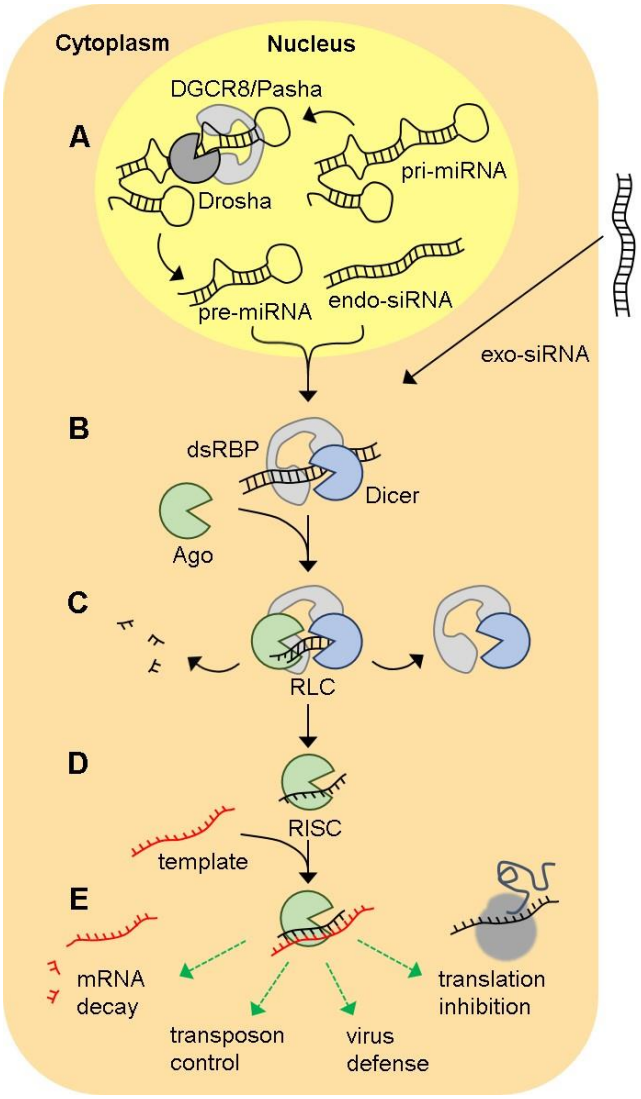



Figure 1.2: Schematic pathway of RNA interference. **A)** Pri-miRNA are processed by the microprocessor complex comprising Drosha and DGCR8 or Pasha. The pre-miRNA and endgoneous siRNA get exported from the nucleus via exportin-5. **B)** Pre-miRNA and both endo- and exogenous siRNA are bound by Dicer and a dsRBP and are processed into smaller double-stranded fragments (19-23 nt). **C)** Upon binding of an Argonaute protein the RISC loading complex (RLC) forms. Strand selection occurs based on thermodynamic stability of the RNA duplex termini. **D)** After degradation of the passenger strand the guide strand is loaded onto Ago to form the mature RISC. **E)** The RISC binds sequence-specifically to its substrates and exhibits various functions. Green arrows indicate a positive enhancement.

In contrast to miRNAs siRNAs can be both endogenous or exogenous[36,37] (Figure 1.2 A). They either derive from transposons, RNA hairpins, pseudogenes or viral RNA[6,38,39]. Like miRNAs they are processed form long dsRNA precursors but in contrast to miRNA precursors they are perfect matching[40]. In contrast miRNAs may contain not perfectly base-paired sequences, mismatches and loops. SiRNAs are directly processed in the cytoplasm by Dcr and

the mature siRNA is 19-21 nt long, comprises a 5' phosphate and a 2 nt 3' overhang[6,40], both necessary for Ago binding[7]. As for miRNA one strand of the double-stranded siRNA precursor gets degraded (passenger strand)[41] while the guide strand is loaded onto Ago to form the RISC (Figure 1.2 C, D). In general strand selection depends on the thermodynamic stability of the siRNA duplex[42]: The strand with the less stable 5' end is maintained and part of the functional RISC. However, in case of lacking clear thermodynamic asymmetry both strands might get incorporated into a functional RISC[42]. Like the miRISC the siRISC recognizes its targets via sequence complementarity. Upon binding of the respective substrate Ago cleaves the RNA between nucleotides 10 and 11 of the guide siRNA strand[43], the two products get released in a specific order[44] and are then further degraded by other pathways (see section 1.3).

Both mi- and siRNAs seem to play important roles in cancer development[11,24] as well as other diseases[25] and could serve as potential drugs for various diseases. Especially for diseases based on misregulated genes, e.g. cancer or metabolic diseases, and for viral defense siRNAs could be a potential tool[45]. In *Drosophila* for example siRNAs have been proven to be an effective anti-viral molecule[46].

The third class of RNAs are piRNAs which differ in many aspects from the previously mentioned RNA classes: Usually they derive from heterochromatin clusters[47] and prevent transcription of transposons[47,48] which could have a negative impact e.g. on germ lines and disrupt genes. PiRNAs do not require processing by Dcr and are slightly longer than mi- or siRNAs (26-35 nt)[49].

### 1.1.2 Protein key players in RNA interference

RNA processing during the above mentioned RNAi pathways involves various proteins: Drosha[13], Dicer and Ago[50] all possess (endo-)ribonuclease activity[51] and are essential for processing of miRNAs and siRNAs or in case of Ago cleave target RNAs. Dicer is an RNase III enzyme that cleaves unspecificially different substrates. While in humans only one isoform exists, *Drosophila* possesses two Dcr variants[52]. Usually Dcr is present in complex with a double-stranded RNA binding protein (see next section). It is a multidomain protein[53] comprising a PAZ domain required for RNA end binding[54], a DEAD box helicase which unwinds the RNA duplex and two RNase III domains that cleave one strand each[55]. Additionally, it

contains a double-stranded RNA binding domain (dsRBD) which interacts with the substrate but does not contribute to differentiation between different dsRNAs[56]. Though Dicer is a key player in RNAi it seems dispensable for selected sublasses of siRNAs[57].

Argonaute proteins divide into two subclasses: the Ago and PIWI proteins[7]. While Ago proteins occur ubiquitously PIWI proteins are restricted to germline cells[20]. They are specific for piRNAs while Ago proteins bind both si- and miRNA[7,20]. Ago comprises a PIWI domain that is similar to RNase H and harbors thus the catalytic site[58], and a PAZ domain[59] which binds specifically to 3' ends of RNA[54] and therefore ensures proper orientation of the RNA within Ago. The other RNA terminus is bound by a MID domain[60].

### 1.1.3 Double-stranded RNA binding proteins – features and functions

An important class of proteins involved in RNA interference are double-stranded RNA binding proteins which comprise usually two or three double-stranded RNA binding domains[61-64] and exhibit various functions[65].

DsRBDs show a conserved structure of an αββββα-fold[66,67] and bind preferentially to dsRNA[68]. The canonical binding interface comprises the two helices and a flexible loop connecting β-strand 1 and 2[62,63]. Binding occurs to two successive minor grooves and the major groove in between[67] of the RNA. Both structure and binding mode seem to be conserved over all dsRBDs though some show aberrant characteristics[69] that could modulate binding[63,70]. Interestingly, the affinity seems to differ across a large range and can be modified by phosphorylation[71]. However, all dsRBDs show a preference for dsRNA[72] in A-form helix[67] and can tolerate mismatches to a certain extent. Even though a certain specificity for sequence or structure has been described dsRBDs seem to bind in a sequence-independent but shape-specific mode to RNA[64,73]; a plausible explanation since most contacts occur to the phosphate backbone or sugars[68,72] moieties of the nucleic acid. For some proteins like ADAR2 sequence-specific binding to the minor grooves has been reported[74]. The general lack of sequence specificity potentially gives rise to considerable dynamics within the system[75]: Human TRBP has been shown to slide on dsRNA[76].

DsRBPs seem necessary for proper RNAi since they are not only involved in si-/miRNA processing but also do enhance effective loading of the mature si-/miRNA strand onto Ago[22,77] or directly increase the affinity of Dicer for its substrate[78]. Furthermore, they could be essential to sense the thermodynamic asymmetry of an RNA duplex, e.g. *Drosophila* R2D2 in complex with Dicer-2 occupies the stable 5' end[79] and human TRBP on its own binds the stable duplex end[80]. Loqs-PD has been shown to modulate Dicer cleavage and reactivity during precursor siRNA processing[81] and selected dsRBPs can even modulate and straighten duplex RNAs[82] or stabilize unstable RNA structures[83].

In most cases both dsRBDs contribute to RNA binding but the exact interplay between the two is elusive up to date. Cooperativity has been proposed for some systems while in general the dsRBDs seem to bind rather independently[84]. For efficient RISC formation both dsRBDs seem to be required though[61]. In many dsRBPs the dsRBDs are connected by a flexible linker of considerable length (up to 60 residues as for TRBP[84]) that could provide independence of the two domains. This linker could be essential to enable sliding of the domains on the respective RNA[75]. When a third dsRBD is present it usually mediates protein-protein contacts to the helicase domain of Dicer[78,85]. Since not all dsRBPs contain a third dsRBD it is speculated that for these proteins the C-terminus contributes to these interactions.

### 1.1.4 The isoform Loquacious-PD is specific for endogenous siRNA maturation in *Drosophila*

In *Drosophila* Dcr-1 together with the Loqs isoform PB[78,86] is involved in miRNA biogenesis[52,87,88]. Dcr-2 in contrast promotes siRNA maturation[52] and is involved in both exogenous and endogenous siRNA maturation (Figure 1.3). However, depending on the degree of RNA complementarity, exchange between both pathways seems possible[52]. Specificity for either siRNA pathway seems to be conferred by its dsRBP partner[36]: R2D2 bound Dcr-2 processes exogenous siRNA[77] (Figure 1.3 right panel) while the Loqs-PD bound form is specific for endogenous siRNA[89-91] (Figure 1.3 left panel). Recent studies though indicate that the two pathways might not be separated that strictly: A cross talk seems possible and both RBPs might compensate partially the absence of the other or act in a sequential mode on processing and loading[92] (Figure 1.3 B, C). In other studies it has been

shown on the other hand that both Loqs-PD and R2D2 are involved in dicing of the respective RNA substrate and do rather antagonize each other[36]. However, only Loqs-PD seems to enhance endo-siRNA processing[86] and it remains unclear whether R2D2 and Loqs-PD – which share the binding site on Dicer-2 – bind simultaneously to the RNase or mutually elusive[36,86].

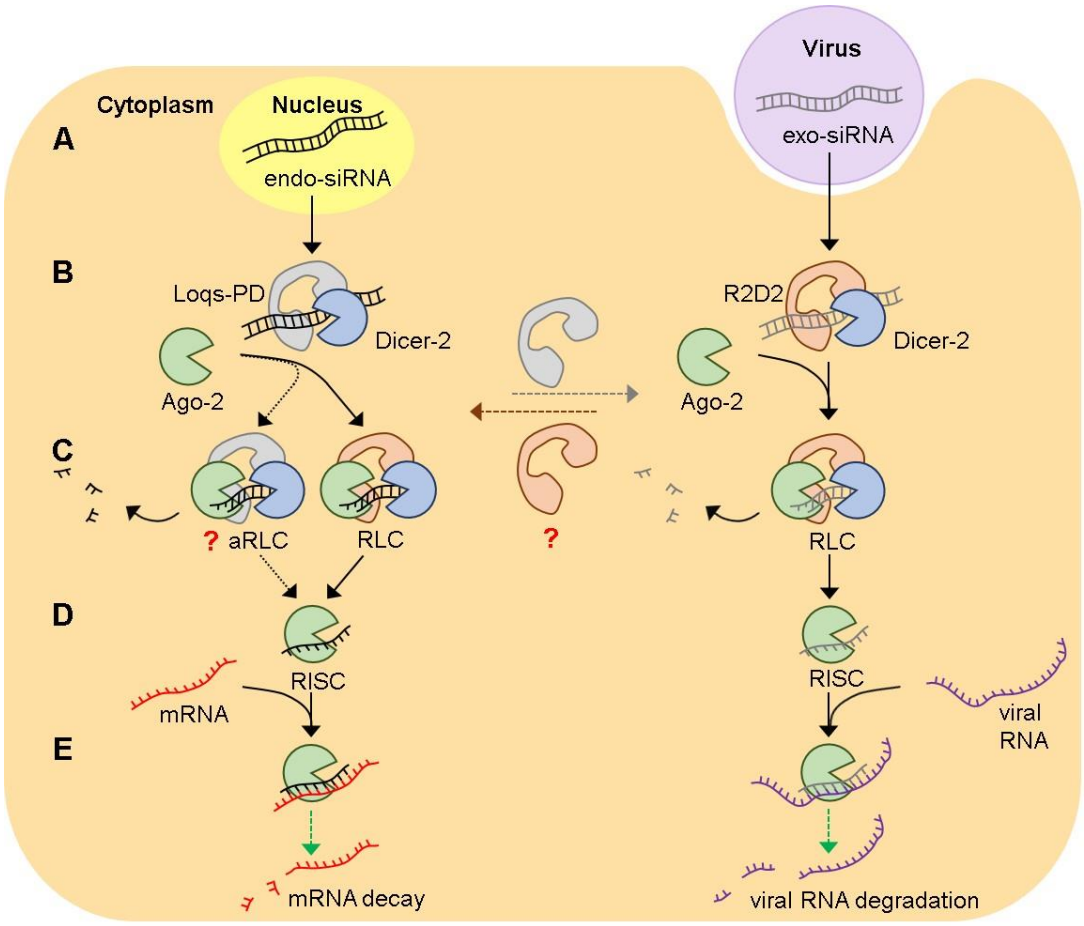



Figure 1.3: Pathways of siRNA maturation in *Drosophila melanogaster*. **A)** Endogenous siRNAs derive from transposons or pseudogenes. They are exported from the nucleus while exogenous siRNAs derive from viruses. **B)** Both siRNAs are bound by Dicer-2 and a dsRBP. Note that R2D2 binds exo-siRNAs while Loqs-PD is specific for endo-siRNAs. The siRNAs are processed into smaller fragments. **C)** Upon recruiting Ago-2 the RLC forms, strand selection takes place and the guide strand is loaded onto Ago. While it is known that the two dsRBPs are specific for binding and processing of the respective siRNAs their role for the RLC is not well defined. A cross-talk between both pathways seems possible and R2D2 could be part of the siRNA-RLC. An alternative RLC (aRLC) could involve Loqs-PD. **D)** After release of the other proteins the mature RISC forms. **E)** It binds sequence-specifically to its target and both RISCs promote RNA degradation. Endo-siRISC promotes mRNA decay to regulate protein synthesis while the exo-siRISC enhances viral RNA degradation. Green arrows indicate a positive enhancement.

R2D2 is furthermore involved in RISC loading and it remains elusive so far whether Loqs-PD could exhibit a similar function[36] and form an alternative RLC (aRLC) (Figure 1.3 C). While initially miRNA was thought to be loaded onto Ago-1 after processing whereas both siRNAs are bound by Ago-2 to form the mature RISC[93], this model has been revised. The different Ago proteins do not seem to be specific for either class of the RNA but rather exhibit different functions[94]. Furthermore, it seems plausible that different Ago proteins act sequentially within the same pathway[95]. Loading – but not strand selection or cleavage – of the mature siRNA onto effector RISC seems to involve a complex machinery of several chaperones[96,97]. The different isoforms of Loqs and R2D2 seem to have preferential RNA substrates and protein binding partners but can replace each other to a certain extent[88].

## 1.2 RNA localization

### 1.2.1 Mechanisms of RNA localization

RNA localization provides a tool for translational regulation by controlling spatial gene expression in different cell compartments which is used by many organisms. The asymmetric distribution of mRNAs[98] limits protein expression locally and thus allows for example to define the body axis[99,100]. Furthermore, mRNA localization is crucial in embryo development, cell polarization[101], axon growth and regeneration[102] and provides numerous benefits to the organism: Regulating availability of mRNA is very efficient as only few mRNA molecules need to be transported but can serve as templates for multiple proteins at the target site[103]. Clusters of mRNAs allow the efficient synthesis of large protein complexes and the direct assembly thereof and the protein production can be adjusted to specific needs and as a response to external stimuli[102]. Prominent examples of mRNAs that are carefully localized are *oskar*, *gurken*[104], *nanos* and *bicoid*[99,100] in *Drosophila* embryos and oocytes. These RNAs determine cell polarity and thus define the body axis[105] and its compartments.

Across most species distinct mechanisms to localize mRNAs have evolved: Some mRNAs can get locally enriched by binding to proteins that are targeted to specific sites or are degraded to prevent translation in predetermined regions[106] (see 1.3.1). In these cases RNA localization is closely linked to RNA degradation pathways, e.g via the CCR4-NOT complex[107]. Usually mRNAs are actively transported to their respective target sites[108] (Figure 1.4). Therefore *cis*-elements within the mRNA[109,110] and specific RNA-binding proteins[111] guide and control mRNA transport via motor proteins. Motor proteins, namely dyneins, kinesins and myosins[112], usually dimerize[112,113] and migrate along mictrotubules or actin filaments in an ATP-dependent manner[114] (Figure 1.4 B). These proteins either migrate in plus or minus direction and thus enable directional transportation of the RNA cargo[112]. A sequential transportation involving multiple different motor proteins is also possible and allows fine tuning of RNA transport.

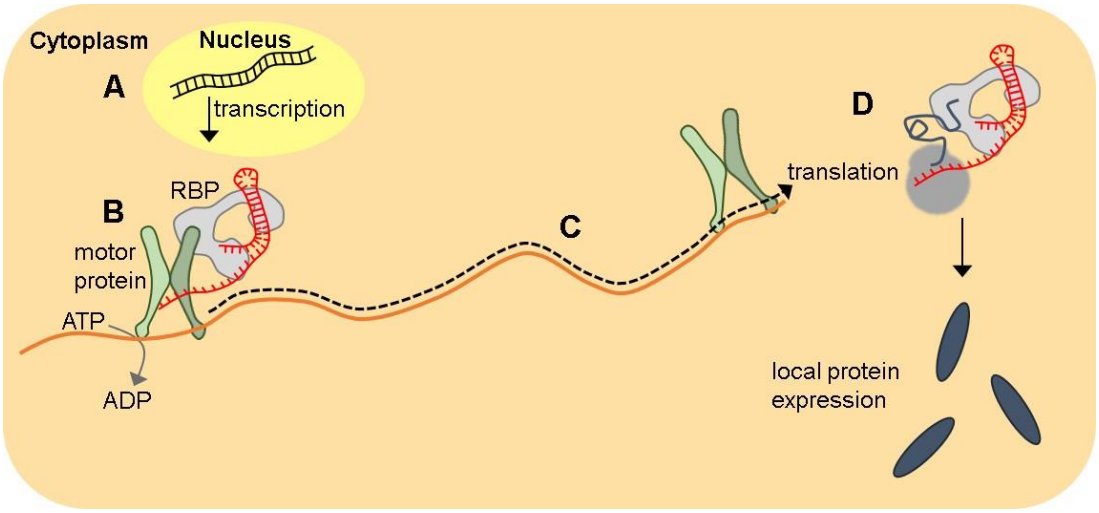



Figure 1.4: Principle of active mRNA transport and local protein expression. **A)** After transcription the mRNA is exported to the cytoplasm. Binding of RBPs can already occur at this stage. **B)** RBPs bind to specific localization signals located usually in the 3'-UTR of the mRNA and recruit motor proteins. **C)** In an ATP-dependent manner the motor proteins promote directional transport along microtubules or actin filaments. **D)** At the target site translation is initiated which leads to a locally restricted protein synthesis.

### 1.2.2 Disease related RNA binding proteins in RNA localization

RNA binding proteins like She2p[115], She3p[116], Puf6p[115], Vg1[117] or Khd1[118] involved in RNA localization comprise various conserved[119] as well as novel[120] RNA binding domains and properties[111]. They recognize specific localization sequences within the RNA to determine cargo fate[104,117]. Like FMRP[121] they often aggregate or cluster to form RNP-granules[122]. This aggregation is reversible but can be affected by mutations which usually cause neurodegenerative diseases[111] when occurring in dendrites or other neuronal cells. Prominent examples for mutation caused diseases are ALS or fragile X-syndrome[123] with the two affected proteins ataxin-2 and FMRP respectively.

### 1.2.3 Staufen – a dsRNA recognizing protein crucial for brain development

In mammals and *Drosphila* one RNA-binding protein involved in mRNA localization is Staufen[124]. Staufen occurs as two orthologues in mammals, Staufen-1 and 2, which exhibit different functions. While Staufen-1 is found ubiquitously[125] Staufen-2 expression is limited to specific tissues, e.g. the brain[126]. There Staufen-2 is crucial for neurogenesis, memory formation[127] and plays a role in neurodegeneration. Additonally, it ensures proper

asymmetric distribution of mRNAs in dendrites and is furthermore involved in long-term depression via metabotropic glutamate receptors while Staufen-1 exhibits an opposing function in promoting long-term potentiation[128]. Staufen-2 potentially regulates tissue growth and size[129]. Careful regulation of the system and the two proteins is thus crucial for proper brain function and development. Both Staufen-1 and 2 occur in different isoforms (2 and 4 respectively)[130] and comprise four to five dsRBDs[131,132] and a tubulin-binding domain (TBD) which binds to the motor proteins[130]. Similarly to Loqs or other dsRBPs dsRBD1, 3[133] and 4[132] of Staufen have been described to bind dsRNA. Staufen dsRBD2 mediates proper mRNA localization via microtubuli while dsRBD5 initiates its translation[132]. Analysis of Staufen-2 bound mRNAs in *Drosophila* and mammals revealed Staufen-recognized structures (SRS) within the RNA 3'-UTR that were required for Staufen binding[134]. Interestingly, RNA recognition and localization are thus mediated by duplex RNAs or otherwise structured RNA like Alu elements[135,136] instead of the usually recognized single-stranded RNA (see 1.2.1). Binding for *Drosophila* Staufen dsRBD3 occurs via the canonical binding interface but helix1 could provide specificity for stem loop structures[133]. Since sequence or shape specificity of dsRBPs seems to be depending on the system (see 1.1.3) a model for RNA recognition of Staufen-2 could provide new insights into dsRNA binding.

## 1.3 RNA degradation

### 1.3.1 RNA degradation pathways and the importance of the exosome for RNA homeostasis

An RNA life cycle is completed by its degradation. Three pathways are known among pro- and eukaryotes, all of them involve endo- or exoribonucleases as enzymes[137] and require decapping and/or deadenylation of the RNA[138,139]. A very efficient way to degrade RNA is endoribonuclease mediated[140]: usually two cleavage products are generated which can be degraded simultaneously by components of the other two pathways (Figure 1.5 A). Important examples for this kind of RNA processing are Dicer, Drosha and Ago[137,141-143], which underlines that RNAi pathways in fact are RNA degradation pathways.

The deadenylation-dependent pathway is the most prominent in eukaryotes. However, most eukaryotes possess a deadenylation-independent path as well (Figure 1.5 B)[139]. Here degradation is initiated by the removal of the m7G cap from the 5' end by Dcp1 and 2[144] which is induced by binding of the Sm complex (LSm) to the 3' end of the RNA[145]. Afterwards the exoribonuclease Xrn1 degrades the RNA from the 5' to 3' end[139]. In one of the deadenylation-dependent pathways (Figure 1.5 C, $C_i$) this mechanism is the same, only preceded by removal of the 3' poly(A)tail by different enzymes (e.g. CCR4-NOT, PAN, PARN)[146]. A second deadenylation-dependent pathway involves the exosome (Figure 1.5 $C_{ii}$)[139,146,147]: After deadenylation the RNA is cleaved and the remaining methyl-guanosine gets degraded by the scavenger decapping machinery (involving DcpS)[146,148]. In contrast to Xrn1 the exosome contains Rrp6, a 3' to 5' distributive exoribonuclease[149], but also comprises Rrp44 which functions as both an endoribonuclease[150] and a 3' to 5' processive exoribonuclease[151,152]. For specific RNA substrates the Sm complex might be involved here as well[153].

The exosome functions in various pathways that require controlled RNA turnover: It is not only essential in degradation of superfluous RNA and thus RNA homeostasis[154,155], but also in RNA processing, maturation and quality control[147,156]. By controlling 3' end processing[157] and RNA packaging[158] it provides an additional layer to gene expression regulation. Often the exosome is functionally coupled to splicing[159-161] and is thus efficiently involved in translation regulation. Furthermore, the exosome is crucial to eliminate aberrant transcripts[155] and exhibits functions during DNA damage repair, genome maintenance[162] and

double-strand brakes (DSB): The absence of certain exosome factors (e.g. Rrp6) leads to insufficient DSB repair[163] and the exosome is also crucial to prevent DSB by solving R-loops[164,165], a major cause for DSB. Besides that it contributes to the antibody class switch recombination in B cells[164].

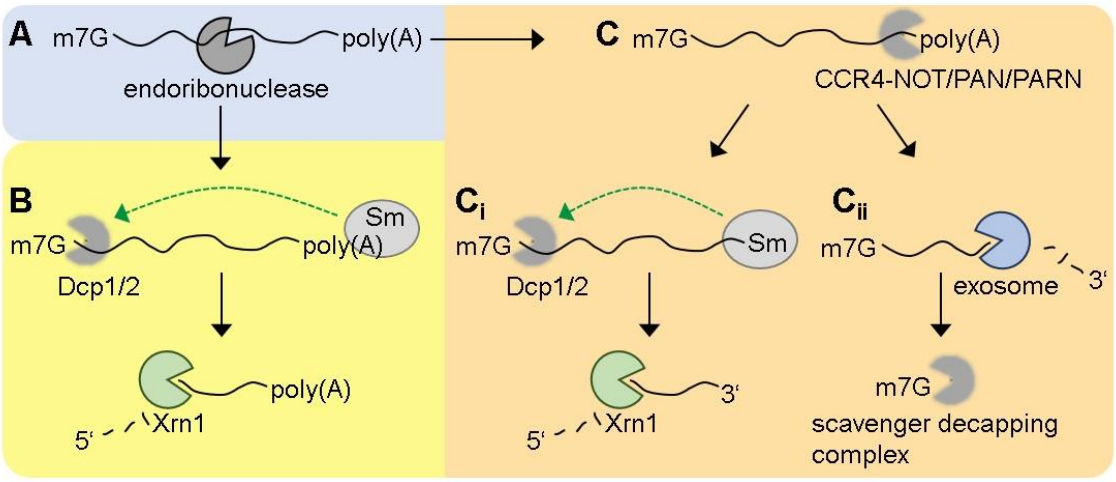



Figure 1.5: Different eukaryotic RNA degradation pathways. **A)** Endoribonuclease mediated degradation leaves two cleavage products that are degraded via the **B)** deadenylation-independent pathway and the **C)** deadenylation-dependent pathway. Note that **$C_i$)** and B are identical apart from the prior removal of the poly(A)tail in C. In contrast to Xrn1 in B and $C_i$, degradation via the **$C_{ii}$)** exosome occurs in 3' to 5' direction. Green arrows indicate a positive enhancement.

Apart from mRNA turnover and its role for genome stability the exosome regulates different non-coding RNAs. It is involved in trimming of precursor RNAs and e.g. in yeast in the removal of RNAs from cryptic unstable transcripts (CUTs, cryptic transcription)[166], which arise from bidirectional promoter transcriptions, promoter upstream transcripts (PROMPTs)[167,168] in humans and enhancer RNAs (eRNAs)[165,169]. It regulates no-go decay[139], non-stop decay[170] and non-sense mediated decay[171,172] and recognizes unstable and incorrectly processed and folded RNAs, e.g. CUTs derived from ribosomal DNA repeats and telomeres[173]. Note that several mRNAs get already degraded within the nucleus by Rat1p[174]. By regulating RNA metabolism in general the exosome contributes to homeostasis and can prevent autoimmunity, which can be boosted by excess of RNAs[175,176]. Degradation via the exosome implies unfolding of a possibly structured RNA and subsequent degradation to individual nucleotides, which is both done by either the exosome itself or its cofactors.

### 1.3.2 The exosome is highly regulated by *trans*-acting factors

The so called “degradosome” of prokaryotes is well understood in terms of structure and mechanism[177] and comprises both RNase and helicase activity[178]. In contrast to that the eukaryotic exosome is more complex and comprises nine different proteins, of which six form heterodimers which get stabilized by a cap[147]. Since this core has a low catalytic activity[151,179] it serves as a platform for the endoribonuclease Rrp44 and Rrp6 which perform the cleavage reaction. Additionally, this core channel modulates the catalytic activities of the nucleases[152]. Depending on the substrate several accession modes are available to guide the RNA to these catalytic subunits[180,181] which might involve unwinding directly by Rrp44[182]. It is known that the platform binds several regulators; among them proteins that unfold and unwind structured RNAs[177] which enables the eukaryotic exosome to degrade double-stranded and structured RNAs as well.

The exosome is a key player in RNA metabolism and therefore needs to be carefully regulated[156]. As for many regulatory proteins or enzymes regulation can occur via post-translational modifications (PTM), e.g. phosphorylations[183]. Other regulatory factors act in *cis* or *trans*, either within the RNAs itself or as protein-cofactors. RNAs can comprise both positive and negative regulatory elements, which are usually secondary structure elements most often found in the 3’-UTRs[184]. However, certain elements do occur in 5’-UTRs or the coding regions as well[185]. One important regulatory RNA element are AU-rich elements[143,186] (ARE; adenylate-uridylate rich element, 3’-UTR) which is for example recognized by HuR that stabilizes the RNA structure[187] and prevents exosome access. Other RNA stabilizing proteins are Nab2, Pub1[188] and LSm/Sm[189,190], which protect the RNA from exosome binding and unfolding and are thus negative regulators. Sm however is also needed for proper functioning of the exosome, e.g. it controls processing of precursor-tRNAs[191]. Destabilizing proteins in contrast are often recruiters of the exosome and mediate contacts between the exosome and its target RNA. These positive regulators are also referred to as primary exosome-specificity factors (ESFs)[158] and are necessary to confer specificity to the exosome, which is required since the exosome as a helicase[177] is unspecific. Several of these ESFs-like negative regulators do as well[188] bind to AREs and recruit other factors like RHAU (DHX36, a DEAH-box helicase)[192] and DDX5 (DEAD-box helicase)[193] in humans or in *S. cerevisiae* the U3 protein 18 (Utp18), Nop53, Nrd1, Nab3 and Sen1[158].

Two important positive exosome regulators are no single proteins, but complexes instead, and different forms can be found among all eukaryotes from yeast to humans: The SKI and the TRAMP (Trf4/Air2/Mtr4 polyadenylation complex) complexes promote exosome activity[168]. The SKI complex is specific for the cytoplasm[194] and mainly induces degradation of RNA derived from unusual termination during translation. It contains a DExD/H RNA helicase[194,195] and unwinds the (structured) RNA substrates. In contrast, the TRAMP complex occurs only within the nucleus[156] and regulates turnover of nuclear non-coding RNAs as well as aberrant mRNAs. Up to date it is known that TRAMP distinguishes aberrant RNAs from correct ones based on their modifications, e.g. TRAMP preferentially processes hypomodified iMet-tRNAs [196]. In humans the NEXT complex (Nuclear Exosome Targeting complex) pursues similar tasks like TRAMP[168].

### 1.3.3 The positive exosome regulator Mtr4 unwinds RNA substrates

The TRAMP complex comprises one RNA binding protein (Air1 or 2)[197], one poly(A)polymerase (Trf4 or 5)[198] and the DExH helicase Mtr4[199,200] (Figure 1.6). After substrate binding the poly(A)polymerase component attaches a short poly(A) stretch to the 3' end of the RNA[201], thereby increasing the affinity of Mtr4 for the respective RNA[202,203] which subsequently unwinds the RNA in an ATP-dependent manner and directs it to the exosome for degradation. Interestingly, this system contains a negative feedback mechanism since Mtr4 itself restricts the addition of adenosines to the RNA to a maximum of four moieties[201,204,205]. Currently there is good reason to believe that polyadenylation occurs after unfolding through Mtr4[200], though an experimental proof is missing and it contradicts the theory of increased affinity of Mtr4 for polyadenylated RNAs. Besides it was shown that polyadenylation occurs in the absence of Mtr4[198]. While the TRAMP complex pursues further tasks like tRNA editing[206] Mtr4 on the other hand may also act independently of TRAMP in association with other factors[207] like C1D which is involved in DNA damage response and the human Mpp6[208].

The Mtr4 protein consists of two conserved RecA-like domains, a winged helix and ratchet domain which form the core of the protein[199]. While it is known that the RecA-like domains bind ATP/ADP, initiate the formation of the TRAMP complex and harbor the catalytic

site, little is known about the functions of the winged helix and ratchet domains. Most likely they are involved in RNA binding[203] or mediate contacts to other protein factors. A unique feature of the Mtr4 protein is the arch domain; an insertion comprising a stalk made of four helices and a KOW (Kyrpides-Ouzounis-Woese)[209] domain at its end. Similar arch domains exist in rRNA binding proteins like L24 which binds 23S rRNA[199]. Also in Mtr4 the KOW domain has been shown to bind RNAs[210] (Figure 1.6) and removal of this insertion leads to an overall lowered activity of Mtr4 in RNA unwinding[203]. Furthermore, the KOW domain is supposed to activate the exosome[197] and boost RNA degradation.

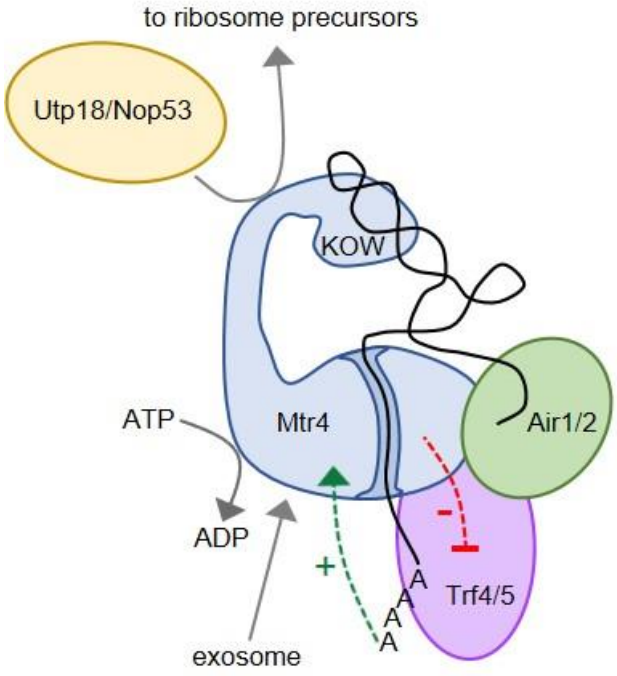



Figure 1.6: Scheme of the TRAMP complex consisting of the poly(A)polymerase Trf4/5 (purple), the RNA binding protein Air1/2 (green) and the helicase Mtr4 (blue). Trf4/5 adenylates the target tRNA which increases the affinity of Mtr4 (green arrow). Mtr4 limits adenylation (red line) and unfolds the RNA in an ATP-dependent manner. Utp18/Nop53 can recruit the exosome to ribosomal precursors via Mtr4 as docking platform.

Mtr4 does not only bind tRNAs and components of the TRAMP complex, it also interacts with other protein cofactors like Utp18 and Nop53 (Figure 1.6)[211]. Both proteins contain an "arch interaction motif" (AIM) that binds to the Mtr4 arch domain. Utp18 is part of the 90 S pre-ribosome while Nop53 binds pre-60 S particles[211]. Both proteins are involved in trimming of ribosomal precursor RNAs. Interestingly, Nop53 is the yeast homologue of PICT1, a human regulator of p53[211,212]. The recruitment of Mtr4 via AIM containing proteins

to e.g. the ribosome thus shows the importance of exosome mediated RNA homeostasis. Furthermore, Mtr4 interacts via its N-terminus with Rrp6 and Rrp47 and thus directly contacts the exosome[213].

## 1.4 NMR spectroscopy

### 1.4.1 A tool for structural biology

Besides X-ray crystallography and electron microscopy nuclear magnetic resonance spectroscopy (NMR) has evolved as a major tool in structural biology. While a majority of available protein structures have been determined by X-ray crystallography NMR provides various benefits over the other two methods: Since the sample is usually in the solution state only NMR allows analyzing dynamics and motions of a system while X-ray crystallography and electron microscopy require either crystals or immobilized samples. In addition, different setups permit the analysis of specific residues or even atoms and respectively quick and easy accessible experiments can provide information on ligand binding sites, molecular weights and the structural integrity of a biomolecule. Furthermore, for NMR is a non-invasive technique samples can be reused afterwards. However, NMR does come with a few drawbacks: The sample amount is usually quite large since it is an insensitive method and for a complete structure determination multiple time-consuming experiments need to be recorded. Analysis of the spectra can be laborious and, though decent progress was made over the last years, signal overlap limits the maximum size of biomolecules suited for analysis. In these cases an integrated structural biology approach is most promising combining benefits of different techniques.

### 1.4.2 Theory of NMR spectroscopy

NMR makes use of the intrinsic properties of the nuclei within the sample: In a simplified model every nucleus rotates around its own axis and thereby creates a magnetic moment called μ. This spin is divided into discrete states and characterized by the spin quantum number I. I can vary in steps of ½ from zero and determines the angular momentum of a nucleus (Table 1.1). This momentum scales with the gyromagnetic ratio, a nucleus specific constant describing the susceptibility for a magnetic field, and yields the magnetic moment (Equation 1).

$$\mu = \gamma \mathrm{I} \qquad (1)$$

Only nuclei with a spin quantum different than zero are thus visible in NMR. In practice, however, only nuclei with a spin quantum of a fraction of an integer are commonly used in biomolecular NMR spectroscopy. Full integer spin quantum numbers cause multiple effects and complicate analysis. Table 1.1 gives an overview of nuclei mostly used in biomolecular NMR spectroscopy. Deuterium is listed because it is often used to render parts of the samples 'invisible' and to facilitate analysis but it is usually not detected. Note that protons provide the highest gyromagnetic ratio and are therefor most sensitive in NMR. Apart from protons and phosphor, which is an additional probe available in nucleic acids, the natural abundance of the other nuclei is very low. For that reason, when detecting either of these nuclei, it is required to isotopically label the sample, meaning to enrich the isotope.

Table 1.1: Nuclei commonly used in biomolecular NMR spectroscopy and their properties. For each isotope the spin quantum number, the gyromagnetic ratio and its natural abundance is shown. Adapted from Cavanagh et al, 2006[214].

| Nucleus | Spin quantumnumber | Gyromagnetic ratio [(T*sec)$^{-1}$] | Natural abundance [%] |
|---|---|---|---|
| $^{1}$H | ½ | 2.6752*10$^{8}$ | 99.98 |
| $^{2}$H | 1 | 4.107*10$^{7}$ | 0.02 |
| $^{15}$N | ½ | -2.712*10$^{7}$ | 0.36 |
| $^{13}$C | ½ | 6.728*10$^{7}$ | 1.11 |
| $^{31}$P | ½ | 1.0841*10$^{8}$ | 100.00 |

The magnetic moment of a nucleus is able to interact with an external magnetic field. In a strong magnetic field $B_0$ the spins align with the applied magnetic field and occupy one of two states, either + ½ or – ½, parallel or antiaparallel to $B_0$, called α and β (Figure 1.7 A). The two states differ in energy levels and the energy difference between them is described as the following:

$$\Delta E = \gamma \hbar B_0 \qquad (2),$$

where ħ is the reduced Planck constant. The α-state is energetically favoured and the distribution of the spins between the α- and β-state is determined by the energy difference in equation 2: The larger the difference the more spins are found in the low energetic α-state (Figure 1.7 C). This difference in occupation gives rise to a net (or bulk) magnetization which reflects the sum of all spins (Figure 1.7 B). Note that all spins precess independelty around the magnetic field (Figure 1.7 A). Therefor no xy magnetization is observable. In a simplified model the net magnetization is used to create a detectable signal. Thus, high energy difference between the two states is desirable. Since the population difference is usually very low and in the range of parts per million (ppm) different variables are modified in NMR spectroscopy to increase the difference and obtain a maximum signal (Figure 1.7 C): First, nuclei with a high gyromagnetic ratio are used for detection, mostly $^{1}H$. Second, a stronger magnetic field $B_0$ can be used and third a decrease in temperature will lead to a higher occupation difference since the population of the states follows the Boltzmann equation:

$$\frac{N_\alpha}{N_\beta} = e^{\frac{2\mu B_0}{k_B T}} \qquad (3)$$

In equation 3 $N_\alpha$ and $N_\beta$ reflect the number of spins found in the respective state, $k_B$ is the Boltzmann constant and T the temperature.

In an NMR experiment a radiofrequency pulse rf is applied which comprises a magnetic component that interacts, similarly to $B_0$, with the spins in the sample (Figure 1.7 D). These spins, usually in z-orientation when in the α- or β-state, are rotated by the perpendicular rf pulse into the xy-plane and are all in phase, meaning coherent. The rf pulse applied needs to meet (or be at resonance with) the Larmor frequency of the nuclei which is the precession frequency of the nuclei around the z-axis.

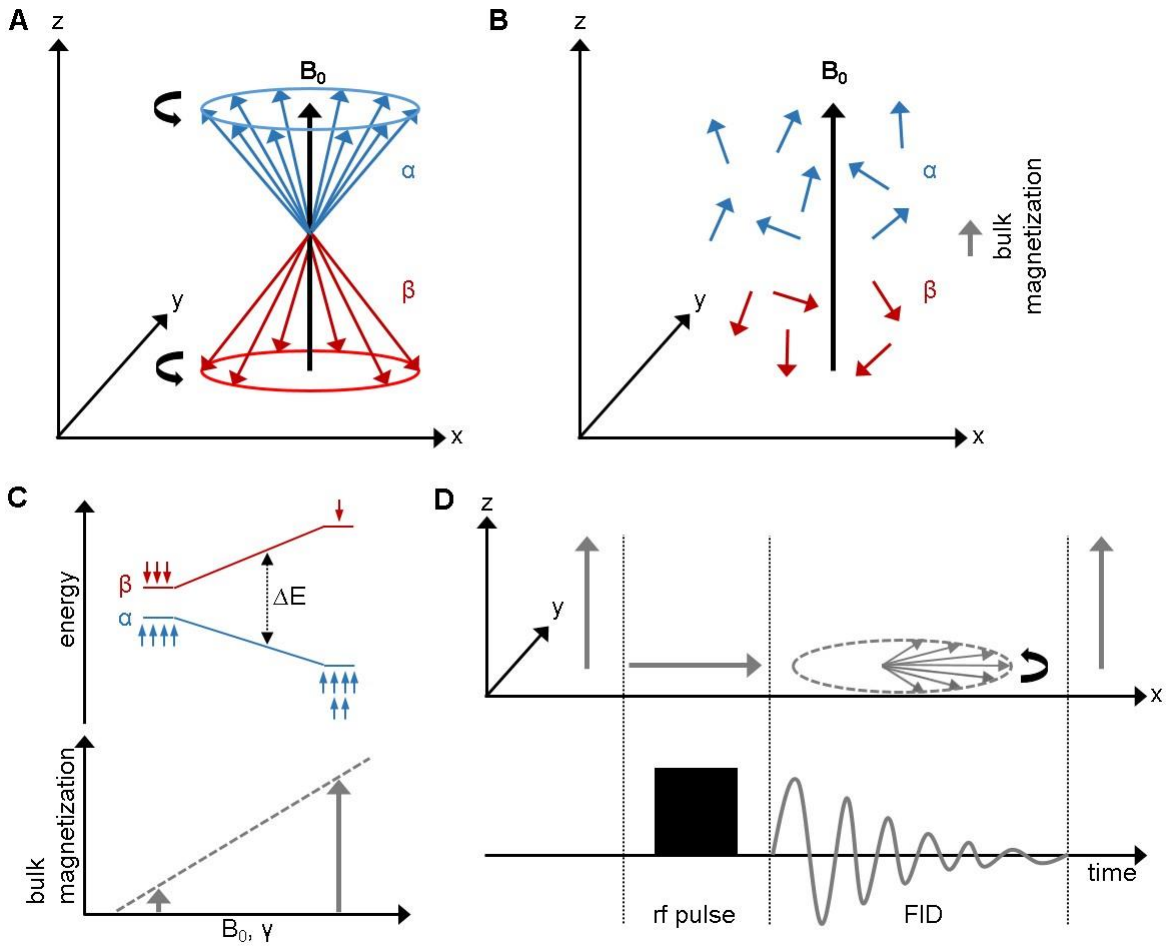



Figure 1.7: Schematic principle of NMR spectroscopy. **A)** Spins align either parallel (α, shown in blue) or antiparallel (β, shown in red) to the external magnetic field $B_0$ and precess around the z-axis. **B)** Note that there is no pure α- or β-state, the spins are tilted. The excess of spins with an α-state contribution give rise to a small net bulk magnetization along z (grey arrow). **C)** The two spin states have different energies and the difference increases with the magnetic field, the gyromagnetic ratio or at lower temperatures. The larger the difference the higher populated is the energetically favoured α-state. This population difference increases the net bulk magnetization (grey arrow). **D)** The bulk magnetization along z is rotated into the xy-plane by an rf pulse. After the pulse the spins precess in the xy-plane and loose their phase coherence. This precession is measured as the oscillating free induction decay (FID). After the experiment the magnetization returns to equilibrium along z.

In this transverse plane the spins can be further manipulated by other pulses and gather different information. The spins rotate in the xy plane which gives rise to a time signal. As soon as no rf pulse is applied the system relaxes. This relaxation is caused by different rotation velocities of the spins which causes a loss of phase coherence. Therefor the apparent and detectable net magnetization in the xy plane decreases, the so-called free induction decay (FID) (Figure 1.7 D). At the same time the spins rotate back from the transverse plane

into z direction until they reach equilibrium. This equilibrium corresponds to the bulk magnetization before the rf pulse was applied.

The free induction decay is an oscillating signal with a cosine modulation and an exponential decay. It contains information of all spins and needs to be deconvoluted. This deconvolution is achieved by a Fourier transformation, a mathematical function that converts a time-dependent signal into a frequency-dependent one. Thus, a single FID can be converted into a complete NMR spectrum with signals at frequencies corresponding to the precession frequency of all spins. Ideally, isolated nuclei of the same type will all have the same Larmor frequency and thus result in one single signal. In reality, however, the precession frequencies differ depending on the chemical environment and possible interactions with neighbouring nuclei. In a biomolecule all nuclei experience shielding and deshielding effects based on surrounding electrons, inductive effects and ring currents. These influences vary the frequencies and shift signals within the NMR spectra. The difference between the observed precession frequency and the Larmor frequency is called chemical shift. Since the chemical shift most often is very small and the Larmor frequency is depending on the magnetic field $B_0$ NMR spectra are converted to ppm and referenced to standard chemicals, e.g. TMS, according to the following formula:

$$\delta_{ppm} = 10^6 * \frac{v - v_{TMS}}{v_{TMS}} \qquad (4)$$

In equation 4 δ is the chemical shift, v corresponds to the resonance frequency of the sample spin and $v_{TMS}$ to the frequency of TMS. The chemical shift is very sensitive to its environment and thus provides an excellent tool to monitor changes in the sample.

### 1.4.3 Relaxation processes

As mentioned in 1.4.2 the FID is a decaying oscillating signal. The decay is caused by the bulk magnetization returning to equilibrium, meaning aligning to the external magnetic field along z. This process is called longitudinal (or T1) relaxation and is based on spin-lattice interactions. It describes the loss of xy magnetization and build up of the original z magnetization (Figure 1.8 B). The loss of coherence within the xy plane is called transversal (or T2) relaxation and is caused by spin-spin interactions (Figure 1.8 A). T2 relaxation

determines the length of the FID that contains information and describes dephasing of the spins that were, after the 90° pulse, aligned within the plane. Both relaxation mechanisms rely on dipole-dipole interactions and anisotropy. The spins experience local magnetic fields from surrounding atoms that vary and contribute to relaxation. These local fields are fluctuating and influenced by the overall tumbling of the molecule in solution. The tumbling of a molecule is described by its correlation time, the time it takes for the molecule to move one radian. It is depending on the molecular weight of the molecule: Larger molecules tumble slower. Tumbling is a random process and can be described by spectral densities. The spectral density of a molecule is depending on its correlation time and determines the contribution of different motion frequencies to the overall tumbling. For small molecules many frequencies contribute to motions while for large molecules mainly low frequencies do. Low frequencies increase the efficiency of T2 relaxation and decrease that of T1 relaxation. Therefor large molecules show fast T2 relaxation and slow T1 relaxation. Both relaxation processes can be described by the following equations:

$$M_z(t) = [M_z(0) - M_z^0]e^{-\frac{1}{T_1}t} + M_z^0 \qquad (5)$$

$$M_x(t) = M_x(0)e^{-\frac{1}{T_2}t} \qquad (6)$$

$M_z$ and $M_x$ are the longitudinal and transversal magnetization respectively. T1 and T2 are times that describe the two exponential processes. Usually in biological systems T1 is larger than T2 and mainly determines the length of the interscan delay required to enable full z magnetization prior to the next experiment (Figure 1.8).

Note that for proteins T1 (after reaching a minimum at low molecular weights) increases with higher molecular weight while T2 decreases. The molecular weight dependency of both processes limits NMR spectroscopy: Fast T2 relaxation causes broad signals while slow T1 relaxation requires long interscan delays. However, the two relaxation mechanisms can be intentionally utilizied for NMR experiments: T1 relaxation is used in the NOESY experiment while T2 relaxation is important for the ROESY (see 1.4.6).

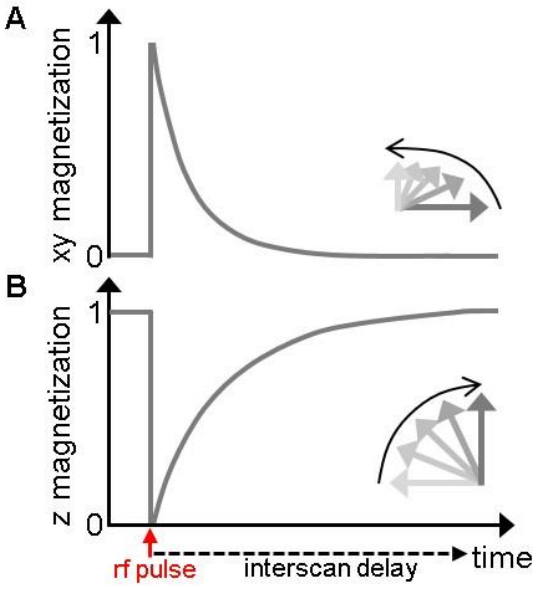



Figure 1.8: T1 and T2 relaxation are exponential processes. **A)** The rf pulse (red) transfers magnetization into the xy plane leading to maximum transveral magnetization. The spins loose phase coherence (grey arrows) which leads to an exponential decay of detectable transverse magnetization (T2 relaxation). **B)** After the initial rf pulse z magnetization is completely converted into transverse magnetization. Therefor longitudinal magnetization is at zero. Over time the spins return to equilibrium (grey arrows) and build up the initial z-magnetization (T1 relaxation). T1 relaxation mainly determines the minimal interscan delay required for full z-magnetization prior to the following experiment.

### 1.4.4 Benefits and mechanisms of magnetization transfer

Nuclei in a biomolecule interact with each other and 'sense' surrounding nuclei as described in 1.4.2. In general two possibilities for interaction can be distinguished: Spins either interact via bonds or through space. Spin-spin interactions via bonds are called J-couplings (scalar) and describe the influence of the nuclear spin of one nucleus on another transferred via electrons in bonds. Usually couplings are only visible across two or three bonds and are called geminale or viscinale couplings respectively. Coupling nuclei give rise to multiplets in the NMR spectra. The size J of a coupling is given in Hertz and correlates with the interaction strength. These couplings can be used to transfer magnetization between nuclei and the transfer efficiency depends on the coupling constant J.

As described in 1.4.2 protons with the high gyromagnetic ratio yield maximum signals in the NMR spectra. Due to high complexity in biomolecular NMR spectroscopy multidimensional NMR experiments need to be recorded in order to obtain information.

Therefor nitrogen and carbon as additional probes are required. It is possible to detect either of the two nuclei but the low γ will lead to a poor signal intensity. Thus, advanced pulse schemes have been developed that allow to transfer magnetization between different nuclei, e.g. transfer magnetization from a proton to nitrogen and back. This is done in a $^{1}H,^{15}N$-HSQC experiment. In this case the spectrum will contain information about both the proton and nitrogen and will show amide correlations and be two-dimensional. This reduces signal overlap and thus facilitates spectra analyses. The *insensitive nuclei enhanced by polarization transfer* (INEPT) is a building block that is used very often in different experiments to transfer magnetization between two nuclei (Figure 1.9). For a proper understanding product operators are required but are beyond the scope of this chapter. Therefore, the INEPT will be described qualitatively.

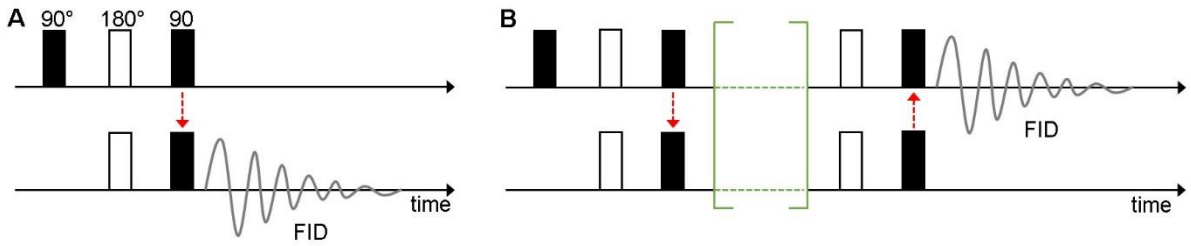



Figure 1.9: Schematic pulse sequence of INEPT magnetization transfer. **A)** The basic INEPT building block comprises a 90° pulse on both nuclei that transfers magnetization after coupling could evolve between the first and second 90° pulse. The FID can be recorded on the second nucleus. The red arrow indicates magnetization transfer. **B)** Two INEPT transfers allow forth and back transfer of magnetization. This is the basic concept of a 2D-NMR experiment and benefits from the high gyromagnetic ratio during detection of $^{1}H$. The green brackets indicate the possibility to modify the magnetization by including different pulse schemes.

After rotating the bulk magnetization of nucleus A (with a high γ) into the xy plane by the first 90° rf pulse, a 180° pulse, a so called spin echo, is applied to both nuclei A and B (figure 1.9 A). During that spin echo the two nuclei experience coupling to each other while the individual offsets are refocused and the following 90° pulses on both nuclei transfer the magnetization from nucleus A to B. This magnetization can now be observed in the free induction decay and the resulting signal will have an increased intensity compared to an experiment where only nucleus B was used. To further optimize the experiment a second INEPT step is very often included at the end of the pulse sequence to transfer magnetization back to protons for observation (Figure 1.9 B). Often a decoupling sequence during acquisition

removes the coupling between the hetero-atom and protons. This causes multiplets to turn into singulets which simplifies analysis and increases the signal-to-noise ratio.

The second option for spins to interact is through space. Here two spins interact via dipolar relaxation which means that both nuclei affect each others' z magnetization. This so called nucleaur Overhauser effect (NOE) can be detected for distances r between spins of up to 5-6Å and is used in NOESY spectra. The cross relaxation efficiency is distance-dependent and is described as

$$\mathrm{NOE} \sim \frac{1}{r^6} \qquad (7)$$

Furthermore, the NOE is time-dependent as it correlates with the cross-relaxation rate constant σ of the two nuclei and the cross-relaxation time τ, the so called mixing time. In NOESY experiments magnetization is therefor transferred automatically from a nucleus A to nucleus B after the magnetization of A has been modulated. Note that longer mixing times can lead to spin diffusion which means that magnetization is transferred from A to B to a third nucleus C or even further. This can complicate spectra analysis but allows to detect complex systems, e.g. secondary structures in RNAs.

### 1.4.5 Water suppression techniques

In aequeous solutions water gives rise to an intense signal in the NMR spectra at approx. 4.7 ppm. It saturates the receiver and thus decreases the intensity of sample signals. Therefor water suppression schemes need to be applied and are available. One possibility is to apply a presaturation on water by a low power continuous wave pulse before the first 90° pulse that leads to equal occupation of α- and β-state of the solvent and thus no solvent net magnetization can be transferred into the xy plane. Unfortunately, all sample signals with similar resonance frequencies as water, e.g. protein $H_\alpha$, will be saturated as well und thus remain invisible. Besides that presaturation can cause baseline artefacts and the saturated solvent might exchange with exchangeable protons from the sample, like amino and imino protons, and reduce their signal intensity. An alternative is the watergate sequence (water suppression by gradient tailored excitation) where the water signal is dephased in the xy plane. A pulsed field gradient (PFG) dephases all resonances after the first 90° pulse and a

following 180° pulse inverts all resonances except for water. Another PFG is applied and all inverted resonances get refocused while only water does not. The drawbacks are similar to the presaturation. Besides these two schemes further puls sequences have been developed to minimize the water signal, like a water flip-back and the jump-and-return pulse schemes. An alternative is the use of fully deuterated solvent which can ideally be measured without any water suppression sequence. This is especially useful when resonances near the water signal should be observed. However, exchangeable protons will not be visible.

### 1.4.6 2D homonuclear NMR spectroscopy

In 2D NMR spectroscopy magnetization is transferred between two nuclei to yield a two-dimensional spectrum. As mentioned before this reduces signal overlap and facilitates analysis. In homonuclear spectra correlations of spins of the same nuclei are observed either through bond or through space. Important experiments are for example the COSY (correlation spectroscopy) and the TOCSY (total correlation spectroscopy) which allow to determine connectivities of bonded protons. Besides that homonuclear experiments comprise through space experiments as well: In a $^{1}H,^{1}H$-NOESY correlations of protons in close proximity but not coupled via bonds are observed. An alternative experiment is the ROESY which has a modified pulse sequence but contains the same information as the NOESY. The NOE can be zero and therefor be not detectable while the ROESY always yields positive signals. These spectra give valuable distance information that are essential for structure calculations.

A 2D homonuclear NOESY can be sufficient for a partial assignment of nucleic acids: Imino protons are detected and linked to neighbouring residues due to cross peaks in the spectra. Therefor a sequential walk is possible and the assignment can be further extended to the imino-amino cross peak region and other spectra (see 4.1.2). Note that only residues with h-bonded imino protons can be assigned that way because all other imino protons exchange with the solvent and can hence not be detected.

### 1.4.7 2D heteronuclear NMR spectroscopy

Heteronuclear NMR spectra provide information of coupling spins of different atoms. The most useful experiments are HSQC experiments (Heteronuclear single-quantum correlation spectroscopy). Usually the correlation of a proton to one heteroatom, either nitrogen or carbon, is observed to obtain a $^{1}H,^{15}N$ or $^{1}H,^{13}C$-HSQC respectively. The $^{1}H,^{15}N$-HSQC is called fingerprint spectrum of a protein since they are unique for a given protein at defined experimental conditions and provide a fast impression of its integrity, structural changes or binding events. Note that the amide protons are prone to exchange with the solvent which complicates certain applications, e.g. solvent PRE measurements. An HMQC (Heteronuclear multiple-quantumcorrelation spectroscopy) can be recored to obtain the same information.

### 1.4.8 3D NMR – protein backbone assignment experiments

3D NMR experiments provide an even better resolution than 2D spectra. To record 3D experiments the magnetization needs to be transferred twice, e.g. from a proton to nitrogen, and subsequently to carbon as it is done in an HNCACB. Numerous experiments have been developed and allow backbone and side chain assignment of proteins and each of them comes with different pros and cons.

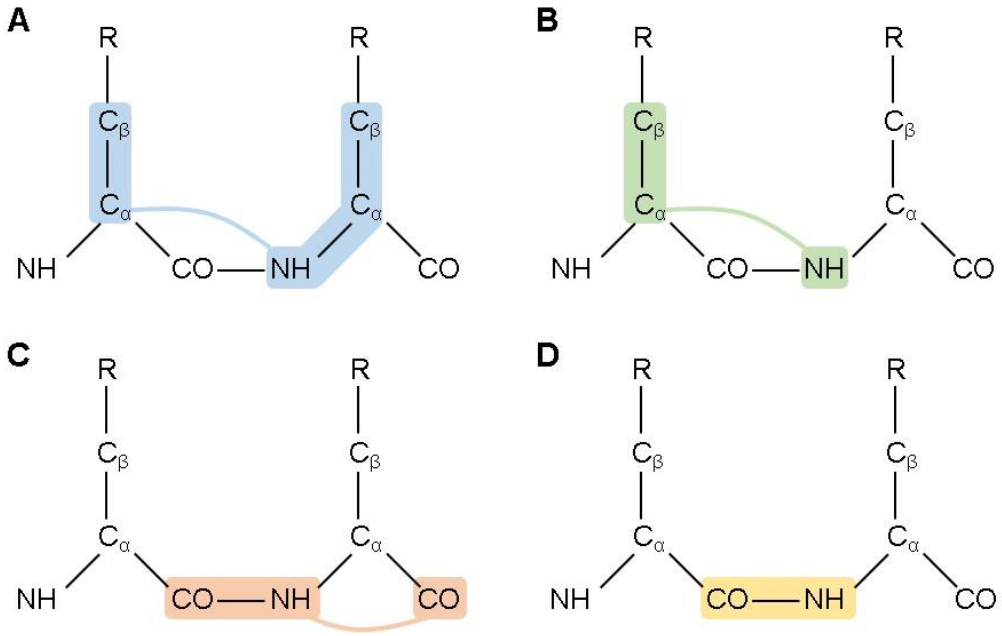



Figure 1.10: 3D experiments for protein backbone assignment. The cartoons show the connectivities obtained from the different NMR experiments discussed in the text highlighted in colour. **A)** The HNCACB links the amide group, the $C_\alpha$ and $C_\beta$ of a residue with those of the preceeding amino acid. **B)** The HNcoCACB provides information of the amide group of the residue with $C_\alpha$ and $C_\beta$ of the previous. Note that both A) and B) are available as HNCA versions where the $C_\beta$ is missing. In return these experiments provide a higher sensitivity. **C)** The HNcaCO links the carbonyl groups to the amide group while **D)** the HNCO connects an amide group to the preceeding carbonyl group.

A good set to assign the protein backbone could comprise a $^{1}H,^{15}N$-HSQC, HNCACB and HNcoCACB. The HNCACB links the $C_\alpha$ and $C_\beta$ of one residue with those of the preceding (meaning N-terminal) amino acid (Figure 1.10 A) while the HNcoCACB implies a step to filter magnetization transfer via the CO group in the peptide bond (Figure 1.10 B). Therefor only the preceding $C_\alpha$ and $C_\beta$ are visible allowing to distinguish those from its own carbon atoms. This set can be sufficient for suitable proteins but both 3D experiments are quite insensitive due to the transfer via $C_\beta$ atoms. The information can be complemented by HNCA and HNcoCA experiments which are similar to the previous set but avoid $C_\beta$ atoms and provide in return a higher sensitivity. Further experiments like HNCO (Figure 1.10 C) and the HNcaCO experiments (Figure 1.10 D) can confirm the assignment but are usually not sufficient on their own. Both experiments provide information of amide protons linked to preceeding (HNCO) or the preceding and its own (HNcaCO) carbonyl group. Other experiments and variations (like

CBCAcoNH, HccoNH) are available, but these experiments represent a good standard set and were used in this thesis.

### 1.4.9 Paramagnetic Relaxation Enhancement

Section 1.4.3 describes relaxation processes that always occur in an NMR experiment. Relaxation can be intentionally applied and analyzed. The introduction of paramagnetic centers leads to enhanced relaxation and thus a faster decay of the signal. Paramagnetic compounds contain an unpaired electron (Figure 1.11 A) that possesses a strong magnetic moment and interferes with the magnetic moment of the surrounding spins. Through dipole-dipole interactions it enhances the relaxation of these spins and leads to a lower signal intensity.

Paramagnetic compounds are metals, radicals or synthesized spin labels. For proteins various spin labels and metals are available and can be attached to the protein either via cysteines, unnatural amino acids or specific tags like DOTA-tags and lanthanide binding tags. Nucleic acids can also be labelled but require either incorporation of thiouridines or chemical synthesis. By comparing a paramagnetic (that is in presence of a free electron) and a diamagnetic sample (either a reduced paramagnetic or not spin-labelled sample) signals of residues in close proximity to the spin label show a reduced signal intensity (Figure 1.11 B, C). This so-called PRE effect (paramagnetic relaxation enhancement) is observable for distances up to approx. 20Å and by calibrating the intensities can be transformed into distance restraints. These are medium-range distance information and can thus be usefully combined with short-distance information from NOESY experiments for structure calculations. PREs also provide valuable information on dynamics and low-populated states. However, a careful analysis needs to be done since dynamics will always provide a population and time average.

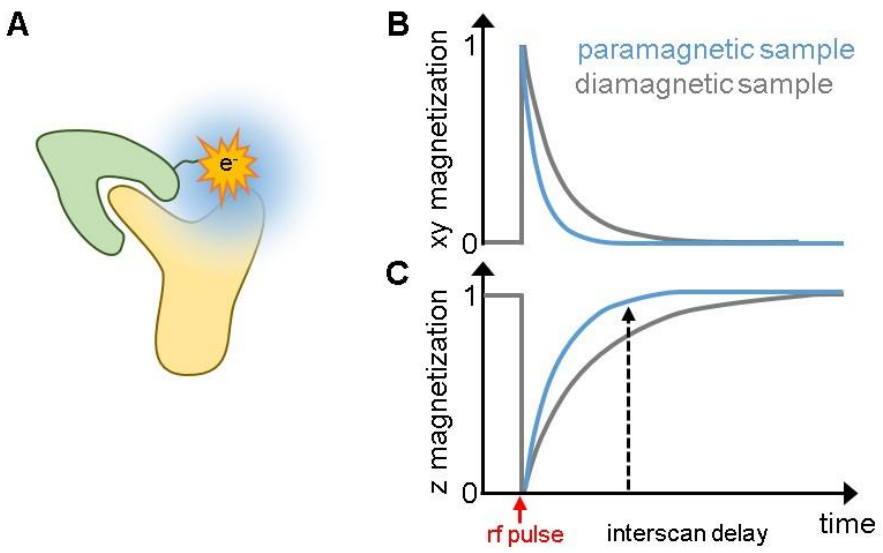



Figure 1.11: Principle of paramagnetic relaxation enhancement. **A)** A covalently attached spin label harbors an unpaired electron $e^-$. The PRE effect, a reduction in signal intensity, can be observed for all residues within a distance of approx. 20Å of this spin label. Thereby potential binding sites on interaction partners can be mapped. **B)** A comparison of magnetization loss in the xy plane (T2 relaxation) of the paramagnetic (blue) and diamagnetic (grey) sample shows that spin labels enhance relaxation and thereby decrease the dectectable signal intensity. **C)** Z magnetization or longitudinal relaxation is affected the same way and equilibrium magnetization is reached faster in presence of a spin label. Hence the interscan delay needs to be long enough to allow full magnetization recovery also for the diamagnetic sample. When the interscan delay is too short (dotted black arrow) it will cause artefacts.

PRE experiments are error prone due to several reasons: First, labelling of the sample might interfere with its structural integrity and remain incomplete. Hence, the sample would contain a mixture of para- and diamagnetic species which causes artefacts and wrong interpretation of the data. Second, mobility of the spin label itself needs to be taken into account with simulations and third, a longer interscan delay is required since for short delays initial z magnetization might have recovered completely for the paramagnetic species but not the diamagnetic one. This would reduce the maximum signal in the following experiment.

Paramagnetic compounds do not need to be covalently attached to the biomolecule but can be added to the solvent instead. This leads to signal loss of residues at the molecule's surface. This so-called solvent PRE gives thus information on solvent accessible areas and provides similar information as obtainable from H/D-exchange experiments.

# 2 Scope of the Thesis

Up to date several functions have been described for dsRBDs during mi- and siRNA maturation[65]: They are required for recognition as well as accurate positioning and cleavage of miRNA precursors[14,17]. Furthermore, dsRBPs recruit Dicer to its target, increase its affinity or enhance loading of the mature mi/siRNA onto Ago [22,77]. DsRBPs enable Dicer to processive processing of long dsRNAs and modulate substrate specificity[81]. For several dsRBPs a function in sensing RNA asymmetry has been described[79,80]. Even though the exact binding mode and binding interface is known, the substrate specificity seems to vary and whether dsRBPs bind sequence-dependently or -independently or recognize specific shapes is still a matter of current debate[64,73,74]. The proposed specificity contradicts the observation that several dsRBPs slide on duplex RNA and bind dynamically[76]. In this thesis Loqs-PD was analyzed for its RNA binding behavior and the protein-RNA complex was characterized. Potential functions were derived from these findings and the possible dynamics further analyzed. Functions and dynamics provide valuable insight into RISC formation and increase our understanding of this complex process.

Since most dsRBPs comprise multiple dsRBDs the individual dsRBDs might pursue distinct tasks within the protein. Therefore, the question arises whether the two dsRBDs 1 and 2 of mouse Staufen-2 contribute to RNA binding or have different functions. Using NMR the two domains were tested for potential RNA binding and possible binding sites were characterized in this work.

Another project aimed to determine the RNA substrate specificity of Mtr4 KOW. So far the KOW domain has been shown to bind tRNAs[199,210] but sequence or structure requirements for target recognition remain elusive. Therefore, different RNA types were tested in NMR experiments and their binding sites on KOW were determined. Furthermore, since Mtr4 and the exosome are recruited to ribosomal precursors via Nop53, the binding site of Nop53 was examined. The results shed some light on RNA recognition by Mtr4 and the recruiting mechanism via Nop53.

# 3 Materials and Methods

## 3.1 Chemicals

| | |
|---|---|
| $^{15}NH_4Cl$ | Cortecnet, Voisins-Le-Bretonneux, France |
| 99% Glycerol | VWR, Darmstadt, Germany |
| Acetic acid | Merck, Darmstadt, Germany |
| Agar | Sigma-Aldrich, Steinheim, Germany |
| APS | Sigma-Aldrich, Steinheim, Germany |
| Biotin (D-(+)) | Carl-Roth, Karlsruhe, Germany |
| Boric acid | Carl-Roth, Karlsruhe, Germany |
| $CaCl_2$ | Merck, Darmstadt, Germany |
| $CoCl_2$ x 6 $H_2O$ | Sigma-Aldrich, Steinheim, Germany |
| Coomassie Briliant Blue G250 | Applichem |
| $CuCl_2$ x 2 $H_2O$ | Sigma-Aldrich, Steinheim, Germany |
| DMSO | Merck, Darmstadt, Germany |
| DTT | Sigma-Aldrich, Steinheim, Germany |
| EDTA | Merck, Darmstadt, Germany |
| EtOH | VWR, Darmstadt, Germany |
| $FeCl_3$ x 6 $H_2O$ | Sigma-Aldrich, Steinheim, Germany |
| Glucose | Merck, Darmstadt, Germany |
| $H_3BO_3$ | Carl-Roth, Karlsruhe, Germany |
| HCl | VWR, Darmstadt, Germany |
| HEPES | VWR, Darmstadt, Germany |
| Imidazole | Merck, Darmstadt, Germany |
| IPTG | Carl-Roth, Karlsruhe, Germany |
| Kanamycin | Carl-Roth, Karlsruhe, Germany |
| $KH_2PO_4$ | Merck, Darmstadt, Germany |
| L-Ascorbic acid | Sigma-Aldrich, Steinheim, Germany |
| LiCl | Carl-Roth, Karlsruhe, Germany |
| $MgCl_2$ | VWR, Darmstadt, Germany |
| $MgSO_4$ | VWR, Darmstadt, Germany |
| $MnCl_2$ x 6 $H_2O$ | Merck, Darmstadt, Germany |
| $Na_2HPO_4$ | VWR, Darmstadt, Germany |
| NaCl | VWR, Darmstadt, Germany |
| $NaClO_4$ | Sigma-Aldrich, Steinheim, Germany |
| $NaH_2PO_4$ | VWR, Darmstadt, Germany |
| NaOH | Merck, Darmstadt, Germany |
| $Ni(III)Cl_2$ | Merck, Darmstadt, Germany |
| NTPs (ATP, CTP, GTP, UTP) | Sigma-Aldrich, Steinheim, Germany |
| PEG8000 | Pomega, Madison, USA |
| SDS | VWR, Darmstadt, Germany |
| Sodium acetate | Sigma-Aldrich, Steinheim, Germany |
| Spermidine | Sigma-Aldrich, Steinheim, Germany |
| Thiamin | Carl-Roth, Karlsruhe, Germany |
| Tris-aminomethane | VWR, Darmstadt, Germany |
| Tris-HCl | Amresco, Ohio, USA |
| Tryptone | Merck, Darmstadt, Germany |
| Urea | Carl-Roth, Karlsruhe, Germany |
| Yeast extract | Merck, Darmstadt, Germany |
| $ZnCl_2$ | Sigma-Aldrich, Steinheim, Germany |
| β-Mercaptoethanol | Sigma-Aldrich, Steinheim, Germany |

| | |
|---|---|
| $^{15}N,^{13}C$ NTPs (ATP, CTP, GTP, UTP) | Silantes, Munich, Germany |
| 20% Acrylamide/Bis-acrylamide (19:1) | Serva, Heidelberg, Germany |

## 3.2 Consumables

| | |
|---|---|
| 6x DNA loading dye | New England Biolabs, Ipswich, USA |
| Amicon MWCO 10k, 15mL | Sigma-Aldrich, Steinheim, Germany |
| Amicon MWCO 3k, 15mL | Sigma-Aldrich, Steinheim, Germany |
| DNA stain G | Serva, Heidelberg, Germany |
| Falcon tubes 15, 50mL | Greiner bio one, Kremsmünster, Austria |
| Filter 0.22/0.45µm | Sortarius, Göttingen, Germany |
| GeneRuler 100bp | ThermoFisherScientific, Waltham, USA |
| GeneRuler 1kb | ThermoFisherScientific, Waltham, USA |
| Lysozyme | Serva, Heidelberg, Germany |
| Ni-NTA resin | ThermoFisherScientific, Waltham, USA |
| Pipette tips 10, 200, 1000µL | StarLab, Berlin, Germany |
| Reaction tubes 1.5, 2.0mL | Eppendorf, Hamburg, Germany |
| TEV protease | Arie Geerlof, HMGU, Germany |
| Transferpette S 10, 100, 200, 1000µL | Brand, Wertheim, Germany |
| Wizard® Plus SV Minipreps DNA Purification System | Promega, Madison, USA |
| Wizard® SV Gel and PCR Clean-Up System | Promega, Madison, USA |
| NcoI | Promega, Madison, USA |
| BamHI | Promega, Madison, USA |
| T4 DNA Ligase | Promega, Madison, USA |

## 3.3 Devices

| | |
|---|---|
| Äkta FPLC | GE Healthcare, Little Chalfont, GB |
| Benchtop 2UV Transilluminator | UVP, Cambridge, GB |
| Blue Power 500 | Serva, Heidelberg, Germany |
| C1000 Thermal Cycler | Bio-Rad Laboratories GmbH, Munich, Germany |
| C1000 Thermocycler | BioRad, Munich, Germany |
| Centrifuge 5415 R | Eppendorf, Hamburg, Germany |
| Centrifuge 5810 R | Eppendorf, Hamburg, Germany |
| Geldoku | Bio-Rad Laboratories GmbH, Munich, Germany |
| HighLoad S75 16/60 Superdex column | GE Healthcare, Little Chalfont, GB |
| HPLC 1260 Infinity | Agilent Technologies, USA |
| K76 sonotrode | Bandelin electronic GmbH & Co. KG, Berlin, Germany |
| Mini-Sub Cell GT | Bio-Rad Laboratories GmbH, Munich, Germany |
| MR Hei-Mix L magnetic stirrer | Heidolph Instruments GmbH & Co.KG, Schwabach, Germany |
| Nanodrop Spectrophotometer ND-1000 | peqLab Biotechnologie GmbH, Erlangen, Germany |
| New Brunswick Scientific Incubator | Eppendorf, Hambug, Germany |
| New Brunswick Ultra-Low Temperature Freezer | Eppendorf, Hamburg, Germany |
| NewClassic MF fine scale | Mettler Toledo, Switzerland |
| NMR centrifuge 1011 | Hettich, Tuttlingen, Germany |
| NMR tubes | Norell, Morganton, USA |
| PowerPac HV 5000V, 500mA, 400W | Bio-Rad Laboratories GmbH, Munich, Germany |
| Rotamax 120 | Heidolph Instruments GmbH & Co.KG, Schwabach, Germany |

| | |
|---|---|
| ScoutPro 600g scale | OHAUS Europe GmbH, Greifensee, Switzerland |
| Shigemi tubes | Shigemi Inc., Allison Park, USA |
| SilentCrusher M | Heidolph Instruments GmbH & Co.KG, Schwabach, Germany |
| SmartSpec Plus Spectrophotometer | Bio-Rad Laboratories GmbH, Munich, Germany |
| Sonotrode UW 2200 | Bandelin, Berlin, Germany |
| Syringes 2, 5ml | Braun, Melsungen, Germany |
| Ultimate 3000 UHPLC$^{+}$ focused | Thermo Fisher Scientific, Waltham, USA |
| Vortexer | VWR, Darmstadt, Germany |
| ZelluTrans dialysis bags 2kDa, 5kDa | Roth, Karlsruhe, Germany |

## 3.4 NMR Spectrometers

**AV 600 Cryo**

- 600 MHz magnet
- Bruker Avance III HD console

**AV 900 Cryo**

- 900 MHz magnet
- Bruker Avance III HD console
- Cryo-TXI / $^{1}H$, $^{13}C$, $^{15}N$ probe, z-gradient

**AV 950 Cryo**

- 950 MHz magnet
- Bruker Avance III HD console
- Cryo-TXI / $^{1}H$, $^{13}C$, $^{15}N$ probe, z-gradient

## 3.5 Softwares and webservers

### 3.5.1 Software

Topspin 3.5, Bruker BioSpin GmbH

NMRFAM-Sparky 1.2, NMRFAM

Microsoft Excel 2013

Microsoft Word 2013

Pymol 1.8.6.0, Schrodinger, LLC

### 3.5.2 Webservers

All Webservers were accessed last time on 01.06.2017.

Echidna.biocenter.helsinki.fi/dali_server/start

atdbio.com/tools/oligo-calculator

biotools.nubic.northwestern.edu/OligoCalc.htmL

web.expasy.org/protparam/

swissmodel.expasy.org

## 3.6 Buffers, solutions, templates and constructs

### 3.6.1 Vectors

Table 3.1: Vectors used in this thesis.

| Plasmid | Promotor | Resistance | Tag | Cleavage site |
|---|---|---|---|---|
| pETM-11 | T7-lac | Kanamycin | N-6xHis-ZZ | TEV |

### 3.6.2 *E. coli* strains

Table 3.2: *E. coli* strains used in this thesis.

| ***E. coli* strain** | **Genotype** | **Application** |
|---|---|---|
| DH10 β | F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara leu) 7697 galU galK rpsL nupG λ- | Cloning<br>Amplification of plasmids |
| BL21 (DE) | F- ompT hsdSB(rB-, mB-) gal dcm rne 131 (DE3) | Protein expression |

### 3.6.3 Media and plates

Table 3.3: Media and plates and their applications.

| **Medium / Plate** | **Composition** | **Application** |
|---|---|---|
| LB | 10 g (w/v) Tryptone, 5 g (w/v) Yeast extract, 10 g (w/v) NaCl per liter | Expression of unlabelled protein<br>Minipreparation of plasmids |
| LB plate | LB medium, 15 g Agar per liter | Transformation |
| Trace elements (100x per liter) | 5 g EDTA pH 7.5, 0.83 g $FeCl_3$ x 6 $H_2O$, 84 mg $ZnCl_2$, 13 mg $CuCl_2$ x 2 $H_2O$, 10 mg $CoCl_2$ x 6 $H_2O$, 10 mg $H_3BO_3$, 1.6 mg $MnCl_2$ x 6 $H_2O$ | Expression of $^{15}N$ labelled protein |
| M9 minimal medium stock (10x) | 60 g $Na_2HPO_4$, 30 g $KH_2PO_4$, 5 g NaCl, 5 g $^{15}NH_4Cl$ | Expression of $^{15}N$ labelled protein |
| Minimal medium | 100 ml M9 (10x), 10 ml TE (100x), 20 ml 20 % (w/v) glucose, 1 ml 1 M $MgSO_4$, 0.3 ml 1 M $CaCl_2$, 1 ml Biotin (1 mg/ml), 1 ml Thiamin (1 mg/mL), 1 ml Kanamycin (50 mg/ml) | Expression of $^{15}N$ labelled protein |

### 3.6.3 Primer for linker mutant cloning

Table 3.4: Primers used to clone Loqs-PD linker mutants.

| Linker mutant | Primer | Sequence (5'-3') | Tm [°C] |
|---|---|---|---|
| All mutants (NcoI) | Loqs-PD$_{linker}$_for | CTTCACGGCCATGGGGGCAG | 60.0 |
| All mutants (BamHI/EcoRI) | Loqs-PD$_{linker}$_rev | GGAGCTCGAATTCGGATCCGGTACC | 62.6 |
| Loqs-PD$_{\Delta NC\Delta 10}$ | Linker$_{\Delta 10}$_for | ACT GGG GCC AAT GCC ACA GGC GGA GGA GAT GCC AGC | 69.8 |
| Loqs-PD$_{\Delta NC\Delta 10}$ | Linker$_{\Delta 10}$_rev | GGC ATT GGC CCC AGT CAC CGA CGG ACC AGC GGA GCT GCT | 71.2 |
| Loqs-PD$_{\Delta NC\Delta 20}$ | Linker$_{\Delta 20}$_for | GCT GGA GGA GGA GAT GCC AGC GAC AAG ACC GTT GGT | 67.1 |
| Loqs-PD$_{\Delta NC\Delta 20}$ | Linker$_{\Delta 20}$_rev | ATC TCC TCC TCC AGC GGA GCT GCT AGG CGA TTC CGG CAG | 69.8 |
| Loqs-PD$_{\Delta NC\Delta 30}$ | Linker$_{\Delta 30}$_for | TCG CCT GAC AAG ACC GTT GGT AAT CCG ATT GGC TG | 62.9 |
| Loqs-PD$_{\Delta NC\Delta 30}$ | Linker$_{\Delta 30}$_rev | GGT CTT GTC AGG CGA TTC CGG CAG CTG CGC GCC GAT C | 74.4 |

### 3.6.4 Protein constucts used

Table 3.5: Overview of all protein constructs used. For each protein the origin organism and the wildtype residues are listed.

| Protein | Organism | Wildtype residues |
|---|---|---|
| Loqs dsRBD1 | *D. melanogaster* | 129-211 |
| Loqs dsRBD2 | *D. melanogaster* | 245-322 |
| Loqs-PD$_{\Delta NC}$ (dsRBD1-2) | *D. melanogaster* | 129-322 |
| Loqs-PD$_{\Delta N}$ (dsRBD1-2-N) | *D. melanogaster* | 129-359 |
| Loqs-PD$_{\Delta NC\Delta 11}$ | *D. melanogaster* | 129-322$_{\Delta 224-234}$ |
| Loqs-PD$_{\Delta NC\Delta 21}$ | *D. melanogaster* | 129-322$_{\Delta 219-239}$ |
| Loqs-PD$_{\Delta NC\Delta 31}$ | *D. melanogaster* | 129-322$_{\Delta 214-244}$ |
| Loqs-PD$_{\Delta NC\Delta 41}$ | *D. melanogaster* | 129-322$_{\Delta 209-249}$ |
| Staufen 2 dsRBD1 | *M. musculus* | 1-93 |
| Staufen 2 dsRBD2 | *M. musculus* | 75-208 |
| Staufen 2 dsRBD1-2 | *M. musculus* | 1-208 |
| Mtr4 KOW | *S. cerevisiae* | 666-818 |
| Mtr4 KOW extended (stalk) | *S. cerevisiae* | 638-842 |
| Nop53 AIM peptide | *S. cerevisiae* | 58-92 |

### 3.6.5 Buffers and solutions for protein purification

Table 3.6: Buffer composition and solutions for protein purification and analysis. For each buffer the application is listed on the right.

| Buffer / Solution | Composition | Application |
|---|---|---|
| Lysis buffer | 1 M NaCl, 50 mM Tris pH 8.0, 5 mM β-Mercaptoethanol, 5 % Glycerol | Loqs purification 1st Ni-NTA |
| Wash buffer 1 | 1 M LiCl, 50 mM Tris pH 8.0, 5 mM β-Mercaptoethanol, 2 mM Imidazole, 5 % Glycerol | Loqs purification 1st Ni-NTA |
| Wash buffer 2 | 1.5 M NaCl, 50 mM Tris pH 7.2, 5 mM β-Mercaptoethanol, 2 mM Imidazole, 5 % Glycerol | Loqs purification 1st Ni-NTA |
| Elution buffer | 1 M NaCl, 50 mM Tris pH 7.0, 5 mM β-Mercaptoethanol, 1 M Imidazole, 5 % Glycerol | Loqs purification 1st Ni-NTA |
| Imidazole | 2 M Imidazole pH 8.0 | Loqs purification 1st Ni-NTA |
| TEV buffer | 1 M NaCl, 50 mM Tris pH 7.0, 5 mM β-Mercaptoethanol, 5 % Glycerol | Loqs TEV cleavage<br>Loqs dialysis<br>Loqs purification 1st Ni-NTA |
| Wash buffer 3 | 1 M NaCl, 50 mM Tris pH 7.0, 5 mM β-Mercaptoethanol, 5 % Glycerol + 20, 50, 100 mM Imidazole | Loqs purification 2nd Ni-NTA |
| SEC buffer | 1 M NaCl, 50 mM Tris pH 7.0, 5 mM β-Mercaptoethanol, 3 % Glycerol | Loqs purification SEC<br>Storage |
| SDS staining solution | 2.5 g Coomassie Blue, 450 ml EtOH, 400 ml $H_2O$, 100 ml acetic acid | SDS-PAGE |
| SDS destaining solution | 450 ml EtOH, 400 ml $H_2O$, 100 ml acetic acid | SDS-PAGE |

### 3.6.6 In vitro transcription templates

Table 3.7: The DNA templates used for RNA in vitro transcription. The DNA sequence is listed in 5’ to 3’ direction.

| RNA / top strand | Template strand 5’ - 3’ |
|---|---|
| T7 top strand | TAATACGACTCACTA |
| 21 bp siRNA 1 (guide) | AAGGAGATCATTTTGAAAGCTCCTATAGTGAGTCTATTA |
| 21 bp siRNA 2 (passenger) | AAGGAGCTTTCAAAATGATCTCCTATAGTGAGTCGTATTA |
| Flipped 21 bp RNA 1 | AAGGCGATCATTTTGAAAGCTACTATAGTGAGTCGTATTA |
| Flipped 21 bp RNA 2 | AAGTAGCTTTCAAAATGATCGCCTATAGTGAGTCGTATTA |
| Inside GC 21 bp RNA 1 | AAGTA GATTACGCCGAAATCTACTATAGTGAGTCGTATTA |
| Inside GC 21 bp RNA 2 | AAGTAGATTTCGGCGTAATCTACTATAGTGAGTCGTATTA |
| GC poor 21 bp RNA 1 | AAGTAGATCATTTTGAAAGTTACTATAGTGAGTCGTATTA |
| GC poor 21 bp RNA 2 | AAGTAACTTTCAAAATGATCTACTATAGTGAGTCGTATTA |
| GC rich 21 bp RNA 1 | AAGGAGCTCGCTCCGACAGCTCCTATAGTGAGTCGTATTA |
| GC rich 21 bp RNA 2 | AAGGAGCTGTCGGAGCGAGCTCCTATAGTGAGTCGTATTA |
| Palindromic RNA 16 | AAGACTAGTCGACTAGTCTATAGTGAGTCGTATTA |
| Palindromic RNA 20 | AAGACTAGTCAGCTGACTAGTCTATAGTGAGTCGTATTA |
| Palindromic RNA 24 | AAGACTAGTCGACTAGTCGACTAGTCTATAGTGAGTCGTATTA |
| Palindromic RNA 28 | AAGACTAGTCAGCTAGCTAGCTGACTAGTCTATAGTGAGTCGTATTA |
| Palindromic RNA 32 | AAGACTAGTCGACTAGTCGACTAGTCGACTAGTCTATAGTGAGTCGTATTA |
| Palindromic RNA 40 | GACTAGTCGACTAGTCGACTAGTCGACTAGTCGACTAGTCTATAGTGAGTC GTATTA |

### 3.6.7 Buffers and solutions for RNA transcription and purification

Table 3.8: Composition and application of buffers and solutions used for RNA production and purification.

| Buffer / Solution | Composition | Application |
|---|---|---|
| 10x TRX | 500 mM Tris-HCl pH 7.5, 150 mM $MgCl_2$, 50 mM DTT, 20 mM Spermidine | In vitro transcription |
| Sodium acetate | 3 M Sodium acetate pH 5.5 | RNA precipitation |
| 10x TBE | 890 mM Tris-aminomethane, 890 mM Boric acid, 20 mM EDTA pH 8.0 | RNA-PAGE |
| Gel stock | 20% Acrylamide/Bis-acrylamide (19:1), 8 M Urea, 1x TBE | RNA-PAGE |
| Loading buffer | 1x TBE, 40% (v/v) glycerol | RNA-PAGE |
| HPLC running buffer | 12.5 mM Tris-HCl pH 8.0, 6 M Urea | RNA HPLC purification |
| HPLC elution buffer | 12.5 mM Tris-HCl pH 8.0, 6 M Urea, 0.5M $NaClO_4$ | RNA HPLC purification |
| Sodium chloride | 1 M NaCl | RNA dialysis |

### 3.6.8 Buffers for NMR

Table 3.9: NMR buffer compositions and their applications.

| Buffer | Composition | Application |
|---|---|---|
| Loqs NMR | 20 mM HEPES pH 7.2, 500 mM NaCl | Loqs-RNA interaction |
| RNA NMR | 20 mM $(Na_2H/NaH_2)PO_4$ pH 6.5, 50 mM NaCl | RNA assignment |
| Mtr4 KOW NMR | 20 mM HEPES pH 6.5, 150 mM NaCl, 2 mM β-Mercaptoethanol, 5 % Glycerol | KOW assignment<br>KOW-ligand interactions |
| Staufen NMR | 20 mM $(Na_2H/NaH_2)PO_4$ pH 7.0, 5 mM DTT | Staufen assignment<br>Rgs4 SRS2 RNA assignment<br>Staufen-RNA interaction |

## 3.7 Molecular biology

### 3.7.1 Cell culture for cloning and plasmid isolation

Small cultures of 5-10 ml were inoculated with a single colony from an LB plate. The cultures were incubated at 37°C and 160 rpm overnight. Afterwards the cells were pelleted by centrifugation at 4 °C for 15 min and 12000 rpm.

### 3.7.2 Minipreparation of plasmids

Plasmids were isolated from pelleted cultures (see 3.7.1) according to the manual of the manufacturer. First the cells were resuspended and lysed and after addition of alkaline protease solution the suspension was incubated for 5-10 min. Upon addition of neutralization solution cell debri and genomic DNA was pelleted by centrifugation at top speed for 10 minutes at room temperature. Afterwards the supernatant was transferred into a spin column and centrifuged for 1 min at top speed. The column was washed twice with wash solution and afterwards the plasmid was eluted by adding water and centrifuging at top speed for 1 min.

### 3.7.3 Cloning of Loqs-PD$_{\Delta NC}$ linker mutants

To clone Loqs-PD$_{\Delta NC}$ mutants with varying linker lengths two-step PCRs were made. First one of the flanking primers Loqs-PD$_{linker}$_for or Loqs-PD$_{linker}$_rev was combined with Linker$_X$_rev or Linker$_X$_for respectively. The resulting PCR products comprised one of the dsRBDs and a fragment of the native linker that overlapped partially with the other fragment. The two products were purified via gel extraction and used as a template for the second PCR. Here the two flanking primers were used to amplify both dsRBDs and the respective part of the linker. After purification the PCR product was enzymatically digested with NcoI and BamHI to produce sticky ends. These were compatible with those ends from the standard pETM-11 vector digested with the same enzymes. Therefore, the product could be inserted into the vector by T4 DNA ligation. After transformation in *E. coli* positive clones were confirmed by sequencing.

## 3.8 Protein expression and purification of Loqs

### 3.8.1 Cell culture for protein expression

For protein expression of the different Loqs-PD constructs a 50 ml preculture with kanamycin was inoculated with several colonies from a freshly transformed LB plate and grown overnight at 37 °C. The preculture was completely transferred to 1 l of the respective medium. As soon as the culture reached an OD600 of 0.7-0.9 the culture was incubated with IPTG and further cultivated at 18°C for protein expression overnight. The culture was harvested by centrifugation at 4000rpm and 4°C for 20 minutes. The pellet was transferred to 50 ml falcon tubes and either stored at -20 °C or directly lysed. For $^{15}$N labeling of the protein $^{15}$N labeled minimal medium was used for both preculture and main culture.

### 3.8.2 Cell lysis

Unless otherwise stated, cell pellets were dissolved in Ni-NTA lysis buffer and disrupted using a sonicator with 50 %, 0.5 cycle 8 times while cooling on ice. After sonication the solution was centrifuged at 12000 rpm and 4 °C for 40 min.

### 3.8.3 Protein purification

After sonication and centrifugation the supernatant was applied to approx. 4 ml Ni-NTA twice. The column was washed with approx. four column volumes of wash buffer 1 to remove nucleic acids and 8 column volumes of wash buffer 2. The 6xHis-tagged protein was eluted with 3 column volumes of elution buffer. Afterwards the column was regenerated by washing with water. The elution fraction was dialyzed in dialyze bags using a 10 kDa cutoff for the tandem domain construct or 3 kDa cutoff for the individual domains against TEV buffer. TEV protease was added to the dialysis bags and both cleavage and dialysis were performed overnight at 4 °C. The protein was loaded onto the Ni-NTA column twice again. Afterwards the column was washed with 4 column volumes each of wash buffer 3 with increasing imidazole concentrations. Since the protein was unspecifically bound to both the sticky ZZ-tag and the Ni-NTA cleaved protein could be recovered from the column. Both the TEV-

protease as well as the ZZ-tag were eluted using elution buffer. All fractions containing the cleaved product were pooled and concentrated using either a 10 or 3 kDa cutoff amicon concentrator at 4 °C, depending whether tandem or single dsRBDs were purified. 2 ml of the protein solution were loaded using a 2 ml loop onto a S75 16/60 Superdex size exclusion chromatography column. SEC was performed at a flow rate of 1 ml/min at 4 °C and fractions of 1.5 ml were collected. The protein was analyzed on SDS gels and stored at 4 °C until usage.

### 3.8.4 SDS-PAGE

Prior to SDS-PAGE analysis protein samples of 15 µl were mixed with 6x SDS loading buffer and incubated at 95 °C for 5 minutes. Afterwards the samples were briefly centrifuged and loaded on 10 or 15 well gels. The gels were run at 50 mA per gel for 35 minutes. Afterwards the gels were washed with water and stained with staining solution. After 15 minutes the staining solution was discarded, the gel was washed several times with water and incubated with destaining solution. Usually after further 20 minutes the gels could be analyzed on an imager or the GelDoku.

### 3.8.5 Protein quantification

To determine the protein concentration a Nanodrop was used.

## 3.9 RNA transcription and purification

### 3.9.1 RNA in vitro transcription optimization

To determine the optimal $MgCl_2$ concentration for each individual in vitro transcription, small test transcriptions with increasing $MgCl_2$ concentrations were performed. 10 µl of each test transcription were mixed with loading buffer and loaded onto a 20 % polyacrylamide gel with 8 M urea. The gels were run at constant 10 W for 30-45 min.

### 3.9.2 RNA large scale in vitro transcription

Based on the optimization series the protocol for each RNA was adjusted. The preparative transcriptions were performed in volumes of 5 or 10 ml as described before[215]. Transcription was run at 37 °C for approx. 4 hours. Afterwards the RNAs were precipitated using 0.1 volume of sodium acetate pH 5.5 and 3.5 volumes of cooled ethanol.

### 3.9.3 RNA purification via denaturing HPLC

The precipitated RNAs were pelleted by centrifugation at 4°C and 12000 rpm for 20 min. The supernatant was removed and the pellet air dried for 20 min. Afterwards the RNA was resuspendended in water and loaded via a 5 ml loop onto the HPLC. For separation a DNAPac Pa200 Prep Scale column was used at a flow rate of 10 ml/min at 85 °C and a 0-100 % gradient of $NaClO_4$ was applied. Fractions of 2 ml were collected and analyzed on a polyacrylamide gel. All fractions containing the pure target RNA were pooled and dialyzed for further usage.

### 3.9.4 RNA purification via denaturing polyacrylamide gel

If not via HPLC the RNAs were purified via denaturing polyacrylamide gels containing 8 M urea. After precipitation and removal of ethanol (see previous section) the RNA was resuspended in 4 ml 1 x TBE buffer and mixed with loading buffer. The sample was loaded onto the gel and the gel ran for approx. 15 hrs at 30 W. Afterwards the RNA bands were visualized under UV. The target band was cut out using a scalpel and afterwards electro eluted.

### 3.9.5 Electroelution of RNA gels

To extract the purified RNAs from polyacrylamide gels electro elution was used. The chamber was assembled and the gel loaded in small fragments into the chambers. The system was filled with 1 x TBE buffer and electro elution took place at 4 °C and 200 V. Elution of RNA

was regularly monitored by analyzing samples with the Nanodrop. Elution was essentially complete after 6-7 hrs.

### 3.9.6 Dialysis and lyophylization of RNAs

RNAs after electroelution or HPLC purified were loaded into dialysis bags with a cutoff of 2 kDa and first dialyzed twice against 1 liter of 1 M NaCl and afterwards twice against 1 l water at 4 °C. Using liquid nitrogen the RNA was than frozen in 50 ml falcon tubes. With hot needles small holes were pricked out of the falcon lids and the falcons were placed in a glas vessel. The glas vessel was attached to the lyophilizer and lyophylization was performed until the pellet was completely dried. The RNA was afterwards stored at -80 °C or dissolved in water for usage.

### 3.9.7 RNA quantification

As for the protein samples RNA concentration was determined using a Nanodrop. Usually 100 x dilutions of the concentrated stocks were prepared with water.

## 3.10 Biophysical experiments

### 3.10.1 Static light scattering

To determine the stoichiometry of the Loqs-RNA complex static light scattering was run on 3.8 mg/ml Loqs-PD$_{\Delta NC}$, 1800 ng/µl 21 bp siRNA and a 3:1 mixture (protein:RNA) of 4.1 mg/ml Loqs-PD$_{\Delta NC}$, each in 250 mM NaCl, 20 mM Tris pH 6.5. The samples were first passed via an analytical size exclusion column (GE Superdex200 10/300, Aekta Purifier) at 4 °C and a flow rate of 0.5 ml/min and then analyzed by a coupled Viscotek TDA 305 triple-assay detector. For referencing a 4 mg/ml solution of bovine serum albumin with a molecular weight of 66.4 kDa for the monomer was used and calculations were based on assumed dn/dc value of 0.185 ml/g for proteins, 0.165 ml/g for RNA and 0.177 for 1:1 protein:RNA complexes. The experimentally determined refractive indices and right-angle light-scattering were used

to calculate the molecular masses with the Omnisec software (Malvern Instruments). The SLS runs and analysis were performed by Dr. Arie Geerlof.

### 3.10.2 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) provides valuable information on the overall shape of a molecule, its molecular weight and thus potential oligomerization as well as structural changes. It is easy accessible and with powerful beamlines single experiments can be recorded within seconds. The sample provided in a capillary is exposed to an X-ray. The sample molecules diffract the X-ray beam and the intensity distribution is detected. Like NMR SAXS is a solution technique but due to the free tumbling of the molecules only low resolution data can be obtained. In addition, the high energetic X-ray beam may cause beam damage of the sample. Different plots obtained from SAXS measurements (Guinier, Kratky, P(r)) provide information like the radius of gyration, the Porod volume and therefore the molecular weight, the maximum distance within the molecule and the overall shape.

SAXS experiments for the Loqs-PD project were run on a Rigaku BIOSAXS1000 instrument. It was equipped with a HF007 microfocus generator and a Cu-target at 40 kV and 30 mA. Images were collected with a Pilatus 100 k detector and a photodiode beamstop was used to measure the transmission. Q-calibration was made with a silver-behenate sample. For all measurements multiple 900 sec frames were collected, checked for possible beam damages and afterwards averaged. Rigaku SAXSLab and GNOM software were used for circular averaging, background substraction and calculation of the distance distribution functions. The molecular weights were calculated based on the POROD volumes. SAXS measurements and analysis were done by Dr. Ralf Stehle.

### 3.10.3 Circular dichroism spectroscopy

All CD temperature series were recorded on a JASCO-J715 spectropolarimeter to determine the melting temperatures of RNAs. As buffer the respective RNA NMR buffer was used. For temperature regulation a Peltier-type control system (PTC-348WI) and a Julabo 725 thermostat were used. The power supply was a PS-150J.

## 3.11 NMR spectroscopy

### 3.11.1 NMR sample preparation

After changing buffer to the respective NMR buffer and adjusting the concentration 5 % $D_2O$ were added to the samples for locking. The samples were usually loaded into a 5 mm NMR tube, except for small volume RNA samples where a Shigemi tube was used. For titration experiments small volumes of the ligand were directly added to the NMR tube and the solutions were mixed by carefully inverting the tube several times. To get rid of potential air bubbles the tubes were briefly spun down in an NMR centrifuge. All NMR spectra were recorded at 298 K if not otherwise mentioned.

### 3.11.2 RNA assignment

To assign double-stranded RNA $^{1}H,^{1}H$-NOESY spectra were recorded. For a reasonable resolution at least 128 and 256 points were collected in the direct and indirect dimension respectively, usually more points were sampled. Since for every residue ideally two cross peaks in the imino region should occur (accept for terminal base pairs) a sequential walk connecting all neighboured residues was made. When needed the assignment was further confirmed by analyzing the imino-amino cross peaks. Here uracils and guanosines show characteristic patterns. For the isotopically labelled 21 bp siRNA $^{1}H,^{15}N$-HSQCs were recorded to distinguish uracils and guanosines.

### 3.11.3 Protein backbone assignment

The Mtr4 KOW domain was assigned using HNCACB, HNCA, HNcoCA, HNCO and HNcaCO spectra as described in 4.4.1 using a $^{15}N,^{13}C$ labelled approx. 300 µM sample.

Based on the HNCACB spectra secondary chemical shifts were calculated as follows:

$$\text{secondary chemical shifts [ppm]} = \Delta\delta C_{\alpha} - \Delta\delta C_{\beta} \qquad (8)$$

### 3.11.4 Titrations and calculation of chemical shift perturbations

To titrate ligands $^{1}H,^{15}N$-HSQCs of the free protein were recorded. Upon addition of small amounts of ligand the same experiment was run again. All settings were checked prior to measurement but usually required hardly any adjustment. Ligands were added until no changes in the spectra could be observed to assure full saturation of the protein. Since Mtr4 KOW has a low affinity for its ligands no saturation even in high excess was achieved.

Chemical shift perturbations (CSP) were calculated by using chemical shifts of the free protein and at highest ligand concentration according to formula 9.

$$\mathrm{CSP\ [ppm]} = \sqrt{\Delta H^2 + \left(\frac{\Delta N}{6.51}\right)^2} \qquad (9)$$

Here the spectra of the free and saturated protein were compared.

### 3.11.5 Paramagnetic relaxation enhancement experiments and analysis

For PRE measurements the two siRNA strands were mixed in a 1:1.1 ratio of spin-labelled to non-labelled strand to ensure the spin-labelled strand was completely bound. Protein and dsRNA were mixed in a 1:1.5 ratio to create the complex. $^{1}H,^{15}N$-HSQCs with high numbers of scans were recorded to achieve reasonable signal-to-noise ratios. Afterwards the spin label was reduced by addition of pH adjusted 5-fold excess of ascorbic acid and the same experiment was recorded. 1D spectra were measured to confirm integrity of the spin label and successful reduction. The spectra were processed using exactly the same settings and analyzed using Sparky. To determine the PRE the intensities were analyzed using the Data height function. The PRE effect was calculated as the following:

$$\mathrm{Line\ broadening} = \frac{I_{para}}{I_{dia}} \qquad (10)$$

$$\sigma^{+}_{\mathrm{line\ broadening}} = \left(\frac{I_{para}+N_{para}}{I_{dia}-N_{dia}}\right) - \frac{I_{para}}{I_{dia}} \qquad (11)$$

$$\sigma^{-}_{\mathrm{line\ broadening}} = \frac{I_{para}}{I_{dia}} - \left(\frac{I_{para}-N_{para}}{I_{dia}+N_{dia}}\right) \qquad (12)$$

### 3.11.6 Line broadening experiments and analysis

To evaluate the exchange during RNA binding of Loqs line broadening for different protein constructs, RNA substrates and under different conditions was compared. Initially a $^{1}H,^{15}N$-HSQC spectrum was recorded for the free protein and afterwards the RNA was carefully added in excess and gently mixed by inverting the NMR tube several times. Afterwards another $^{1}H,^{15}N$-HSQC was recorded with either exactly the same setting or an increased number of scans to compensate for the loss of signal intensity. In the latter case this scaling factor was calculated out for the analysis.

For analysis only signals that did not overlap, meaning which were sufficiently separated in the spectra, in neither the free or bound form were used. All other signals were not taken into account. The data height of the signals used was determined using Sparky and the ratio thereof for the free / bound spectra yielded the line broadening factor according to formula 13. The error was calculated as the maximum error range according to formulas 14 and 15 based on the noise of 10000 sampled points within each spectrum. Note that the formulas are identical to those used for PRE calculations (10, 11, 12; see 3.11.5). No internal referencing was made since all unstructured regions (linker and termini) always showed nice ratios of approx. 1 and were thus not affected by RNA binding. This proved the experiments to be comparable.

$$\text{Line broadening} = \frac{I_{free}}{I_{bound}} \qquad (13)$$

$$\sigma^{+}_{\text{line broadening}} = \left(\frac{I_{free}+N_{free}}{I_{bound}-N_{bound}}\right) - \frac{I_{free}}{I_{bound}} \qquad (14)$$

$$\sigma^{-}_{\text{line broadening}} = \frac{I_{free}}{I_{bound}} - \left(\frac{I_{free}-N_{free}}{I_{bound}+N_{bound}}\right) \qquad (15)$$

### 3.11.7 Heteronulcear NOE experiments and data analysis

The heteronuclear NOE experiment provides information on local flexibility. The experiment comprises two spectra: In one all protons are saturated while the other one corresponds to a regular $^{1}H,^{15}N$-HSQC. The intensity ratio gives information on the heteronuclear cross relaxation rate of the amide vectors and the size of this steady-state NOE is an indicator for flexibility and motion of the backbone.

The flexibility of the KOW domain in the free form and bound to ligands was analyzed using heteronuclear NOE experiments. All experiments were set up the same way and both processing and analysis was the same. The data was recorded interleaved and the saturated and non-saturated spectra were split afterwards. The hetNOE values were calculated according to equations 13-15 from 3.11.6.

# 4 Results

## 4.1 SiRNA binding of Loqs-PD

### 4.1.1 Summary of previous results

In a previous study the two individual dsRBD structures of Loqs-PD have been determined by NMR spectroscopy by Thomas Kern and Christoph Hartlmüller. Both dsRBDs exhibit the canonical αββββα-fold with two α-helices packed against a three-stranded β-sheet. $\beta_1$ and $\beta_2$ are connected by a flexible loop. Comparison of relaxation data and overlay of $^{1}H,^{15}N$-HSQC spectra of the single domains and the tandem domain construct Loqs-$PD_{\Delta NC}$ revealed that both domains are independent of each other. Upon titration of a 21 bp siRNA the single domains bound the RNA via the conserved binding mode: Both helices and the loop were involved in RNA binding. A similar pattern of chemical shift perturbations was observed for the tandem domain RNA titration indicating that both domains bind simultaneously to RNA and do not influence each other. This work was done by Thomas Kern. Our collaborators, Stephanie Fesser and Klaus Förstemann, determined affinities for the individual domains and the tandem construct towards different substrates based on EMSAs and fluorescence anisotropy: Both domains bound duplex RNA with nanomolar affinity (approx. 200 nM) and the tandem domains did not show cooperativity (approx. 80 – 100 nM). All constructs showed a weak preference for perfect matching dsRNA and seemed to require a certain length for full affinity binding.

Since many diverse functions of dsRBPs have been described and their contribution to RISC formation is not clearly defined the complex of Loqs-$PD_{\Delta NC}$ and a 21 bp siRNA is described further in the following sections.

### 4.1.2 In vitro transcription and analysis of 21 bp siRNA

The RNA used in this study is based on the *Drosophila* miRNA *bantam* which was slightly varied to function as an siRNA. The thermodynamic stability of both termini was calculated using the atdbio oligo calculator (atdbio.com/tools/oligo-calculator). To analyze

the RNP complex the RNA was transcribed in vitro and purified by myself (Figure 4.1) which brought two advantages: First, larger amounts of the RNA were available to perform all required experiments and second, the RNA could be isotopically labeled which allowed further NMR experiments (Figure 4.2). The RNA was the same as used before by Dr. Thomas Kern except for the two initial bases that were changed to guanines to increase transcription efficiency (compare Figure 4.1 A and Figure 4.8 A).

The in vitro transcription was optimized for both individual strands (Figure 4.1 A) separately. For both strands the target band showed an intensity maximum over a broad concentration range of $MgCl_2$ (Figure 4.1 B). Therefor a $MgCl_2$ concentration of 40 mM was used for all transcriptions of both strands. The RNA was either purified via denaturing PAGE using 8 M urea (see 3.9.4) or HPLC with 6 M urea at 85 °C (see 3.9.3, Figure 4.1 C). Both techniques provided the required resolution to separate the target band from the aborts (Figure 4.1 C). For gel-purified RNA electroelution was necessary to extract the RNA from the gel. After dialysis and lyophylization the two strands were mixed in equimolar amounts and 1D NMR spectra were recorded to check integrity and quality of the sample (Figure 4.1 D).

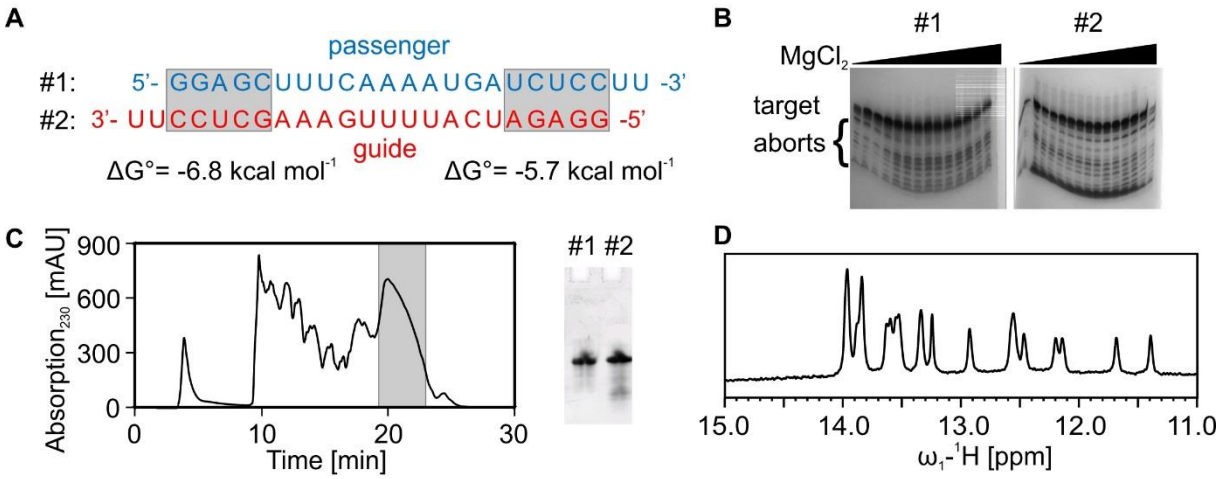



Figure 4.1: Transcription and purification of 21 bp siRNA. **A)** Sequences of the two strands of the 21 bp siRNA. The passenger strand comprising the more stable 5' end is shown in blue, the guide strand in red. Usually the guide strand is loaded onto Ago to form the RISC. The calculated stability of the termini using the first five base pairs (grey boxes) is shown below the sequence. **B)** Denaturing PAGE of transcriptions for strand 1 and 2 using different $MgCl_2$ concentrations. The target product band is most intense and migrates the slowest. All other bands are aborts. **C)** Denaturing HPLC chromatogram of strand 2 purification. The grey box indicates the target product. The two purified strands are shown on a denaturing PAGE on the right. **D)** $^1H$ NMR spectrum of imino region of 21 bp siRNA confirms duplex character.

The assignment of the RNA was required to obtain further information. Therefor a $^{1}$H,$^{1}$H-NOESY was recorded (see 1.4.6, Figure 4.2 B) which contains the same information as the 1D spectrum (Figure 4.1 D) but additional cross peaks between neighbouring residues allow a sequential walk. This sequential walk links residues adjacent within the sequence (Figure 4.2 A) via hydrogen bonds of the imino protons of guanines and uracils (purple dottet line in Figure 4.2 B). Most residues have two crosspeaks corresponding to the previous and the following base pair. Only terminal residues show one crosspeaks since they lack one neighbouring base pair. While all guanines within the 21 bp siRNA show nice dispersion and could be assigned easily the uracils overlap severely. Especially those residues within the two uracil stretches cluster together. The imino-amino cross peak region of the $^{1}$H,$^{1}$H-NOESY (Figure 4.2 C) helped to differentiate between guanines and uracils in the imino region since both bases show characteristic cross peaks in this area. To confirm the base identity a $^{1}$H,$^{15}$N-HSQC was recorded on an isotopically labeled sample (Figure 4.2 D). Here both bases cluster in distinct regions and allow to distinguish them. Note that the shift in the direct dimension is the same as in the $^{1}$H,$^{1}$H-NOESY and the assignment can thus be transferred between the spectra. Since in both spectra several residues overlapped two samples were prepared where either strand was $^{15}$N,$^{13}$C-labeled while the respective pairing strand was not. It was crucial to add the unlabeled strand in slight excess to have the labeled strand completely bound. For both samples a $^{1}$H,$^{15}$N-HSQC was recorded which revealed which residue was part of the passenger or guide strand (Figure 4.2 E). The same samples were used to record a $^{1}$H,$^{13}$C-HSQC (Figure 4.2 F). Making use of uracil crosspeaks (blue dotted line in Figure 4.2 C) which correspond to the pairing and neighbouring adenine H2 the adenines could be assigned (Figure 4.2 F). Thus, easy accessible probes across the entire RNA sequence were assigned and available for further experiments. Additionally, the assignment confirmed the sequence and structure integrity of the RNA. This RNA was used for all following experiments unless otherwise stated.

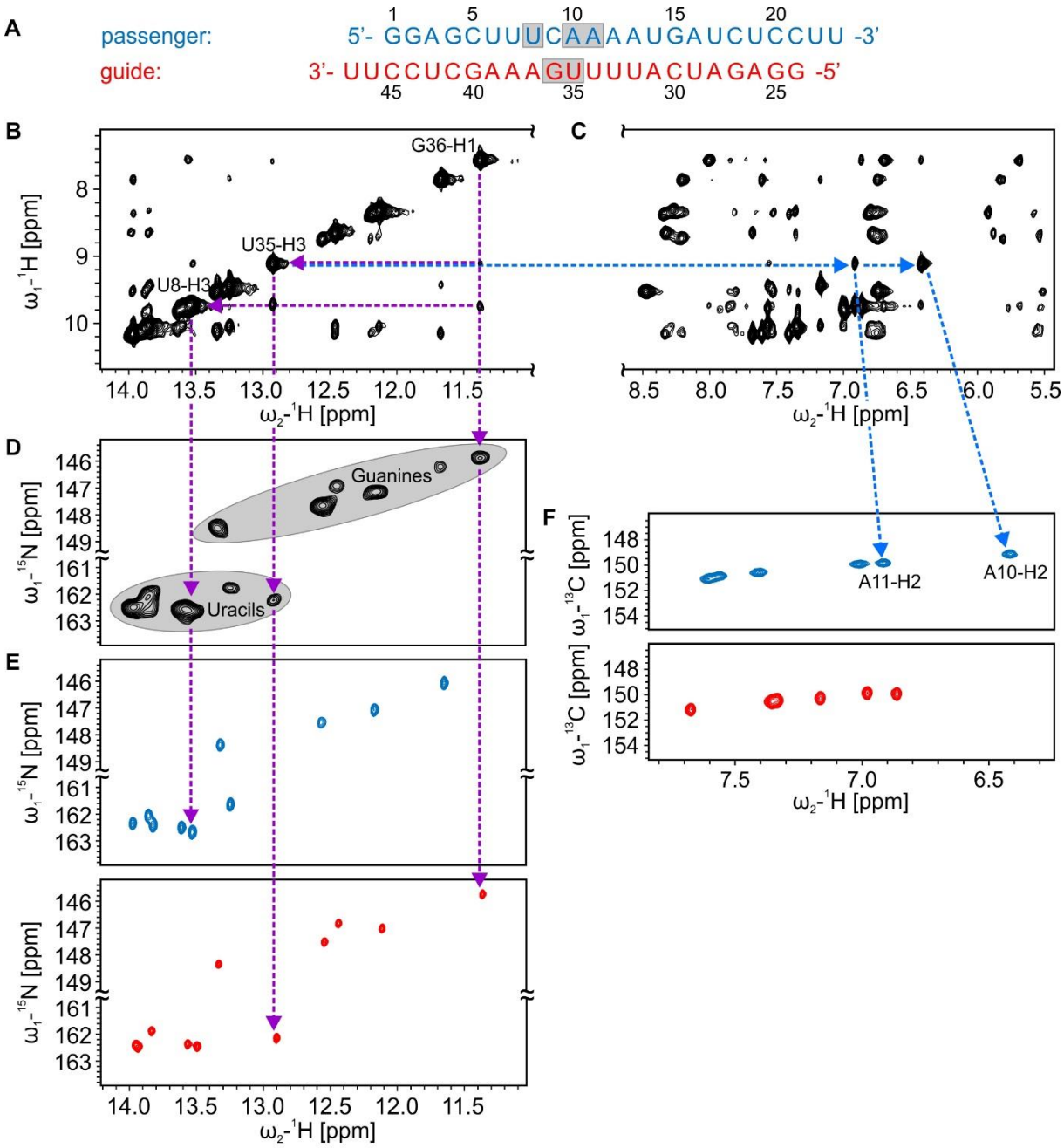



Figure 4.2: Duplex 21 bp siRNA assignment based on homonuclear NOESY spectrum. **A)** Sequence of the 21 bp siRNA. Grey boxes indicate residues for which the assignment is illustrated in this figure. **B)** $^{1}$H,$^{1}$H-NOESY of the imino region of the RNA. Purple dotted lines indicate the sequential walk starting at G36. Note that diagonal peaks correspond to imino protons of guanines and uracils while adenines and cytosines are invisible in this type of spectrum. **C)** The imino-amino cross peak region of the $^{1}$H,$^{1}$H-NOESY shows characteristic peaks corresponding to guanines or uraciles. **D)** $^{1}$H,$^{15}$N-HSQC of isotopically labeled RNA. Guanines and uracils cluster in different regions (shaded in grey) and thus allow to identify residues in B). **E)** To minimize overlap $^{1}$H,$^{15}$N-HSQC of RNAs with either strand labeled were recorded (blue spectrum: passenger strand, red spectrum: guide strand). **F)** $^{1}$H,$^{13}$C-HSQC of the samples used in E). Starting with imino diagonal peaks in B) and going over corresponding cross peaks of A-U base pairs in C) the adenine H2 signals could be assigned in F).

### 4.1.3 Loqs-PD$_{\Delta NC}$ binds dsRNA in a 1:1 complex

As mentioned before (4.1.1) Loqs-PD$_{\Delta NC}$ binds dsRNA via the conserved binding mode (Figure 4.3 A). Addition of RNA caused significant chemical shift perturbations in the respective regions. For a further description of the complex static light scattering (SLS) was used which allows to determine molecular weights and stoichiometries (Figure 4.3 B, see 3.10.1) based on a calibration with BSA. For Loqs-PD$_{\Delta NC}$ a molecular weight of 24.8 kDa was determined while RNA alone yielded 14.3 kDa. The deviation from the theoretical values (21.1 kDa for Loqs-PD$_{\Delta NC}$, 13.8 kDa for the RNA) is acceptable and within the error. Both molecules were monomeric at the given concentration and did not show any sign of degradation. For the mixture of the two with a 3-fold excess of protein a molecular weight of the main species with 37.9 kDa was determined. This corresponds to a 1:1 complex when comparing with the two measurements of the individual components and the result was again within the error. All SLS measurements and analyses were done by Arie Geerlof.

Another suitable technique to determine overall shapes and distances is small angle X-ray scattering (SAXS) (see 3.10.2). SAXS was run on Loqs-PD$_{\Delta NC}$, the 21 bp siRNA and the complex (Figure 4.3 C) as it was done before with SLS. Based on these measurements the maximum distances were determined: The RNA had a maximum distance of approx. 65 Å while the $D_{max}$ for Loqs-PD$_{\Delta NC}$ varied between 100-110 Å. This variation was likely caused by sample quality and possible aggregation. The large distance of the free protein reflects an average population and indicates that the two domains were not attached to each other and sampled different conformations. Interestingly, the complex had a maximum distance of approx. 110 Å. Furthermore, it showed an elongated shape based on the P(r) curve. All SAXS measurements and analyses were done by Ralf Stehle.

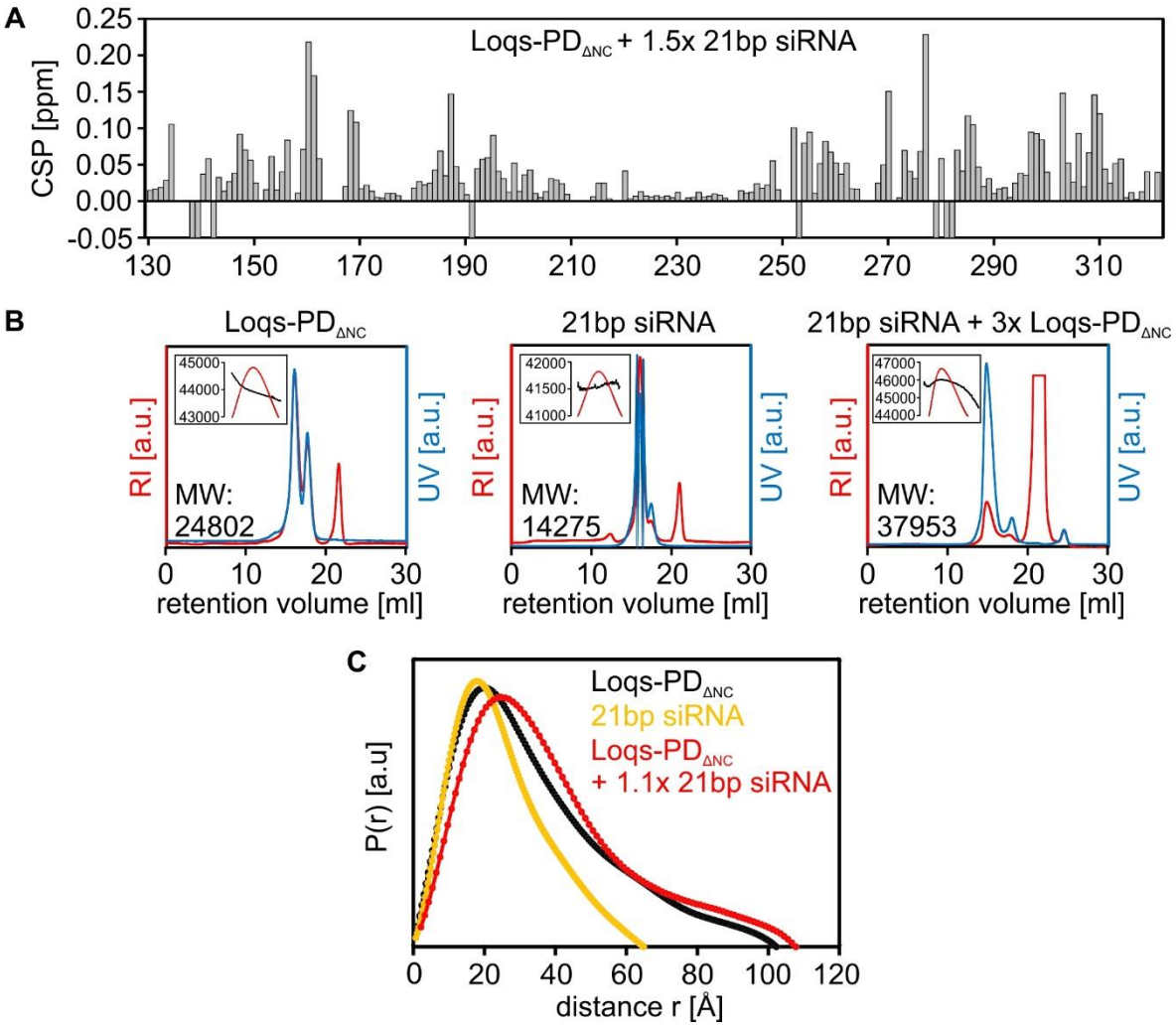



Figure 4.3: Biophysical characterization of Loqs-PD$_{\Delta NC}$ - 21 bp siRNA complex. **A)** Chemical shift perturbations (CSPs) for Loqs-PD$_{\Delta NC}$ upon addition of 1.5-fold 21 bp siRNA. High values indicate stronger shifts in the respective $^{1}H,^{15}N$-HSQC while negative values represent residues that could not be tracked during the titration. **B)** Static light scattering (SLS) for Loqs-PD$_{\Delta NC}$, the 21bp siRNA and a complex thereof. The red curves show the refractive index, the blue curves the UV absorption. In the small boxes the black line indicates the particle size at a given retention volume and is an indicator of sample homogeneity. The derived molecular weights are shown in the graphs and reveal a 1:1 binding. **C)** P(r) curves obtained from small angle X-ray scattering (SAXS) for similar samples as in B). The RNA (yellow) showed a maximum distance of 60 Å while both the protein (black) and the complex (red) comprised distances of 100-110 Å.

### 4.1.4 Both dsRBDs bind RNA independently and compete for binding

Previously recorded relaxation data indicated that the two domains are independent in the free form and titrations suggested the same for the RNA bound form (work by Thomas Kern). To confirm the independence in the RNA bound form both single dsRBDs 1 and 2 were

mixed in an equimolar ratio. The $^{1}H,^{15}N$-HSQC overlaps perfectly with those from the individual dsRBDs which confirms that the domains do not contact each other (Supplementary Figure 6.1). This is in agreement with the observation that the tandem construct spectrum overlaps with the individual domain spectra (Thomas Kern, data not shown).

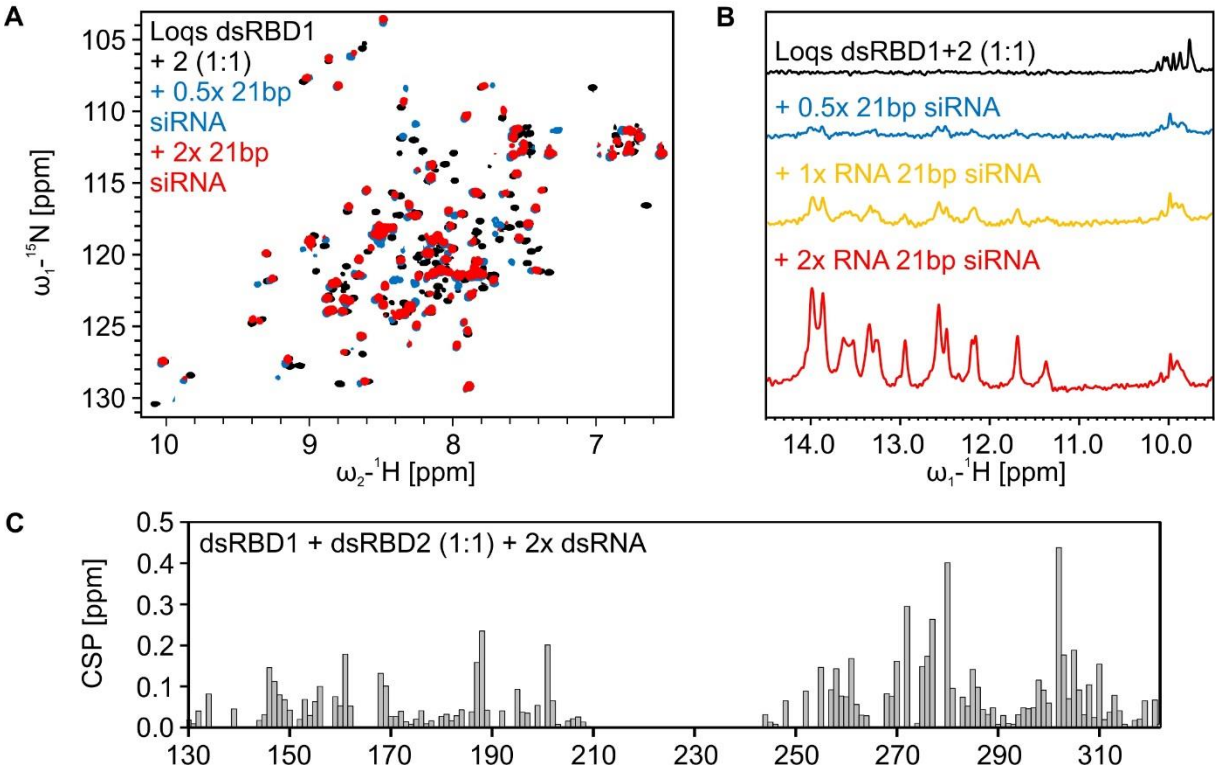



Figure 4.4: Loqs dsRBDs are independent and compete for RNA binding. **A)** $^{1}H,^{15}N$-HSQC spectra of Loqs dsRBD1 and 2 in a 1:1 ratio (black) and in presence of 0.5-fold (blue) and 2-fold (red) 21 bp siRNA. The amount of RNA refers to the total protein amount, meaning the sum of dsRBD1 and 2. **B)** $^{1}H$ NMR spectra of the imino region for free protein (black) and for the same protein:RNA ratios as in A) plus a 1:1 ratio (yellow) confirm the structural integrity of the RNA duplex and show severe line broadening at sub-stoichiometric ratios.

The dsRBD mixture was titrated with the 21 bp siRNA and chemical shifts were observed (Figure 4.4 A). As for the single dsRBD and the tandem domain titrations line broadening occurred in both the protein and the RNA spectra (Figure 4.4 B). The line broadening at sub-stoichiometric concentrations indicates a dynamic binding mode potentially involving sliding of the domains on the duplex RNA. Interestingly, saturation was only achieved at a protein:RNA ratio of 1:1.5 referring to the total protein amount and not the individual domain concentrations. This confirms that binding occurs at a 1:1 stoichiometry

and only a single dsRBD binds an RNA duplex. In the mixture the two dsRBDs thus compete for the RNA substrate. Since saturation for the tandem domain protein was also reached at a ratio of approx. 1:1.5 and both domains bound the RNA these results suggest that the linker is required to ensure that both domains bind the same RNA molecule simultaneously. Possibly the dynamics within the system do not allow more domains to bind the RNA without the linker. The independence in RNA binding is further confirmed by the CSPs (Figure 4.4 C) of the mixed dsRBDs which corresponds to those observed for the single dsRBD titrations (Thomas Kern, data not shown).

### 4.1.5 The linker does not affect Loqs' domain structure

As shown before the linker seems to be required for simultaneous binding of two dsRBDs. Its influence on the complex remains elusive, however. Therefore Loqs-PD$_{\Delta NC}$ linker mutants were generated with varying linker lengths. All linker mutants expressed nicely and could be purified with the wildtype protocol. Overall solubility seemed to decrease with shorter linker and the protein was more expressed in inclusion bodies. A comparison of the $^{1}H,^{15}N$-HSQC of the linker mutants with wildtype Loqs-PD$_{\Delta NC}$ and the CSPs (Figure 4.5 A-D) reveals that shortening the linker hardly influenced the free protein. In fact only neglectable changes within the domains were observed and moderate shifts occured in the regions adjacent to the respective deletion (Figure 4.5 C). The overall CSPs for Loqs-PD$_{\Delta NC\Delta 41}$ were lower. In the spectra several residues shifted into different directions than for the other linker mutants. Obviously, a very short linker exhibits restrictions on the tandem domains.

For two linker mutants SAXS was measured (Figure 4.5 E). The maximum distance got shorter with decreasing linker length. Furthermore, the protein exhibited a more compact overall shape. A short linker thus does not influence the individual domain structure but limits flexibility and leads to a more overall compact state.

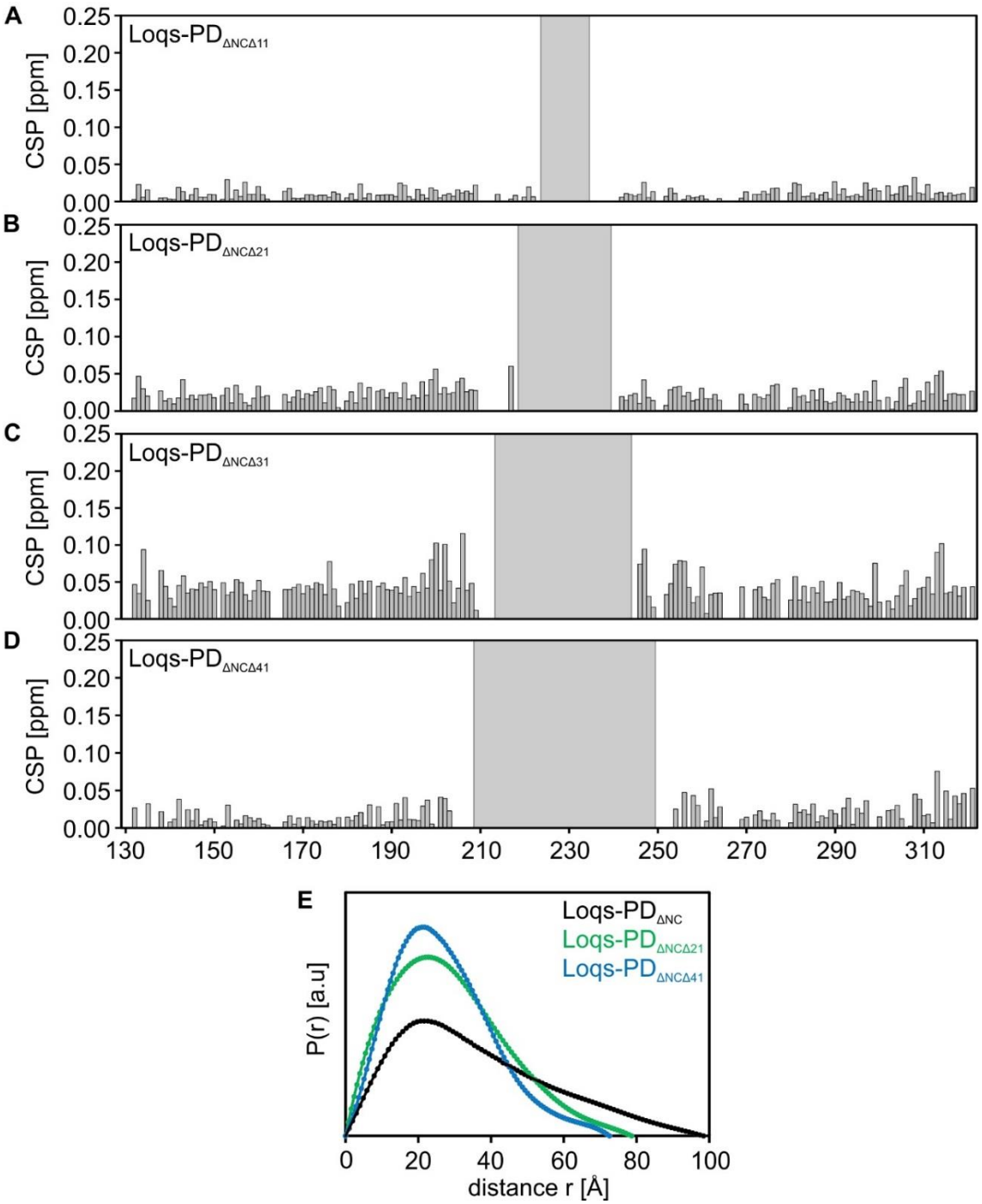



Figure 4.5: Chemical shift perturbations of Loqs-PD$_{\Delta NC}$ linker mutants. Chemical shift perturbations of **A)** Loqs-PD$_{\Delta NC\Delta 11}$, **B)** Loqs-PD$_{\Delta NC\Delta 21}$, **C)** Loqs-PD$_{\Delta NC\Delta 31}$ and **D)** Loqs-PD$_{\Delta NC\Delta 41}$ compared to Loqs-PD$_{\Delta NC}$. The grey boxes indicate the deleted residues. **E)** P(r) curves for Loqs-PD$_{\Delta NC}$ (black), Loqs-PD$_{\Delta NC\Delta 21}$ (green) and Loqs-PD$_{\Delta NC\Delta 41}$ (dark blue).

### 4.1.6 Linker mutants bind RNA via conserved binding interface and interact with Dicer

The same linker mutants were used for RNA titrations. For Loqs-PD$_{\Delta NC\Delta 11}$, Loqs-PD$_{\Delta NC\Delta 21}$ and Loqs-PD$_{\Delta NC\Delta 31}$ similar CSPs were observed (Figure 4.6 A-C) corresponding to the

conserved binding interface. The complex spectra quality decreased and line broadening got more severe for shorter linkers which lead to several signals broadened beyond detection (Figure 4.6 D). For Loqs-PD$_{\Delta NC\Delta 41}$ only the C-terminal residue D322 was observable, all other signals were not detectable. This indicates either dynamic binding or the formation of higher oligomers where the increased molecular weight causes line broadening. The almost complete deletion of the linker in this mutant might cause restrictions that prevent native RNA binding or proper domain orientation within the complex.

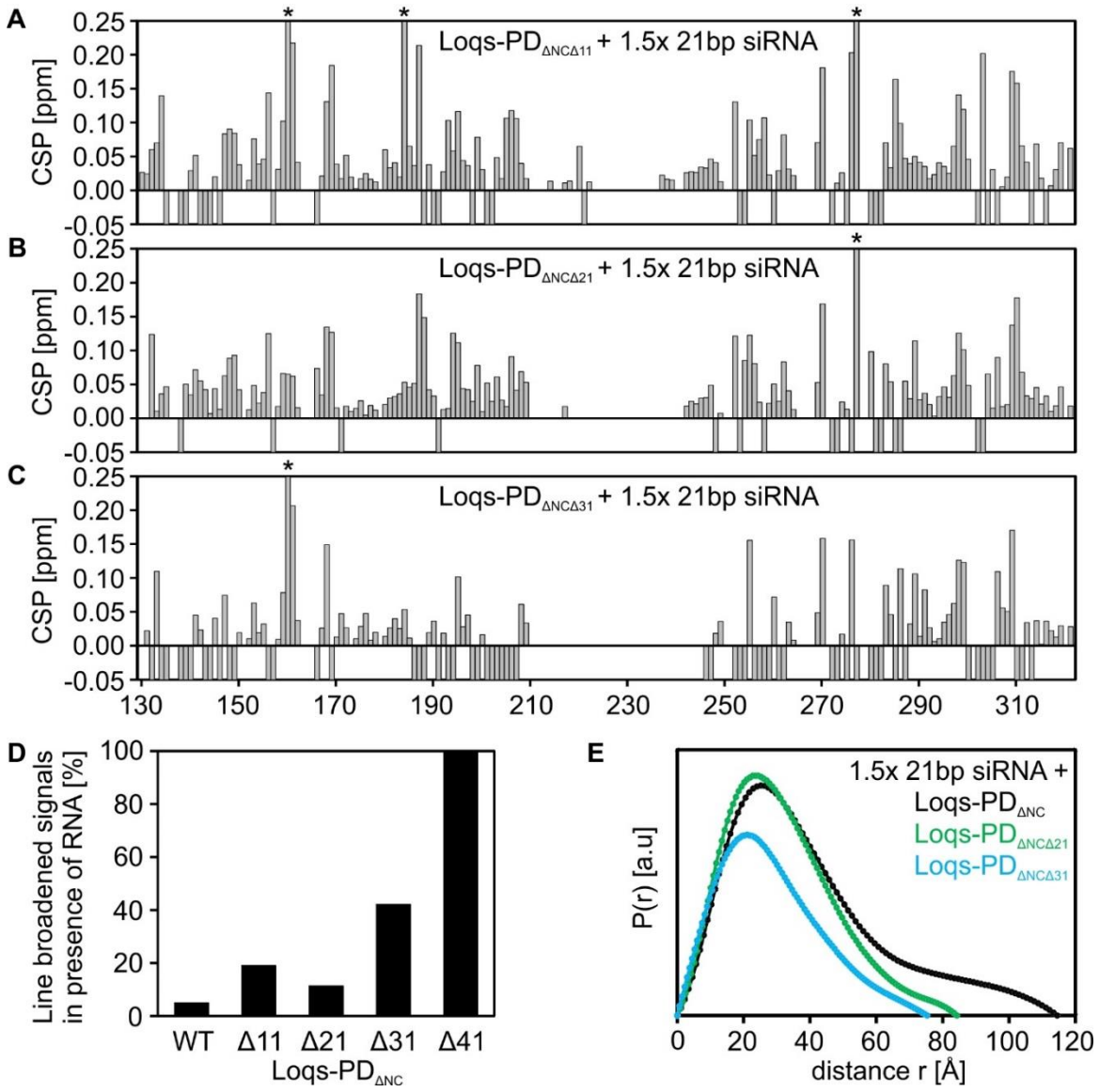



Figure 4.6: RNA titrations of Loqs-PD$_{\Delta NC}$ linker mutants. Chemical shift perturbations of **A)** Loqs-PD$_{\Delta NC\Delta 11}$, **B)** Loqs-PD$_{\Delta NC\Delta 21}$ and **C)** Loqs-PD$_{\Delta NC\Delta 31}$ in complex with 21 bp siRNA. Negative bars indicate not traceable shifts while residues marked with an asterisk exhibit higher CSPs. **D)** Amount of signals that are line broadened beyond detection in presence of RNA. **E)** SAXS P(r) curves for protein-RNA complex with Loqs-PD$_{\Delta NC}$ (black), Loqs-PD$_{\Delta NC\Delta 21}$ (green) and Loqs-PD$_{\Delta NC\Delta 31}$ (blue).

SAXS was measured for protein-RNA complexes with two linker mutants (Figure 4.6 E): The $D_{max}$ decreased with shorter linkers similarly to the free protein (compare to Figure 4.5 E). The complexes thus exhibited a more compact fold and the maximum distance seems to be regulated by the linker. Stephanie Fesser and Klaus Förstemann determined the affinity of Loqs-$PD_{\Delta NC\Delta 41}$ to be comparable to the wildtype protein. Via co-immunoprecipitation and western blots they could also prove that Loqs-$PD_{\Delta 41}$ still interacts with Dicer-2 in cells (Figure 4.7).

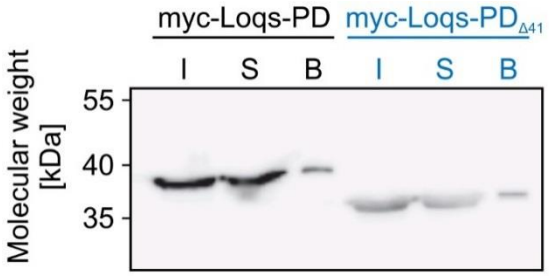



Figure 4.7: Co-immunoprecipitation of Loqs-PD wildtype and Loqs-$PD_{\Delta 41}$ with Dicer-2. Western blot of Co-IP of N-terminally myc-tagged transfected wildtype Loqs-PD and the linker mutant with Dicer-2. A S2 cell line was used which had an genomically encoded $Flag_2$-tag attached to Dicer-2. I: Input, S: Supernatant, B: Bound. Image adapted from Stephanie Fesser and Klaus Förstemann.

Summarizing the results a shorter linker does neither affect the domain structure nor canonical RNA binding. However, below a certain threshold the linker could limit dynamics and flexibility and force the complex into an unfavourable mode or conformation. Hence, the ideal domain arrangement might be perturbed. Despite conformational limitations a short linker does not perturb the interaction with Dicer-2.

### 4.1.7 Loqs stabilizes the RNA duplex and has different effects on both strands

So far the binding site of Loqs on the dsRNA remains elusive. However, Stephanie Fesser and Klaus Förstemann performed cross-linking experiments which revealed all Loqs protein constructs (the single domains as well as the tandem domain construct and Loqs-PD full length) do bind everywhere onto the RNA but preferentially at the termini (Figure 4.8). The RNA used differed slightly from that one used in this thesis (compare to Figure 4.1 A). At different positions a thiouridine was incorporated which allowed cross-linking of the protein.

Quantification of cross-linking was done via signal intensity of the radioactively labeled RNA (Figure 4.8 B, C).

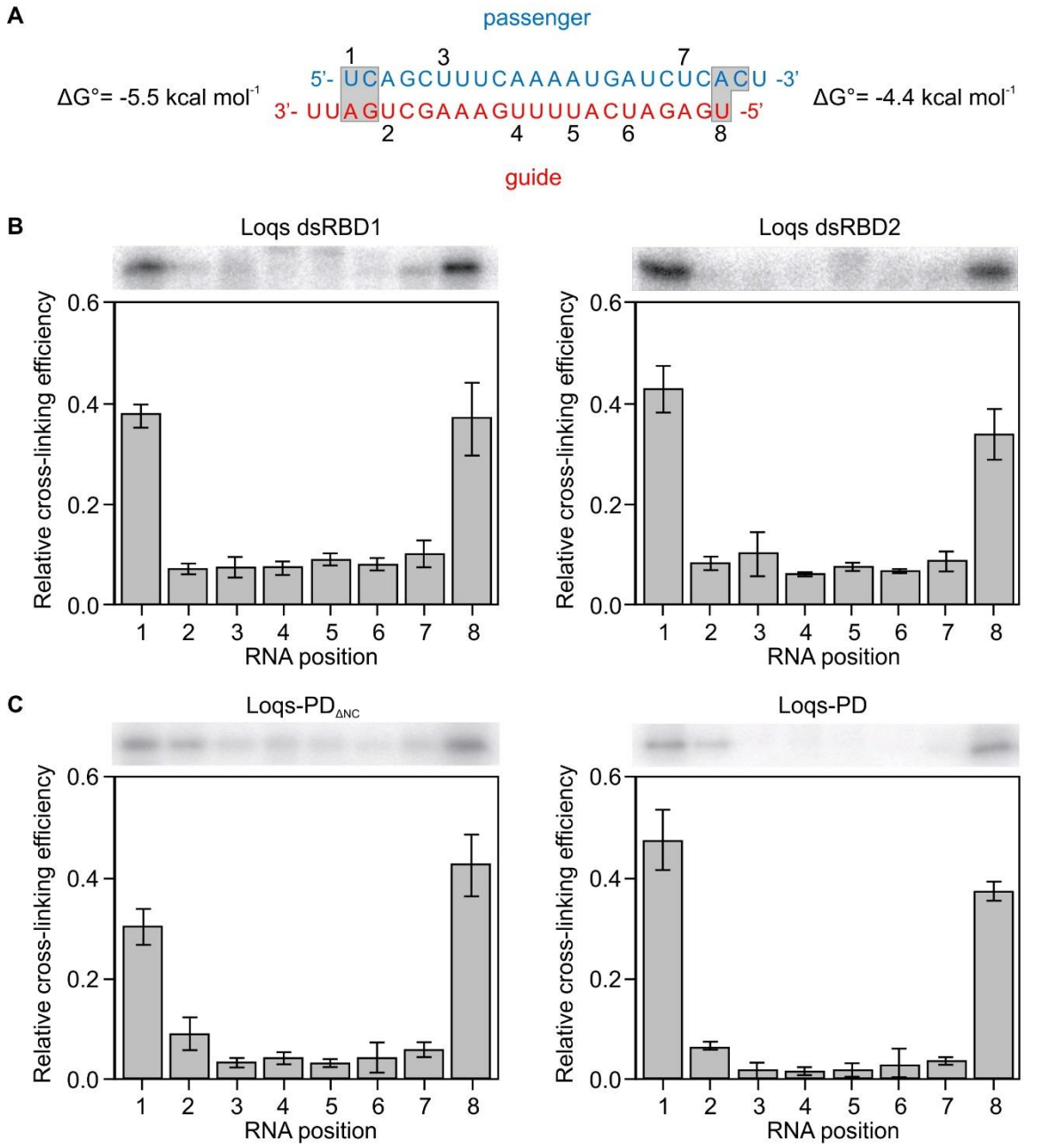



Figure 4.8: Crosslinking of Loqs-PD constructs to siRNA reveals terminal binding preference. **A)** Sequence of the *bantam* derived 21 bp siRNA. As before, the blue strand reflects the passenger strand, the red one the guide strand. The thermodynamic stabilities are given besides. The grey boxes indicate residues that differ from the RNA used for the NMR experiments in this thesis. Numbers 1-8 highlight positions where a thiouridine was incorporated and cross-linking occurred. **B)** Images of the radioactively labelled RNA show intense signals at the RNA termini for both single dsRBDs. The signals were quantified and plotted as relative cross-linking efficiency below. **C)** The same experiment was done for Loqs-PD$_{\Delta NC}$ and Loqs-PD full length. Image adapted from Stephanie Fesser and Klaus Förstemann.

To further describe the binding of Loqs from the RNA's point of view the probes described in 4.1.2 were used in NMR experiments. Both $^{1}H,^{15}N$-HSQC (Figure 4.9 A) as well as $^{1}H,^{13}C$-HSQC spectra (Figure 4.9 B) were recorded for the free RNA and in presence of excess unlabeled Loqs-$PD_{\Delta NC}$. For both experiments two samples were prepared with either strand isotopically labeled to reduce signal overlap (see 4.1.2). All complex spectra suffered from severe line broadening and most signals were broadened beyond detection. Probably higher concentrated samples and longer measurement times could bring back some of the signals but this approach was not further pursued. The loss of the majority of signals strongly points at a dynamic binding mode, potentially involving sliding. Together with the cross-linking data (Figure 4.8) this suggests that Loqs-PD slides on the RNA duplex and may reside more frequently at the termini.

Interestingly, all signals of the passenger strand could be detected while several residues, especially in the $^{1}H,^{15}N$-HSQC, of the guide strand remained clearly visible though line broadened as well (Figure 4.9 A). The experiments were repeated to confirm this finding. Also a few adenine H2 signals could be detected (Figure 4.9 B) close to the noise level. Adenine and guanine H8 signals in contrast vanished completely. All residues that showed a detectable signal intensity in any spectrum cluster in the guide strand (Figure 4.9 C). This suggests that both strands are not equally affected by Loqs and binding seems to be asymmetric. Interestingly, in both strands a cluster of signals that corresponds to either cytosines or uracils H8 remained visible (Figure 4.9 B). Here several peaks overlapped and individual residues could not be assigned. A proper interpretation of these results seems impossible so far since the lack of detectable complex signals for the passenger strand cannot be rationalized. As mentioned before optimization of the experiment could potentially recover some signals, nevertheless the two strands would be differently affected by Loqs binding.

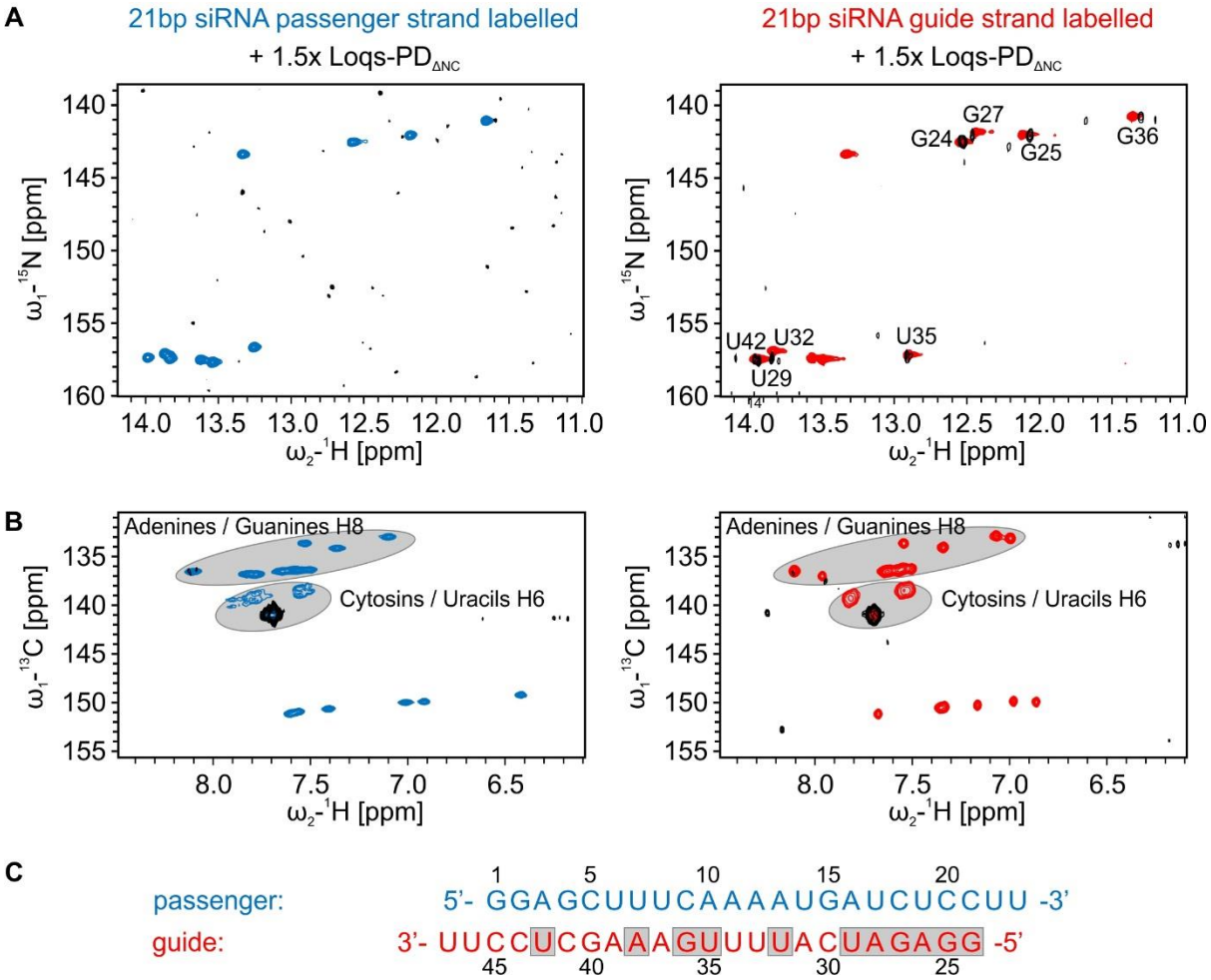



Figure 4.9: Protein bound RNA shows severe line broadening. **A)** $^{1}$H,$^{15}$N-HSQC spectra for free 21 bp siRNA (colored) and in presence of 1.5-fold Loqs-PD$_{\Delta NC}$ (black). The spectrum with the passenger strand isotopically labeled is shown in blue, while the one with the labeled guide strand is shown in red. Detectable signals in the complex are labeled. **B)** $^{1}$H,$^{13}$C-HSQC spectra of the same samples. For the guide strand labeled sample (red) some adenine H2 residues are detectable but are close to the noise and therefore not visible here. Grey spheres indicate clusters of either adenine and guanine H8 signals or cytosine and uracil H6 signals. Interestingly, one cluster of either cytosines or uracils remains visible in the protein bound form. **C)** All residues with detectable signals in either spectrum are marked with grey boxes in the 21 bp siRNA sequence.

HNN-COSY experiments were recorded on the free and bound RNA to further characterize Loqs' influence on the duplex RNA. The HNN-COSY transfers magnetization via H-bonds from an H-bond donor (G, U) to an H-bond acceptor (A, C). Therefore, two signals with opposite phases will be visible at the imino shift (Figure 4.10 A) corresponding to the donor and acceptor. Since magnetization is transferred via the H-bond only stable double-stranded RNA will give signals while unstructured residues remain invisible. The HNN-COSY is

thus a tool to visualize H-bonds and by using the intensity ratio of two signals from a base pair the size of the H-bond can be calculated according to the following formula[216]:

$$\left|{}^{h2}J_{NN}\right| = \frac{\operatorname{atan}\sqrt{\frac{-I_{Na}}{I_{Nd}}}}{2\pi\Delta} \qquad (16)$$

Here I corresponds to the intensity of either the H-bond acceptor or donor and 2Δ is the COSY transfer time.

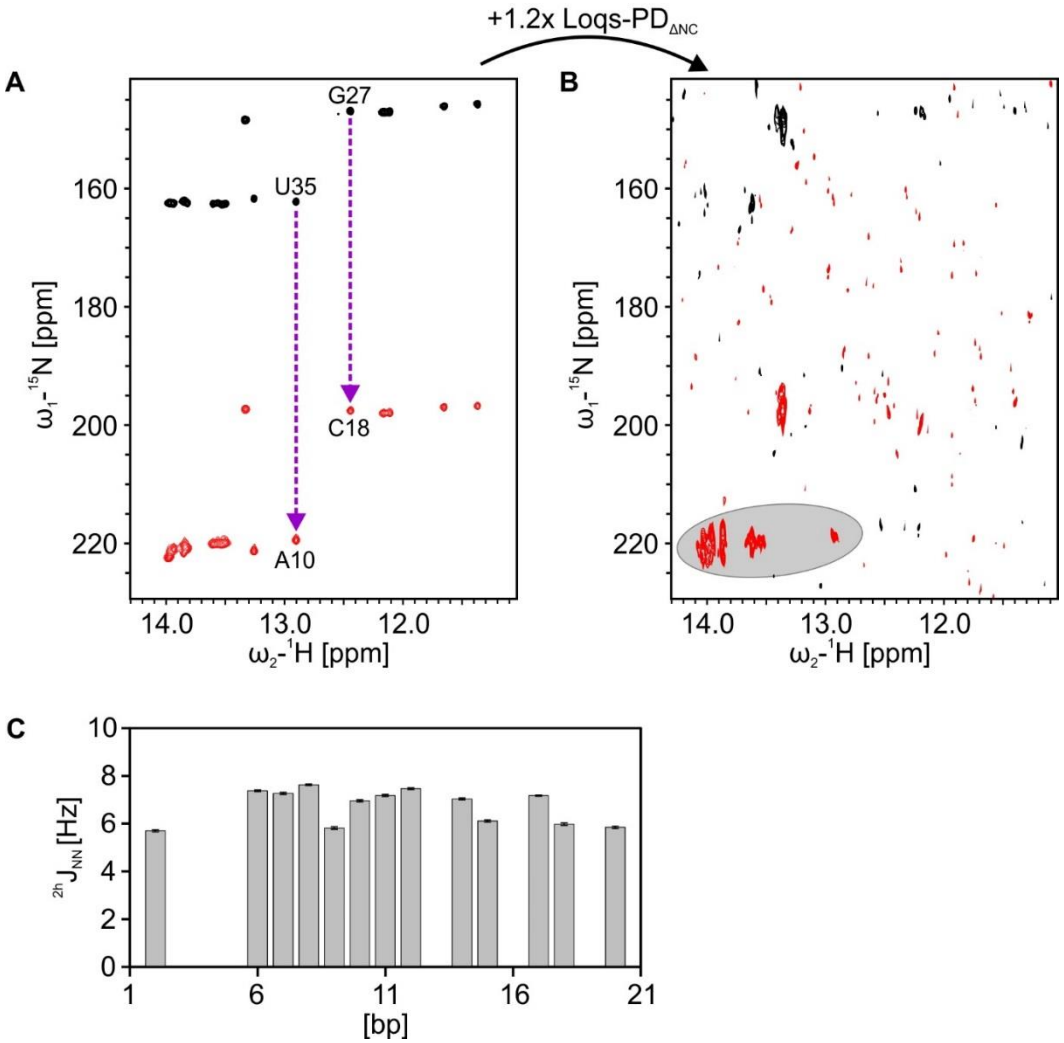



Figure 4.10: HNN-COSY quantifies hydrogen bonds within the RNA duplex. **A)** HNN-COSY of the free 21 bp siRNA. H-bond donor signals are shown in black, acceptor signals in red. Each base pair is represented by two signals at the same imino frequency. As example the pairs G27-C18 and U35-A10 are highlighted. The magnetization is transferred from the donor to the acceptor as indicated by the dotted purple arrows. **B)** In presence of Loqs-PD$_{\Delta NC}$ all signals were severly line broadened. Several acceptor signals gained in intensity compared to the corresponding donor signals; especially adenosine signals increased (grey sphere). **C)** The H-bonds for separated base pairs were calculated and plotted against the base pair number. Since overlapping signals cannot be properly analyzed no H-bond size was calculated for the respective base pairs.

For all base pairs of the 21 bp siRNA two signals were observed in the free HNN-COSY (Figure 4.10 A) indicating that the entire RNA was a stable duplex at 25°C as expected. This is in agreement with CD measurements of the RNA which yielded an approximate melting temperature of 61°C (Supplementary Figure 6.6 B). Upon addition of Loqs-PD$_{\Delta NC}$ most of these signals got line broadened and not all could be detected (Figure 4.10 B). Besides that the line shape and low signal-to-noise ratio allowed only a partial assignment. According to formula 16 the size of the H-bonds for the free RNA was calculated (Figure 4.10 C) and all H-bonds are between 5.7 – 7.6 Hz which reflects a reasonable size. Though the spectrum quality for the complex was not good it allowed a qualitative estimation: Many acceptor signals (red) increased in intensity compared to the donor signals (black) which proved that the magnetization transfer was improved in presence of the protein. This suggests that the duplex RNA is stabilized by Loqs.

### 4.1.8 Intermolecular PREs confirm dynamics and suggest a weak asymmetric binding

Paramagnetic relaxation enhancement (PRE) can provide medium-range distance information (up to approx. 20 Å) and are a versatile tool to analyze protein-RNA complexes and determine binding sites[217] (see 1.4.9). To measure intermolecular PREs within the Loqs-RNA complex the 21 bp siRNA was chemically labeled with a TEMPO spin label. Synthesis and labeling of the RNA was done by Christoph Wunderlich and Christoph Kreutz. Integrity of the RNA and spin label was confirmed by mass spectrometry by the Kreutz group (data not shown) and with 1D NMR (Supplementary Figure 6.2).

Two different samples were prepared with the spin label attached to the 5’ end of either the passenger or guide strand (Figure 4.11 A) and $^{1}$H,$^{15}$N-HSQCs of the complex with Loqs-PD$_{\Delta NC}$ were recorded for the paramagnetic sample as well as the diamagnetic one after addition of ascorbic acid which reduces the spin label (Figure 4.11 B). In both experiments weak line broadening in presence of the spin label could be detected for selected residues. Overall only a few residues were affected but the PRE effect was slightly stronger when the passenger strand was spin labeled (Figure 4.11 C, D). The intensity ratio was calculated according to formula 10 and plotted against the residue number for both experiments (Figure 4.11 E). As already determined visually using the spectra only a few and weak PRE effects

could be observed and for passenger strand labeling the effect was slightly stronger. Since the PRE effect was in total very weak a dynamic binding mode seems plausible: On average the dsRBDs usually were not in close proximity to the spin label and likely diffused on the RNA. However, it seems that the domains slightly preferred the more stable terminus of the duplex over the less stable one. DsRBDs have a preference for perfect matching and thus stable duplex RNAs. Hence this preference could be an intrinsic feature. The effect observed was weak though probably due to the small thermodynamic difference of both ends which differ only in one base pair (Figure 4.1 A).

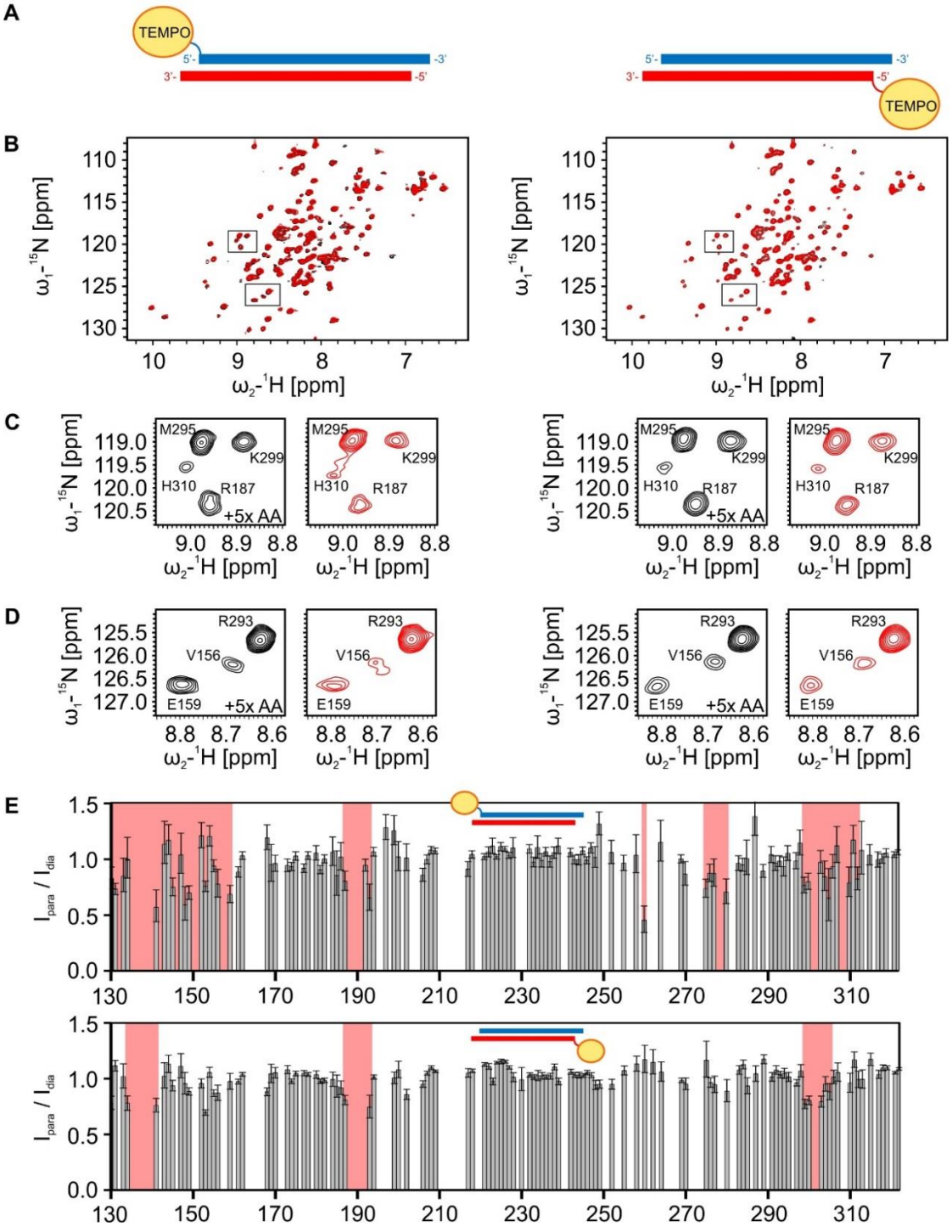



Figure 4.11: Intermolecular PRE measurements of Loqs-PD$_{\Delta NC}$ complexed with spin labeled 21 bp siRNA. **A)** Schematic representation of the two spin labeled RNA used in this experiment. On the left the TEMPO spin label is covalently attached to the passenger strand's 5' end, on the right to the guide strand. **B)** $^{1}$H,$^{15}$N-HSQC spectra of the diamagnetic (black) and paramagnetic complex Loqs-PD$_{\Delta NC}$ with either the passenger strand (left) or guide strand (right) spin labeled. The paramagnetic sample was reduced with 5-fold ascorbic acid. The small boxes show the zoom in shown in **C)** and **D)**. The zoom in show a weak reduction in intensity for selected residues in the paramagnetic sample. **E)** The calculated intensity ratios are plotted against the Loqs-PD residue number. As observed in C) and D) weak effects could be detected. Regions that show a PRE effect are highlighted by red boxes.

### 4.1.9 The C-terminus mediates dimerization of Loqs and interaction with Dicer to form an asymmetric complex

In contrast to Loqs-PA and Loqs-PB the two isoforms PC and PD lack the third dsRBD. Instead Loqs-PD has a modified sequence at its C-terminus. To analyze the influence of these additional 37 residues $^{1}H,^{15}N$-HSQCs of Loqs-PD$_{\Delta NC}$ and Loqs-PD$_{\Delta N}$ were recorded and compared (data not shown). The calculated CSPs (Figure 4.12 A) showed no significant shifts. Note that only residues of Loqs-PD$_{\Delta NC}$ are shown since no additional assignment for the C-terminus was made. The C-terminus thus does not interact with the two dsRBDs or the linker.

Upon titrating 21 bp siRNA Loqs-PD$_{\Delta N}$ showed significant CSPs (Figure 4.12 B) and a similar pattern as the tandem domain construct. Since the yield during purification was very low and line broadening within the complex was severe many shifts could not be tracked and are plotted as negative values. Mostly these residues cluster in RNA binding regions.

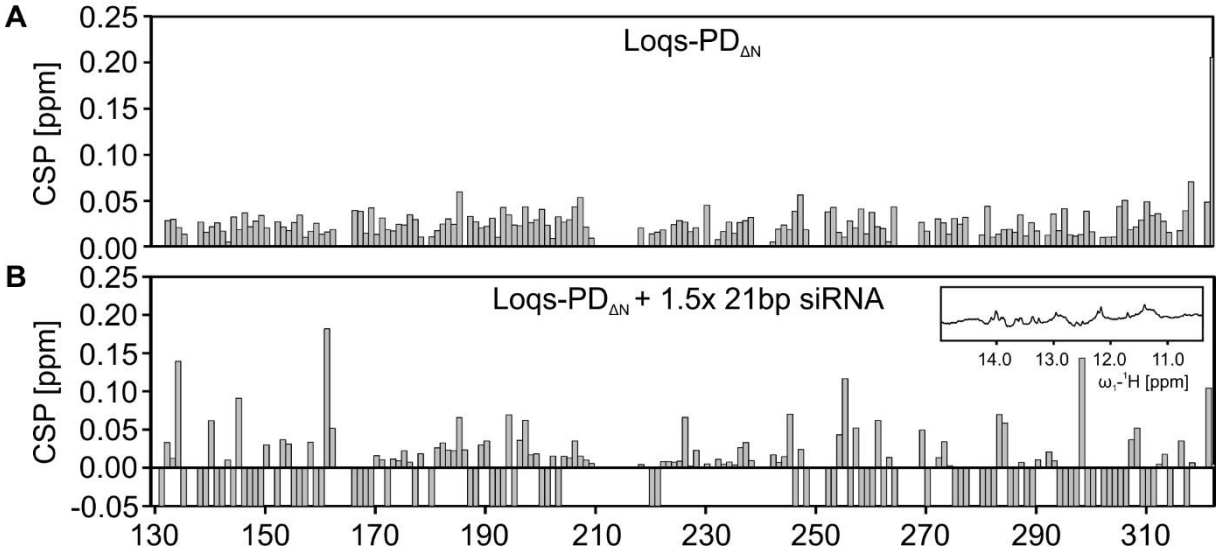



Figure 4.12: Effect of N-terminus on domain structures and RNA binding. **A)** CSPs for Loqs-PD$_{\Delta N}$ compared to Loqs-PD$_{\Delta NC}$. No significant changes were observed suggesting the C-terminus does not interact with the domains. **B)** CSPs for titration of Loqs-PD$_{\Delta N}$ with 1.5-fold 21 bp siRNA. Some significant shifts were observed while most residues experienced severe line broadening and are plotted as negative bars. The small box showes the imino region of the complex sample confirming duplex RNA intregrity.

For both Loqs-PD$_{\Delta NC}$ and Loqs-PD$_{\Delta N}$ SAXS concentration series were measured and the molecular weight was calculated based on the Porod volume (Table 4.1). The molecular weight for Loqs-PD$_{\Delta NC}$ at all concentrations corresponded to a monomer when compared to

the theoretical molecular weight. For Loqs-PD$_{\Delta N}$, however, the molecular weight matched that of a dimer. The C-terminus therefore mediates dimerization of Loqs-PD. This dimerization could possibly explain the line broadening observed in the NMR titration. The stoichiometry of Loqs-PD$_{\Delta N}$ in complex with RNA has not been determined so far.

Table 4.1: SAXS concentration series prove dimerization of Loqs-PD$_{\Delta N}$. For different concentrations of Loqs-PD$_{\Delta NC}$ and Loqs-PD$_{\Delta N}$ the gyration radius, maximum distance, porod volume and the molecular weight have been calculated from SAXS measurements. The theoretical molecular weight is shown on the right.

| Sample | $R_g$ [Å] | $D^{max}$ [Å] | Porod [Å$^3$] | MW [kDa] | theoretical MW [kDa] |
|---|---|---|---|---|---|
| Loqs-PD$_{\Delta NC}$ 1.9 mg/mL | 30.05 ± 0.80 | 104.9 | 29324.9 | 23.8 | 21.1 |
| Loqs-PD$_{\Delta NC}$ 3.6 mg/mL | 32.73 ± 0.86 | 104.6 | 21501.0 | 17.4 | 21.1 |
| Loqs-PD$_{\Delta NC}$ 4.2 mg/mL | 32.18 ± 0.77 | 102.2 | 22139.7 | 17.9 | 21.1 |
| Loqs-PD$_{\Delta NC}$ 4.9 mg/mL | 30.65 ± 1.05 | 104.6 | 23269.3 | 18.8 | 21.1 |
| Loqs-PD$_{\Delta N}$ 1.5 mg/mL | 35.00 ± 1.00 | 167.2 | 52269.1 | 42.3 | 25.3 |
| Loqs-PD$_{\Delta N}$ 3.4 mg/mL | 36.60 ± 0.60 | 169.1 | 64776.1 | 52.5 | 25.3 |
| Loqs-PD$_{\Delta N}$ 4.1 mg/mL | 36.96 ± 1.96 | 177.1 | 63507.2 | 51.4 | 25.3 |
| Loqs-PD$_{\Delta N}$ 4.6 mg/mL | 38.30 ± 1.90 | 177.6 | 65321.1 | 52.9 | 25.3 |

Stephanie Fesser and Klaus Förstemann did further cross-linking experiments with Dicer-2 in presence of either Loqs-PD$_{\Delta NC}$ or Loqs-PD$_{\Delta N}$. Dicer-2 on its own, like Loqs-PD, bound preferentially to the RNA termini and distributed equally to both ends (Figure 4.13 B, C). In presence of excess of Loqs-PD Dicer-2 occupied mainly the less stable end and hardly cross-linked to the stable terminus. This suggests that Loqs-PD and Dicer-2 form an asymmetric complex where probably Loqs-PD binds the stable RNA terminus. An asymmetric complex could be the key step towards strand selection during siRNA maturation. In presence of Loqs-PD$_{\Delta NC}$, however, Dicer-2 bound to both termini equally again. Therefore, the interaction of Loqs-PD and Dicer-2 is mediated via Loqs' C-terminus. Furthermore, this interaction is crucial to fix the ternary complex.

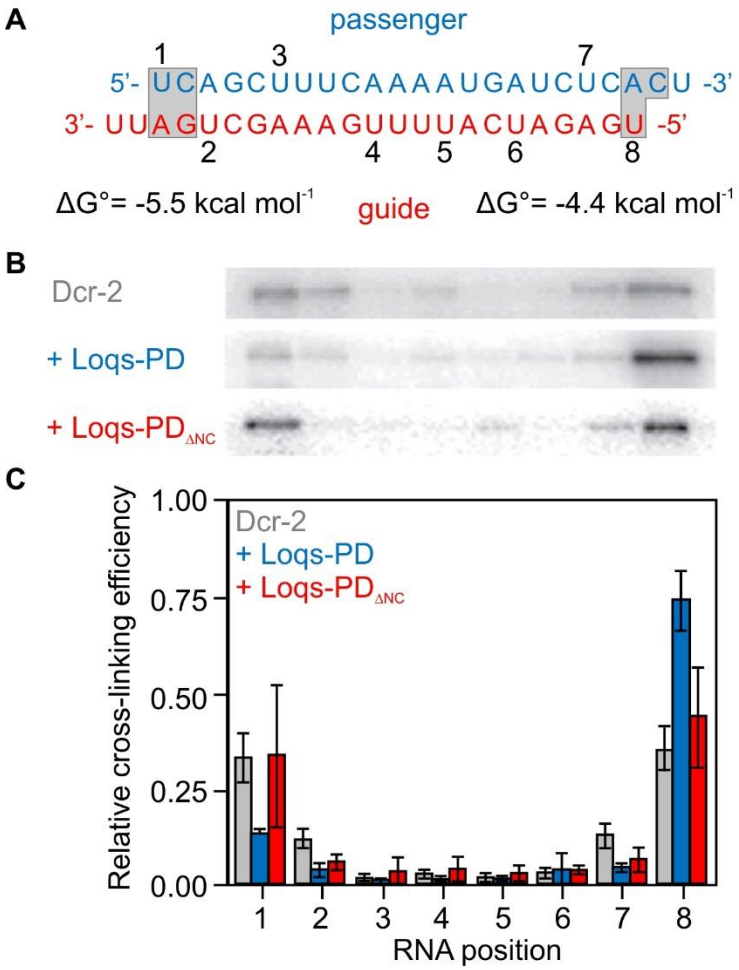



Figure 4.13: Crosslinking of Dicer-2-Loqs-PD complex reveals asymmetric complex formation. **A)** RNA sequence of *bantam* derived siRNA. Numbers 1-8 indicate cross-link positions. Grey boxes highlight residues different from the RNA described in 4.1.2 used for all NMR experiments. **B)** Cross-linking experiments for Dicer-2 alone and in presence of either Loqs-PD or Loqs-PD$_{\Delta NC}$ in excess. **C)** The relative cross-linking efficiency for all experiments shown in B). In presence of full length Loqs-PD (blue) Dicer-2 cross-links preferentially to the less stable terminus. Truncated Loqs-PD (red), however, does not lead to an asymmetric distribution of Dicer-2. Image adapted from Stephanie Fesser and Klaus Förstemann.

## 4.2 Dynamic dsRNA binding by Loqs-PD

### 4.2.1 Chemical exchange as a tool to study dynamics

An NMR signal reflects a time and population average of molecules observed. When a biomolecule samples different conformations, binds a ligand or performs a chemical reaction the involved residues experience different chemical environments. The process of switching between these different states is called chemical exchange. Depending on the velocity of exchange the signals can be line broadened or the different states are detected in separate signals. In slow exchange, when the exchange between two states is slower that the frequency difference of both respective signals ($k_{ex} << \Delta v$) two distinct peaks are observed. For fast exchange in contrast a single averaged signal is detected when the exchange is faster than the frequency difference ($k_{ex} >> \Delta v$).

The same is true for dynamic binding in complexes: When a protein moves along a ligand it contacts different binding registers which leads to chemical exchange. This phenomenon was used by Loth et al[218]. They analyzed sliding of the lac-repressor on dsDNA: Increasing length of the DNA duplex lead to increased line broadening for residues in the protein-DNA interface (a direct consequence of chemical exchange because longer duplexes provide more different binding registers). Sequence-dependent chemical shifts allowed the estimation of a 1D diffusion constant. Different nucleic acids thus could potentially cause different line broadening which can point at dynamic motions of a protein on the nucleic acid.

### 4.2.2 Transcription and analysis of palindromic RNAs

The approach described in 4.2.1 was established for the Loqs-PD-siRNA complex. To analyze the effect of RNA length on line broadening in Loqs spectra different RNAs were artificially designed, transcribed in vitro, purified and analyzed by NMR. An initial set of three RNAs was based on an 8-nt building block comprising a 4-nt palindromic sequence. By adding multiple of these building blocks 16, 24 and 32 bp dsRNAs were designed (Figure 4.14 A). The RNA was palindromic and consisted of repetitive elements to allow self-annealing and thus reduce the workload. Furthermore, even though most literature states RNA binding by

dsRBDs to be sequence-independent, this design allowed an analysis solely based on length since the sequence is always the same.

To gain further data two new RNAs were transcribed comprising the building block mentioned above and additional palindromic sequences of 4-nt to make up 20 and 28 bp dsRNAs (Figure 4.14 A). Transcription of the RNAs was straightforward after $MgCl_2$ optimization. Unfortunately, the palindromic sequence led to lower yields with increasing sequence length which was most likely due to self-annealing of the DNA template or the nascent RNA during transcription. Several attempts by adding DMSO or varying the temperature did not improve transcription efficiency. Therefore, the low yield had to be compensated by larger transcription volumes and a careful purification afterwards to minimize loss. After purification (Figure 4.14 B) via HPLC the RNAs were analyzed by NMR spectroscopy. Both 1D and imino NOESYs were recorded (Figure 4.14, C, D; Supplementary Figure 6.3). All RNAs could be assigned and for RNAs 16, 24 and 32 as well as 20 and 28 all spectra overlap. In the 1D spectra it is obvious that several signals gain in intensity while others remain constant. Constant signals correspond to the termini and thus the increase in intensity reflects the additional building blocks that contribute to the signals. This confirms the structural integrity of the RNAs.

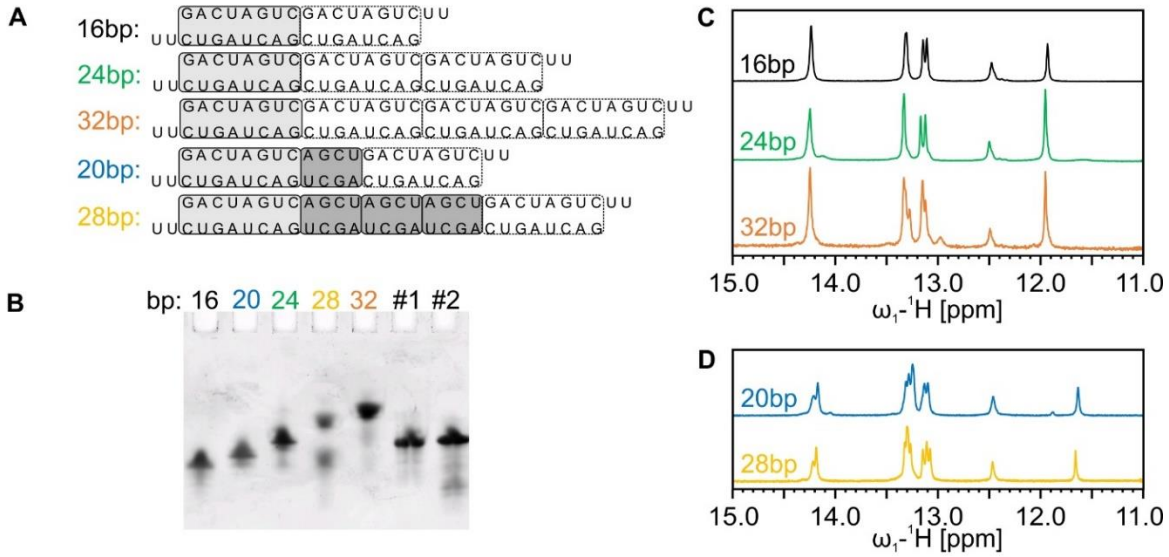



Figure 4.14: Palindromic RNAs used for line broadening analysis. **A)** Sequences of the different palindromic RNAs used. The boxes indicate the repetitive and palindromic sequence used to build up the RNAs. Note that RNA 20 and 28 comprise additional shorter palindromic sequences. **B)** Denaturing polyacrylamide gel of palindromic RNAs and RNA 1 and 2 of the 21 bp siRNA as reference. **C, D)** 1D imino traces of RNA 16, 24, 32 and 20, 28 respectively. The palindromic sequence leads to increasing signals with increasing RNA lengths.

### 4.2.3 Single and tandem dsRBDs slide along dsRNA

To analyze potential sliding of Loqs-PD the two individual dsRBDs and the tandem dsRBD contruct were analyzed with the above mentioned RNAs. Both individual dsRBDs were titrated using RNA 16, 20 and 24 and both CSPs and line broadening were analyzed (Figure 4.15). DsRBD1 and 2 showed very similar shifts for all three titrations, indicating a similar binding mode (Figure 4.15 A, B). While the CSPs for the titrations with RNA 20 and 24 were basically identical the CSPs for 16 bp RNA titration were a little bit smaller (Figure 4.15 C, E). This suggests that either the protein was not saturated or the RNA was too short for full affinity binding. Since RNA was added in 5-fold excess and no further shifts were observed, the protein was saturated but the RNA was too short to provide an optimal substrate. This is in agreement with the low affinity determined by Stephanie Fesser and Klaus Förstemann for short RNA duplexes. Nonetheless, the CSPs pattern was similar for all titrations.

The line broadening was calculated according to equation 13. For both dsRBD1 and 2 line broadening upon complex formation increased with increasing length of the RNA (Figure 4.15 D, F). Here three different classes of residues need to be distinguished: First, for many residues line broadening was linearly depending on RNA length (e.g. 290-300). These residues overall showed moderate line broadening by a factor of 1-4 and low CSPs in the corresponding plot. The low CSPs values indicate these residues were not involved in RNA binding or not even close to the RNA binding interface. Line broadening therefore can only be caused by the increased molecular weight of the complex which lead to a decreased signal-to-noise ratio in the spectra. Second, other peaks show higher CSP values but similar line broadening as the first group of residues (e.g. 310-315). These residues were either not directly involved in RNA binding since CSPs can be secondary effects or they contacted the RNA through their side chains which caused higher CSPs but moderate line broadening in the detected amide backbone. Third, in both domains some residues showed both increased CSPs and strong line broadening (201, 306). These are residues that are directly involved in RNA binding and experience line broadening due to chemical exchange. The high affinity (200 nM) of the individual dsRBDs and the line broadening of most RNA signals (see 4.1.7) suggest that this exchange is due to sliding and not frequent dissociation and subsequent reassociation of the protein.

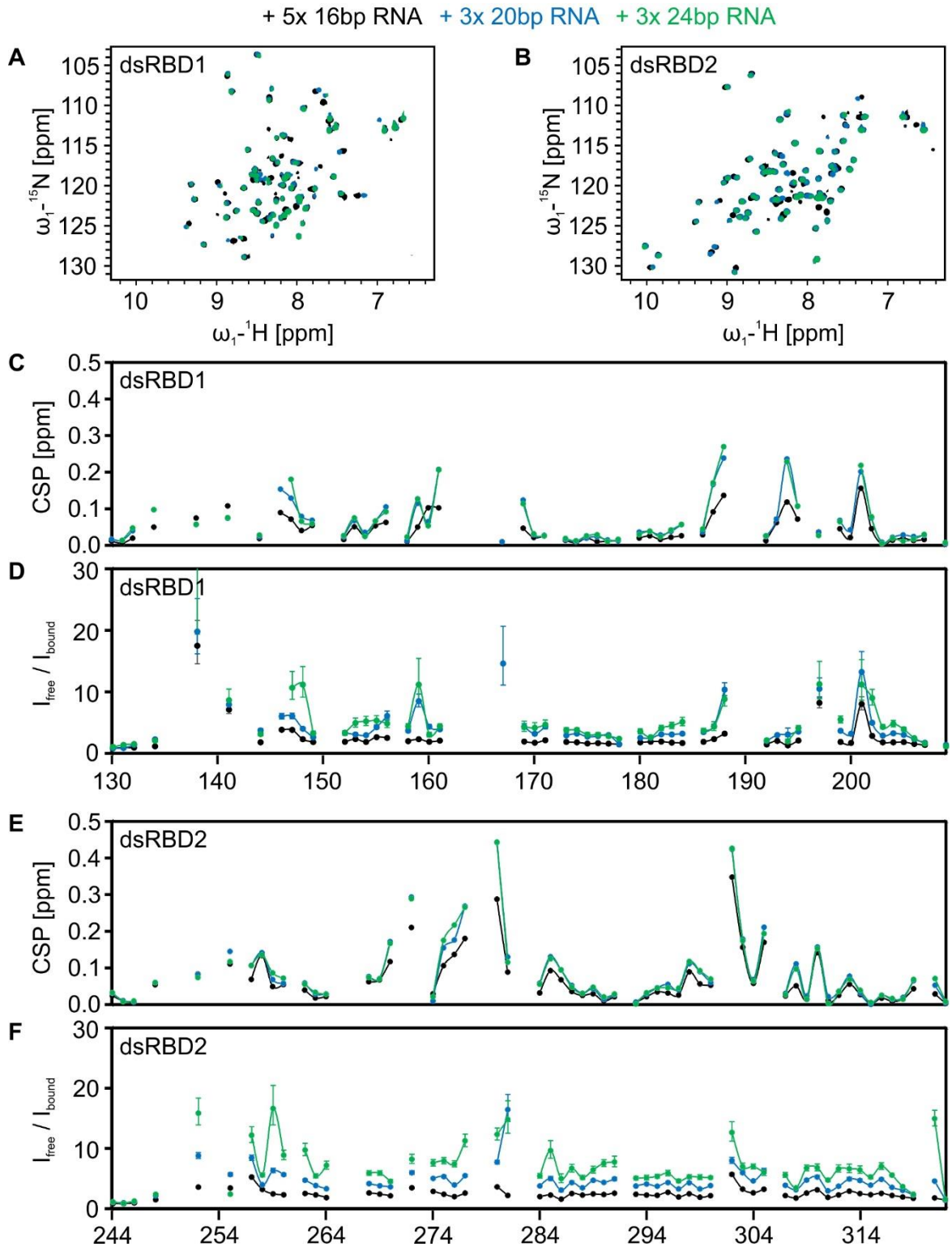



Figure 4.15: Line broadening analysis of single dsRBD RNA titrations. **A, B)** $^{1}H,^{15}N$-HSQCs of dsRBD1 and 2 respectively in presence of 5-fold 16 bp RNA (black), 3-fold 20 bp RNA (blue) and 3-fold 24 bp RNA (green). **C, E)** Calculated CSPs for titrations shown in A) and B) corresponding to dsRBD1 and dsRBD2 binding respectively. **D)** and **F)** show the corresponding intensity ratios of the free over bound form plotted against the residue number.

Using the RNAs 16, 20, 24, 28 and 32 the same experiments were done with the tandem dsRBDs of Loqs (Figure 4.16 A). Due to the size and increased overlap the overall spectral quality got worse, however, a similar effect as for the individual dsRBDs was observed. The CSPs were comparable for all titrations (Figure 4.16 B) but line broadening increased with longer RNAs (Figure 4.16 C). This strongly suggests that both domains do slide in the tandem construct as well.

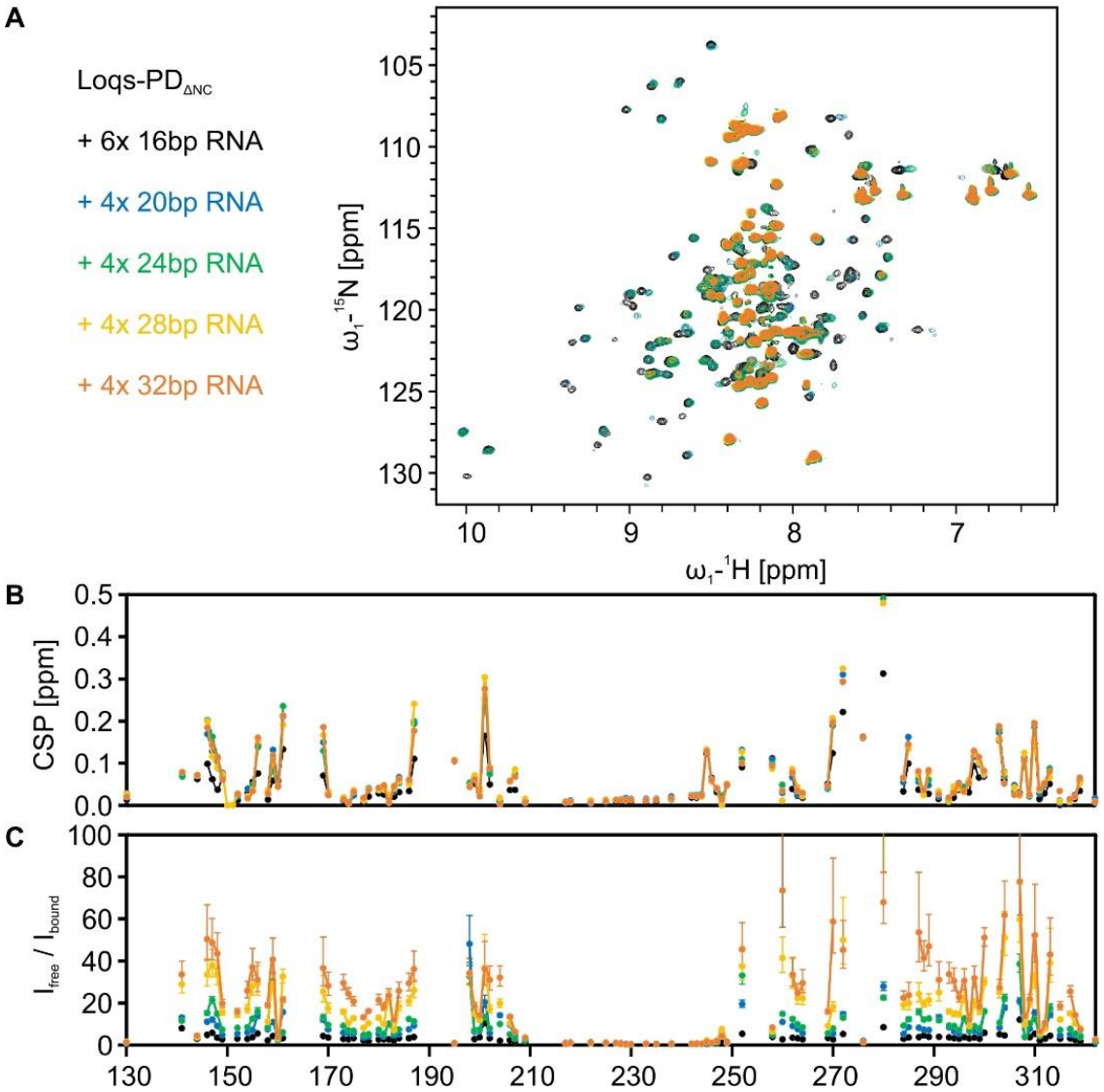



Figure 4.16: Line broadening analysis of tandem dsRBD RNA titrations. **3A)** $^{1}$H,$^{15}$N-HSQCs of Loqs' tandem dsRBDs in presence of 6-fold 16 bp RNA (black), 4-fold 20 bp RNA (blue), 4-fold 24 bp RNA (green), 4-fold 28 bp RNA (yellow) and 4-fold 32 bp RNA (orange). **B)** Calculated CSPs for titrations shown in A). **C)** shows the corresponding intensity ratios of the free over bound form plotted against the residue number.

### 4.2.4 Sliding is independent of RNA sequence composition

Though literature states that RNA binding by dsRBDs in general is sequence-independent, the effect of sequence on line broadening was analyzed. The previous results suggest that Loqs might sense thermodynamic stability of an RNA duplex (see 4.1.7 and 4.1.8) and hence the GC content of or its distribution within an RNA could influence sliding. Therefore, the previously used 21 bp siRNA was varied in sequence to create four new RNAs (Figure 4.17 A). For one the termini were flipped by introducing GCs at the opposite end and even increasing the difference in thermodynamic stability. A GC rich RNA was created by exchanging AU base pairs against GCs and a GC poor RNA by changing GCs to AUs. Another RNA comprised a GC rich patch in the center. After purification all RNAs were analyzed by 1D NMR (Figure 4.17 B) and imino NOESY spectra (Supplementary Figure 6.4) and assigned to confirm the structural integrity of each duplex. Some of these RNAs differ from typical and potent siRNA sequences[219] but were used to analyze the sequence effect on sliding.

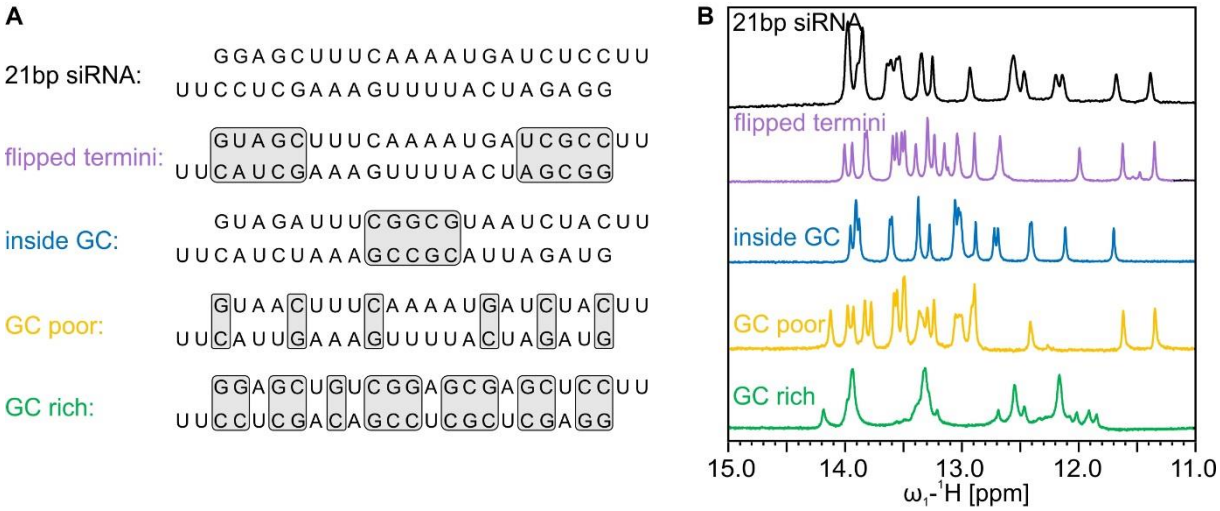



Figure 4.17: Different siRNAs used for line broadening analysis. **4A)** Sequences of the different siRNAs used. Grey boxes highlight changes compared to the 21 bp siRNA. **B)** 1D imino traces of the siRNAs.

As before CSPs were analyzed for all RNA titrations (Figure 4.18 A) and the same pattern was observed for all experiments. The line broadening analysis (Figure 4.18 B) revealed no significant differences for the different substrates used. Sequence composition hence seems to have no or only a marginal effect on sliding and is neglectable compared to the RNA length.

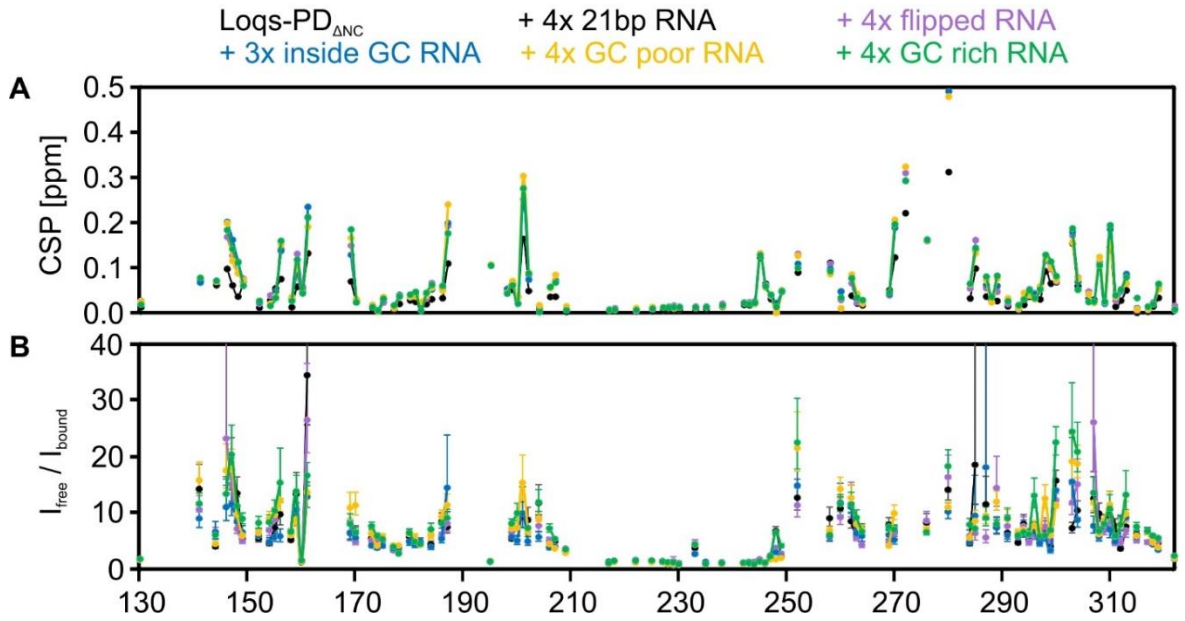



Figure 4.18: Line broadening analysis of Loqs-PD$_{\Delta NC}$ siRNA titrations. **A)** Calculated CSPs for titrations of Loqs' tandem dsRBDs with 4-fold 21 bp RNA (black), 4-fold flipped RNA (violet), 3-fold inside GC RNA (blue), 4-fold GC poor RNA (yellow) and 4-fold GC rich RNA (green). **B)** The corresponding intensity ratios of the free over bound form plotted against the residue number.

### 4.2.5 The linker does not affect sliding

As shown above (4.1.4) the linker assures simultaneous binding of both dsRBDs to the RNA. For other dsRBPs it has been proposed that the linker also contributes to sliding[75]. To test this hypothesis for Loqs-PD, two linker mutants were compared with the wildtype protein (Figure 4.19). The $^{1}H,^{15}N$-HSQC spectra of the complexes showed hardly any differences (Figure 4.19 A). This was confirmed by the similar CSPs (Figure 4.19 B) which were almost identical. A comparison of the line broadening (Figure 4.19 C) showed also no significant effect of the linker length. Thus, sliding is not affected by approx. halving the linker. Note that the two shortest linker mutants were not tested.

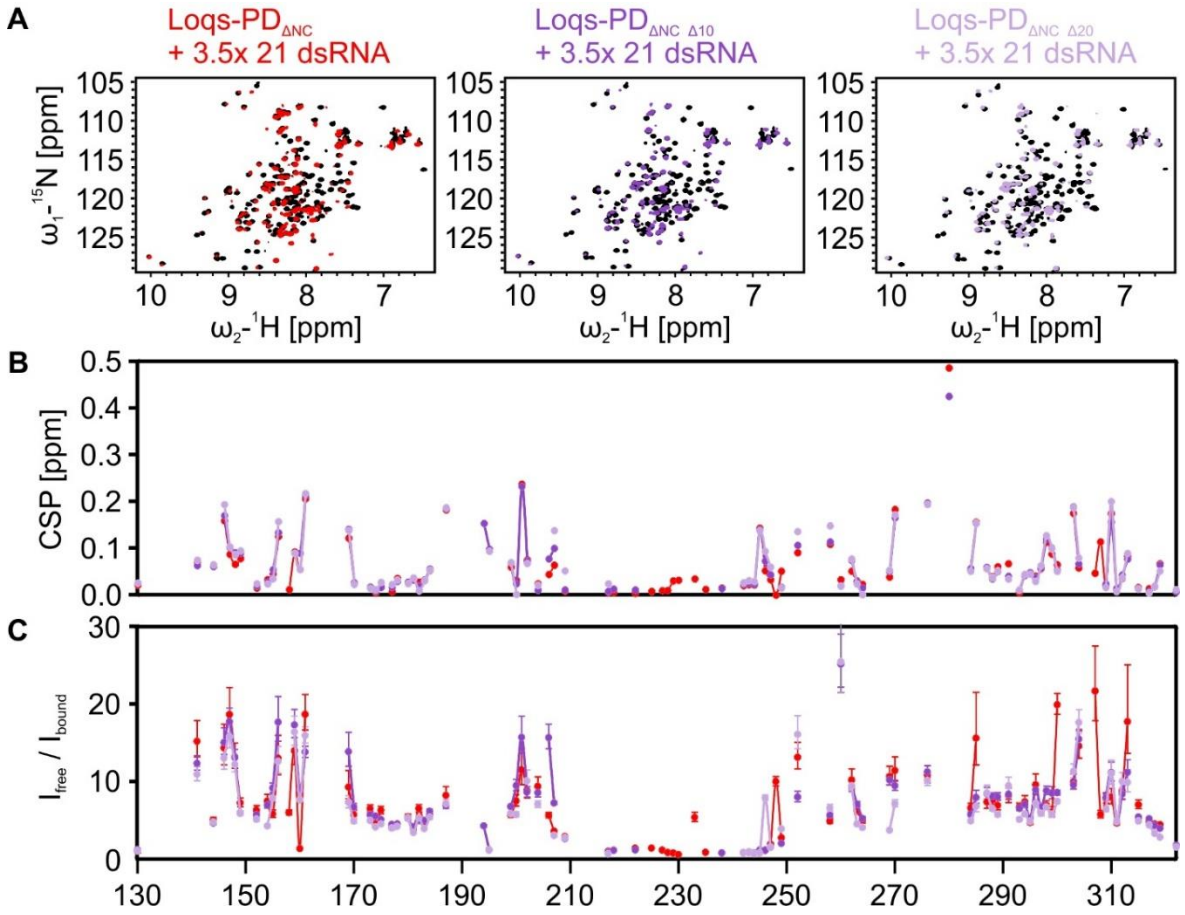



Figure 4.19: Line broadening analysis of Loqs-PD$_{\Delta NC}$ linker mutant RNA titrations. **A)** $^1$H,$^{15}$N-HSQCs of the respective free protein (black) and in presence of 3.5-fold 21 bp RNA shown for the wildtype linker (red), Δ10 linker mutant (violet) and Δ20 linker mutant (light violet). **B)** Calculated CSPs for titrations shown in A. **C)** shows the corresponding intensity ratios of the free over bound form plotted against the residue number.

### 4.2.6 Elevated temperatures enhance sliding

All previous measurements were performed at 25°C which hardly reflects natural conditions. In *Drosophila* (or other insects) the proteins will experience a broader temperature range which could affect their functions while in mammals usually a constant but elevated temperature is maintained. Therefore, different temperatures covering 15, 20, 25, 30 and 37 °C were tested with the standard 21 bp siRNA (Figure 4.20). Since temperature affected the overall spectra (Figure 4.20 A, B) and not all peaks shifted similarly, not all assignments could be transferred between the spectra. Some signals overlapped after temperature changes and were not analyzed for intensities. All experiments were recorded on the same sample and the protein:RNA ratio was exactly the same through all experiments.

Nevertheless, the CSPs were not identical but showed some variations instead, especially in the RNA binding regions (Figure 4.20 C). This might reflect a potential reduction in affinity at elevated temperatures because the CSPs are lower at higher temperatures. The overall line broadening reduced with increasing temperatures (Figure 4.20 D) and for the regions not involved in RNA binding it was again linearly depending on temperature (e.g. residues 170-180).

RNA binding regions in contrast showed a strong reduction in line broadening upon increasing the temperature (Figure 4.20 D). This suggests that sliding increased and the system switched further to the fast exchange regime leading thus to sharper signals and less line broadening.

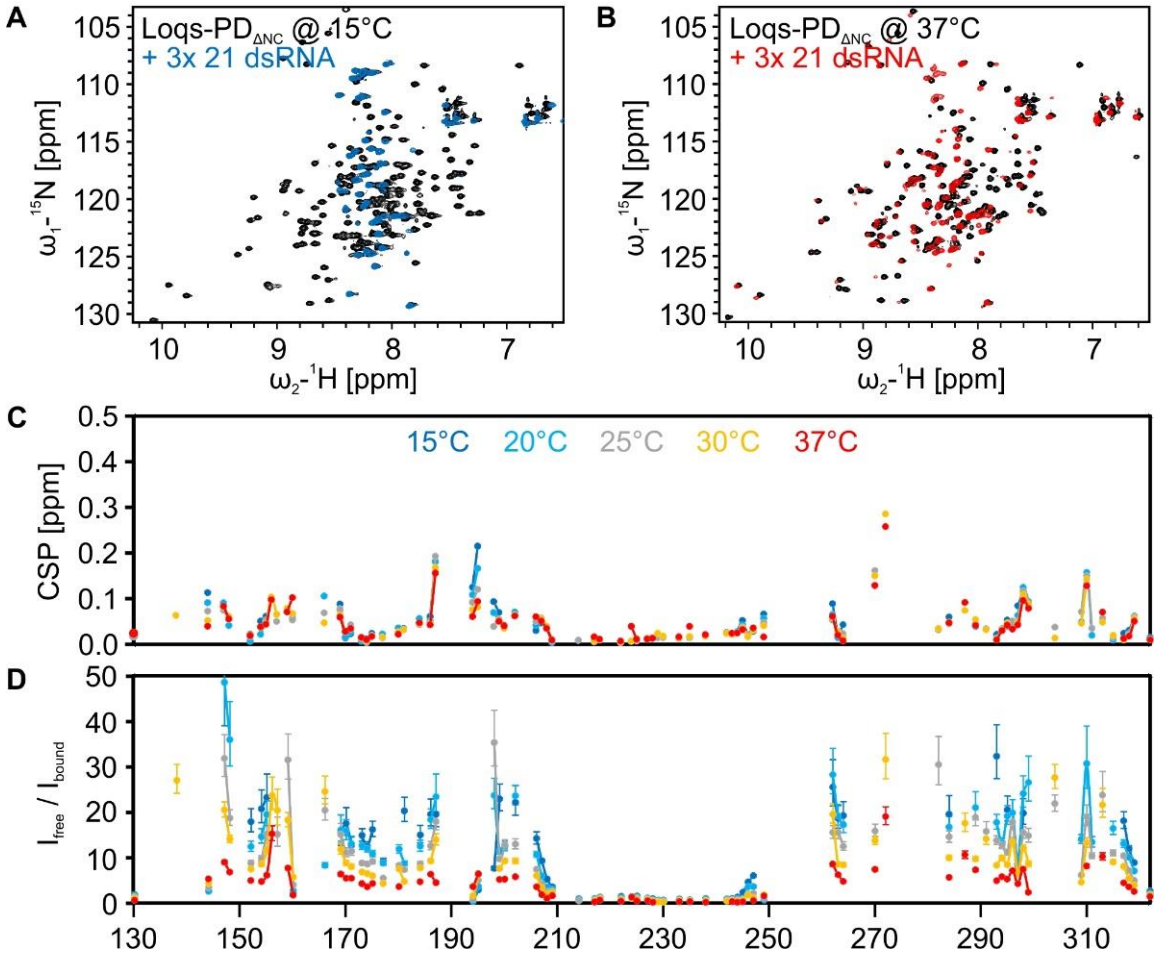



Figure 4.20: Line broadening analysis of Loqs-PD$_{\Delta NC}$ RNA titration at different temperatures. **A)** $^1$H,$^{15}$N-HSQCs of free Loqs' tandem dsRBDs (black) in presence of 3-fold 21 bp RNA at 15 °C (blue) and 37 °C (red). **B)** Calculated CSPs for titrations at 15, 20, 25, 30 and 37 °C. **C)** The corresponding intensity ratios of the free over bound form are plotted against the residue number.

Table 1 shows that the average line broadening of both dsRBDs significantly decreased with raising temperatures. When the inverse of these line broadening factors is plotted against the temperature (Figure 4.21) the y-axis corresponds to the relative exchange speed. It is thus a measure for the relative sliding velocity.

Table 4.2: Average line broadening for dsRBD1 and 2 over selected residues at different temperatues.

| Temperature [°C] | Average line broadening | |
|---|---|---|
| | dsRBD1 (136-205) | dsRBD2 (252-318) |
| 15 | 14.02 | 19.46 |
| 20 | 13.16 | 17.52 |
| 25 | 10.99 | 12.33 |
| 30 | 7.88 | 9.56 |
| 37 | 4.94 | 5.52 |

The two curves show an exponential increase (Figure 4.21) which can be approximated by van't Hoff's rule: For the analyzed temperatures the inverse line broadening increased by a factor of approx. 1.6-2.0 when temperatue increased about 10 °C. Line broadening was solely caused by chemical exchange because the increased molecular weight of the complex was the same at all temperatures. The chemical exchange in an approximation and assuming that no further processes contribute was only caused by sliding of Loqs on the dsRNA. Therefore, the inverse line broadening is a direct measure for the exchange speed and consequently the relative sliding velocity of the protein.

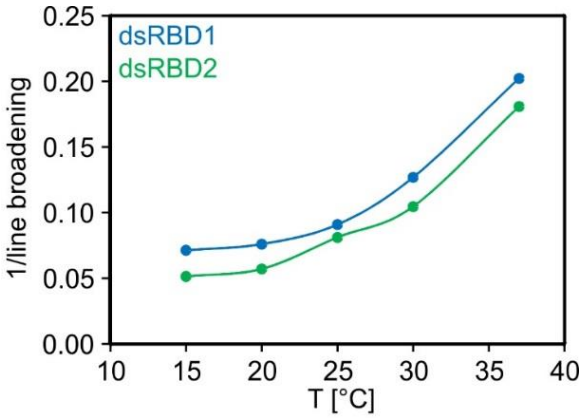



Figure 4.21: Inverse line broadening increases at higher temperatures. The inverse of the average line broadening (table 4.2) plotted against temperature for dsRBD1 (blue) and dsRBD2 (green). Overall sliding increases by a factor of 1.6-2.0 when the temperature increases about 10 °C.

### 4.2.7 Viscosity limits sliding

The viscosity in a cell differs from the one in an aqueous buffer used for NMR. Therefore, different viscosities were tested using the same setup as above to estimate sliding under physiological conditions. According to literature, viscosity in the cytoplasm should be 1-3 times that of water which corresponds to approx. 1-2 cP (Figure 4.22 A). Hence 10 % and 25 % glycerol were added to the samples (Figure 4.22 A). Both samples showed very similar CSPs as the 0 % glycerol sample (Figure 4.22 B, C), indicating that glycerol does not affect the binding interface. However, overall line broadening increased with increasing glycerol concentration (Figure 4.22 D) indicating sliding to be slower. Here the opposite effect of temperature is observed: Higher viscosity shifts the chemical exchange to a slower regime causing more line broadening.

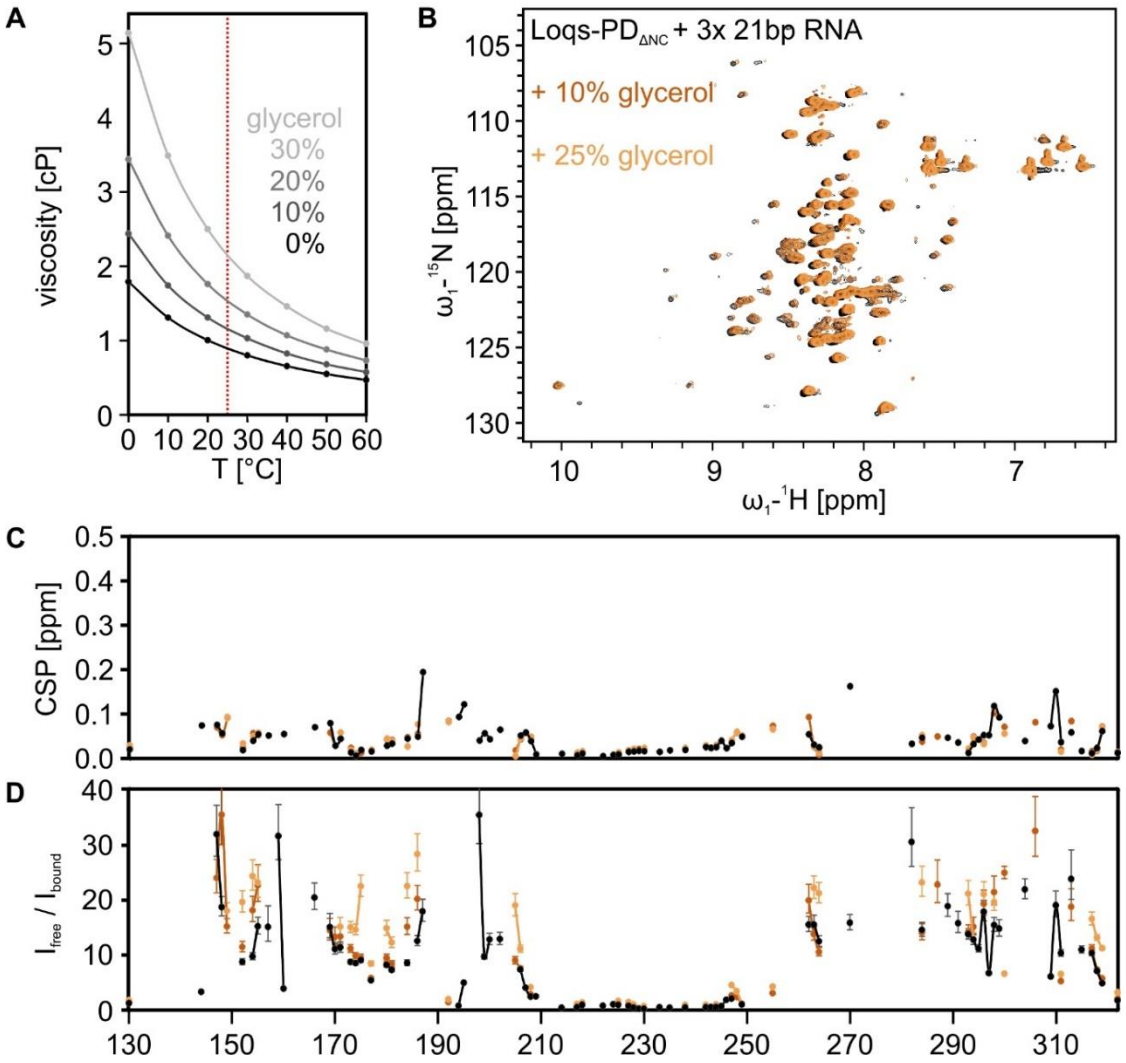



Figure 4.22: Line broadening analysis of Loqs-PD$_{\Delta NC}$ RNA titrations at different glycerol concentrations. **A)** Viscosity of different aqueous glycerol solutions at varying temperatures (based on Segur et Oberstar, 1951[220]). **B)** $^{1}$H,$^{15}$N-HSQCs of Loqs' tandem dsRBDs in presence of 3-fold 21 bp siRNA in standard buffer (black), with 10 % glycerol (brown) and 25 % glycerol (light brown). **C)** Calculated CSPs for titrations shown in B. **D)** shows the corresponding intensity ratios of the free over bound form plotted against the residue number.

By comparing overall line broadening of dsRBD1 for the different experiments (Table 4.3) it is obvious that at 25 % glycerol sliding decreased by a factor of 2. 25 % glycerol corresponds to approx. 1.9 cP – thus the viscosity increased by roughly a factor of 2 compared to pure water (Figure 4.22 A, Table 4.3). Hence, sliding seems to be linearly depending on viscosity. Note that the spectra quality did not allow the analysis of many residues and these results might thus be biased.

Table 4.3: Average line broadening for dsRBD1 and 2 over selected residues at different glycerol concentrations.

| Glycerol [%] | Approx. viscosity [cP] | Average line broadening | |
|---|---|---|---|
| | | dsRBD1 (136-205) | dsRBD2 (252-318) |
| 0 | 0.87 | 9.53 | 13.08 |
| 10 | 1.11 | 13.15 | 13.28 |
| 25 | 1.85 | 18.40 | 17.79 |

## 4.3 Staufen-2 dsRBD1-2 RNA binding

### 4.3.1 Both domains are independent in the free form

To evaluate NMR accessibility of Staufen-2 dsRBD1 and 2 the two individual domains as well as the tandem construct Staufen-$2_{\Delta NC}$ were used to record $^{1}H,^{15}N$-HSQC spectra (Figure 4.23). In all constructs intense and dominant signals at the centre of the spectrum occured, indicating unstructured and flexible regions. At lower contour levels signals corresponding to structured regions showed up. An overlay of the single domain spectra and the tandem domain spectrum showed that most signals nicely overlap. Additional peaks in the tandem domain spectrum are likely to correspond to linker residues which are missing in the individual domain spectra. Since the tandem domain spectrum is thus the sum of the individual spectra it seems likely that the two dsRBDs 1 and 2 of Staufen-2 are independent in the free form and do not significantly interact.

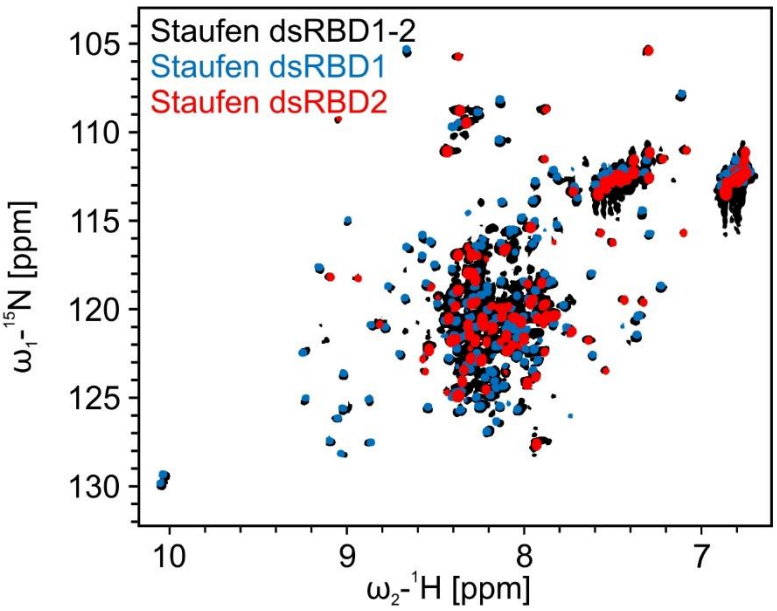



Figure 4.23: Fingerprint spectra of Staufen-2 constructs. $^{1}H,^{15}N$-HSQC spectra of Staufen 2 tandem dsRBD1-2 (black), dsRBD1 (blue) and dsRBD2 (red). Most signals of the individual dsRBDs overlap with those of the tandem construct.

### 4.3.2 Structure homology models indicate conserved fold and extended loop in dsRBD2

The $^{1}H,^{15}N$-HSQC spectra for both individual dsRBDs indicated folded proteins (Figure 4.23). The intense signals in the centre of the spectrum, however, suggest parts of the

proteins to be flexible or unstructured. A backbone assignment based on an HNCACB, HNcaCO and HNCO was started for dsRBD1. The assignment would allow to determine secondary structure elements and map potential RNA binding sites. The $^{1}H,^{15}N$-HSQC spectrum comprised approx. 280 signals which exceeded the number of 95 residues of the construct. For dsRBD2 approx. 192 signals were detected while 146 expected. This indicated possible multiple conformations, oligomerization or aggregation. SAXS experiments of dsRBD1 confirmed a large molecular weight that significantly exceeded a monomeric form (data not shown).

Overall the HNCACB suffered from a poor signal-to-noise ratio for all residues in the structured regions. Only presumably unstructured residues gave intense signals while for the others cross peaks were missing or severely line broadened. Nevertheless, some residues highlighted in green in the following sequence of dsRBD1 could be assigned:

GPMANPKEKTPVCLVNELARFHSIQPQYKLLNESGPAHSKMFSVQLSLGEQTWESEGSSIKKAQQAVAN
KALTESTLPKPVQKPPKSNVNNNPGS

The project was started late during the thesis and therefore the assignment could not be pursued. For sure optimization of the buffer, temperature or deuteration of the samples would improve spectra quality and allow at least a partial assignment.

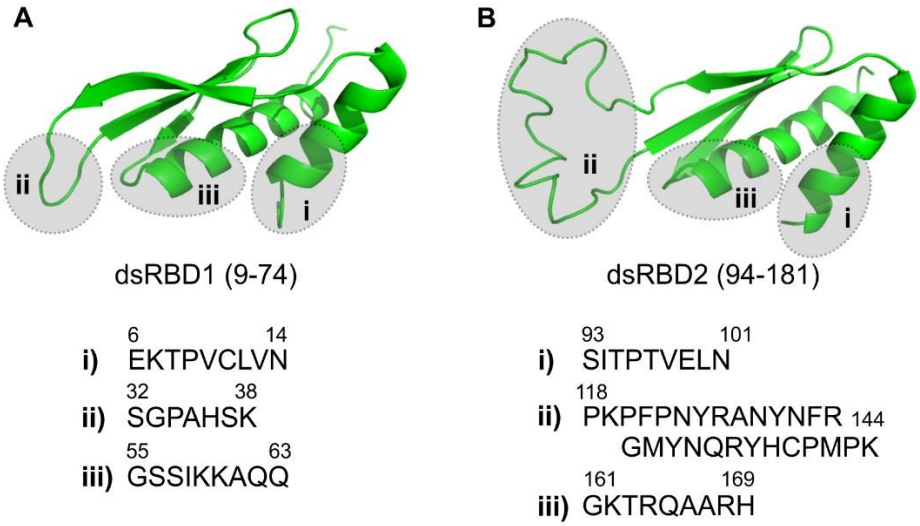



Figure 4.24: Structure homology models of Staufen-2 dsRBD1 and 2. **A)** Sequence-based structure homology model for Staufen-2 dsRBD1 comprised the residues 9-74. **B)** The model of dsRBD2 consisted of residues 94-181 and showed an extended loop between $\beta_1$ and $\beta_2$. For both models the expected conserved binding regions are highlighted by grey spheres. The respective amino acid sequences are shown below.

A structure model was built based on sequence homology (Supplementary Table 6.2) using the webserver swissmodel.expasy.org. After an automatic template search the top 50 hits based on sequence similarity were used for structure modelling. The models for dsRBD1 and 2 with the best GMQE and QMEAN values are shown in Figure 4.24 (Supplementary Figure 6.5). The models comprised residues 9-74 and 96-180 for dsRBD1 and 2 respectively. Both domains showed the canonical αββββα-fold. DsRBD2 has an extraordinary long loop of 21 residues between $\beta_1$ and $\beta_2$ (Figure 4.24 B). For both domains the expected RNA binding interface is highlighted by grey spheres and the corresponding residues are listed below. The sequence similarity thus points at a conserved dsRBD fold and allows a canonical RNA binding mode.

### 4.3.3 The two dsRBDs bind dsRNA dynamically

Both individual dsRBDs were titrated using Rgs4 SRS2 RNA to assess potential RNA binding (Figure 4.25). Prior to titration experiments the structural integrity of the RNA was confirmed by 1D NMR and NOESY spectra (Supplementary Figure 6.6). The spectra of both domains showed severe line broadening upon addition of RNA (Figure 4.25 A, C).

For dsRBD1 (Figure 4.25 A) significant chemical shifts in the spectra were observed while line broadening for dsRBD2 (Figure 4.25 C) was so severe that only few signals remained visible after complex formation. Therefore, both domains do interact with the given stem loop RNA and binding might be dynamic. In both cases saturation was basically achieved at a 1:1 molar ratio which indicates high affinity binding of both dsRBDs. This is in agreement with SPR data from Simone Heber and Dierk Niessing who found the single dsRBDs to bind in the nanomolar range while the tandem domains show clear cooperativity and an increased affinity.

Interestingly, line broadening also occured in the RNA spectra (Figure 4.25 B, D) when protein was in excess which further supports potential dynamics within the complex. A closer look at the imino spectra reveals that the two domains seem to interact differently with the RNA: While for dsRBD1 relatively sharp RNA imino signals were observed at a 1:0.5 protein:RNA ratio (Figure 4.25 B), for dsRBD2 the RNA signals were severly broadened and almost not detectable any more (Figure 4.25 D). Thus, binding of the individual dsRBDs might

occur at a different time scale and involve different dynamics. Note that for both titrations only minor and neglectable shifts in the RNA spectra occurred.

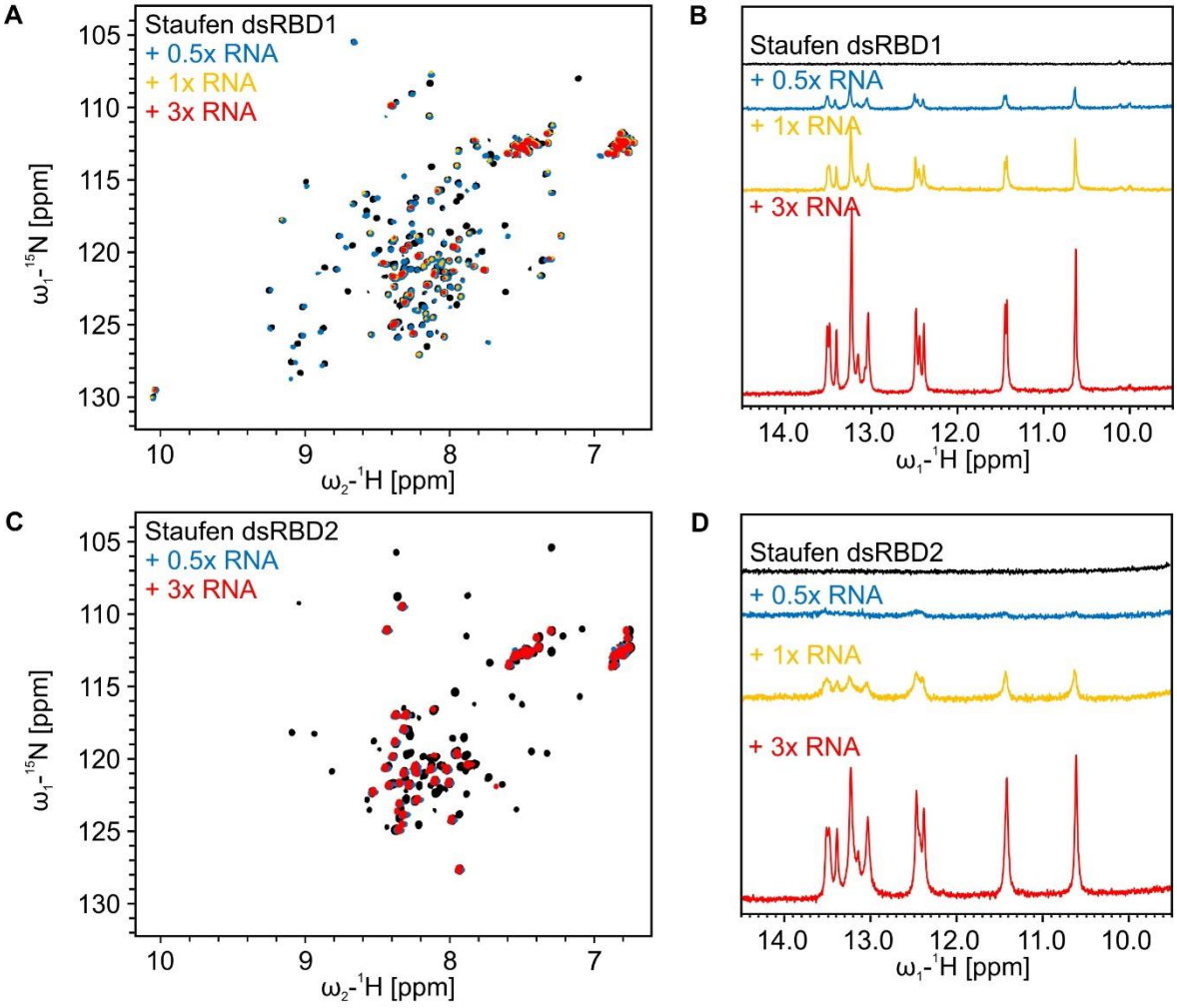



Figure 4.25: RNA titrations of Staufen-2 single dsRBD1 and 2. **A)** $^{1}H,^{15}N$-HSQC spectra of free dsRBD1 (black), in presence of 0.5-fold (blue), 1-fold (yellow) and 3-fold excess of RNA. **B)** The corresponding 1D imino traces of the RNA at the same ratios. **C, D)** The RNA titration for dsRBD2.

So far only a few residues of dsRBD1 could be assigned (see 4.3.2). Some of these residues showed shifts in presence of RNA (Figure 4.26 A) and are thus possibly part of the RNA binding site. Note that Ser32, Gly33 and Ala35 are part of the SGPAHSK sequence shown in Figure 4.24 A in 4.3.2. The shifting residues were mapped onto the structure model (Figure 4.26 B) and all clustered in the predicted loop or α-helix 1. Simone Heber and Dierk Niessing created mutants based on these results and could confirm reduced or even abolished binding

using SPR. Despite the incomplete assignment this suggests that the assignment is correct and that the loop residues of Staufen-2 dsRBD1 do mediate interactions with the RNA.

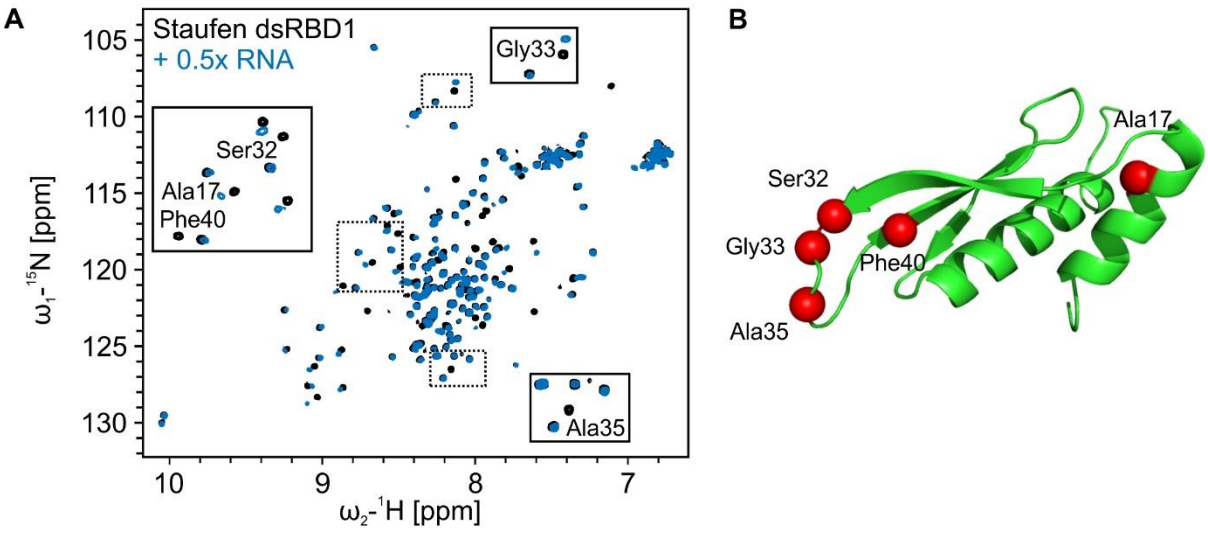



Figure 4.26: Binding site 2 in Staufen-2 dsRBD1. **A)** $^{1}H,^{15}N$-HSQC of free Staufen-2 dsRBD1 (black) and in presence of 0.5-fold stem loop RNA (blue). The small boxes indicate zoom in and highlight residues shifting in presence of RNA. **B)** Residues that could be assigned and showed shifts upon RNA addition are highlighted as red spheres in the structure homology model of dsRBD1.

### 4.3.4 Simultaenous binding of both dsRBDs is independent and dynamic

RNA binding might be different in presence of the other domain and therefore the tandem domain protein was also used for an RNA titration. Similarly to the individual domains the signals got line broadened in presence of RNA and chemical shifts were observable (Figure 4.27 A). By comparison with the individual domain titrations the same residues were affected which indicates that both domains bind RNA in the tandem construct and the binding interface of each individual domain is not affected by the presence of the other domain. The RNA imino signals (Figure 4.27 B) got line broadened as for the dsRBD2 titration. This suggests that also in the tandem domain protein considerable dynamics occur, be it sliding (as proposed for Loqs in 4.2) or high on and off rates.

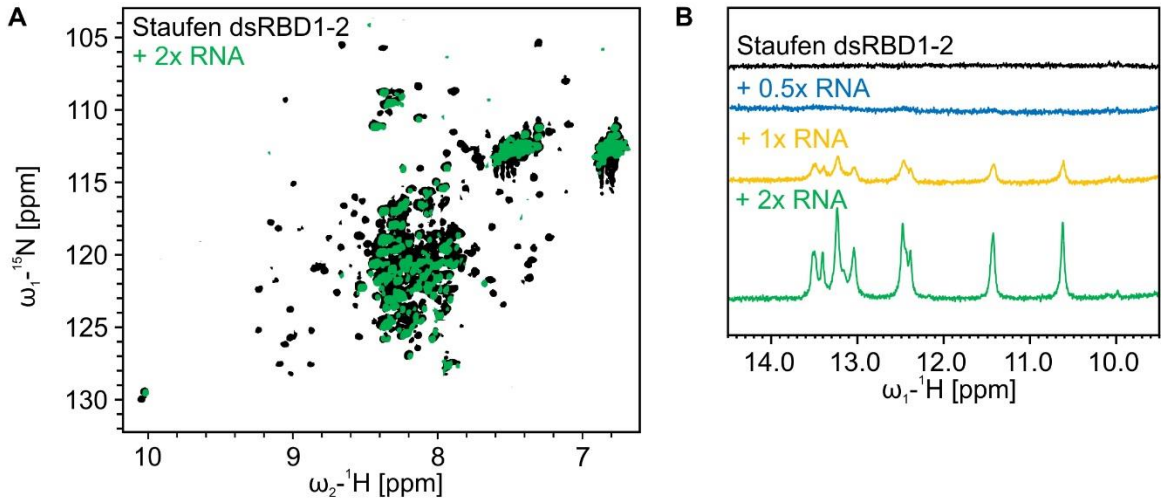



Figure 4.27: Staufen-2$_{\Delta NC}$ RNA titration. **A)** $^{1}$H,$^{15}$N-HSQC of the free Staufen dsRBD1-2 construct (black) and in presence of 2-fold RNA (green). **B)** The RNA imino traces for that titration at different protein:RNA ratios.

## 4.4 Mtr4 KOW ligand binding

### 4.4.1 Backbone assignment of KOW

Due to the excellent quality of the $^{1}H,^{15}N$-HSQC only an HNCACB was used for the initial assignment of the Mtr4 KOW domain. Approximately 50% of the protein could be assigned but the additional magnetization transfer via the $C_\beta$ led to a significant loss of detectable signals. Thus, HNCA and HNcoCA spectra were recorded which indeed filled the assignment gaps by providing information on the previously missing residues and confirmed already assigned residues (Figure 4.28). Remaining gaps are caused by prolines or flexible and overlapping regions. In addition, HNCO and HNcaCO experiments were recorded to confirm the assignment (data not shown). In total approx. 97% of the residues (excluding the artificial N-terminal sequence left after protease cleavage) could be assigned in the HNCACB.

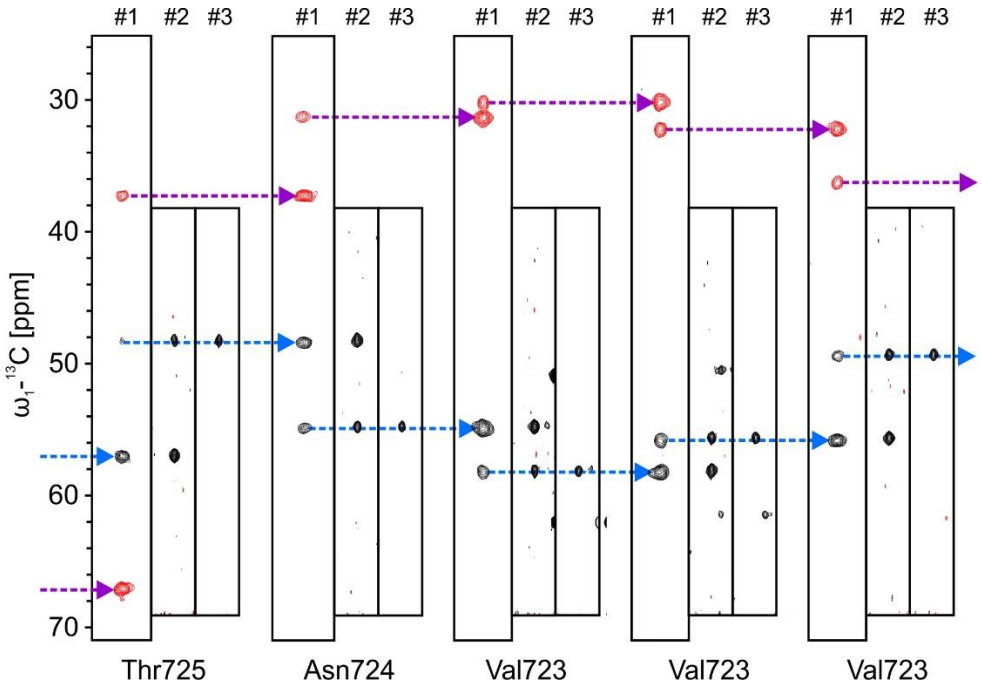



Figure 4.28: Example walk for KOW backbone assignment for the sequence VVVNT. Note that the residue order is opposite (C to N) to the natural one due to the assignment procedure. Black signals show $C_\alpha$, red signals $C_\beta$. Shown are three spectra strips per residue, #1: HNCACB, #2: HNCA, #3:HNcoCA. Blue arrows indicate connections via $C_\alpha$, purple arrows via $C_\beta$.

### 4.4.2 Solution secondary structure elements confirm crystal structure

The calculation of secondary chemical shifts is a versatile and reliable tool to analyze secondary structure elements in solution. It can therefore be used to validate existing crystal structures as in this case. Based on the previously mentioned equation 8 secondary chemical shifts were calculated and plotted against the residue number. The analysis indicates an αβαββαβαβα fold which agrees with the secondary structure elements found in the published crystal structure (Figure 4.29 A; PDB: 2XGJ).

Heteronuclear NOE data provide information on local flexibility. The data recorded for the free protein proved most of the protein to be rigid (Figure 4.29 B). Only the C-terminus (residues 812-818) was highly flexible while the loop (residues 704-711) connecting $\beta_2$ and $\beta_3$ showed moderate flexibility.

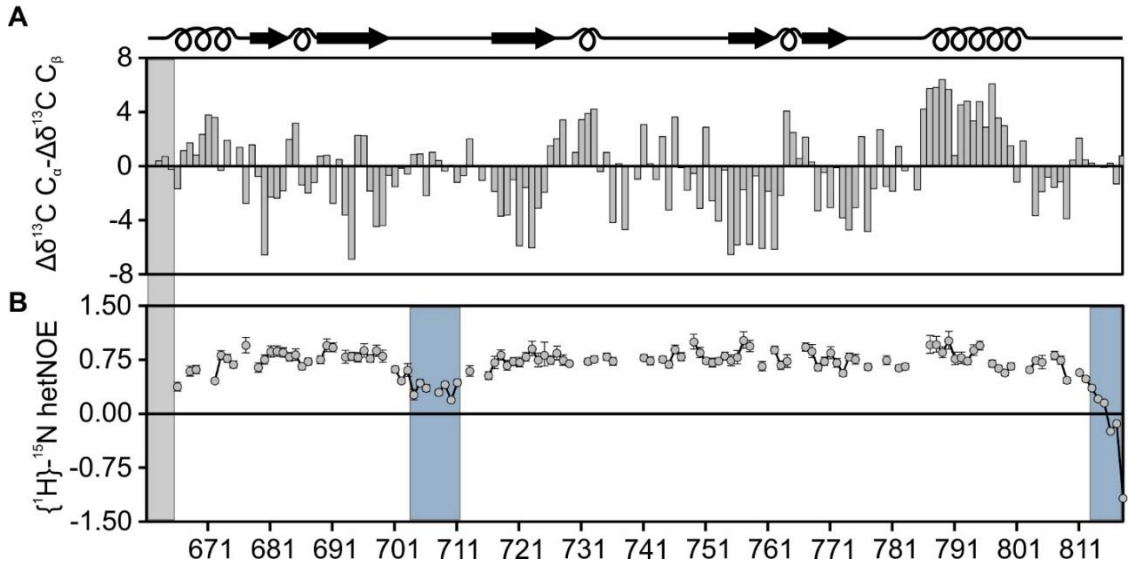



Figure 4.29: Secondary structure elements and flexibility of KOW. **A)** Secondary chemical shifts of Mtr4 KOW plotted against residue number. Positive values indicate α-helical structures, negative values β-strands. Above the plot the secondary structure elements derived from the crystal structure are depicted as cartoons and aligned to the sequence. **B)** Heteronuclear NOE data for the free KOW domain indicate flexible regions in the loop and C-terminus (blue boxes). The grey box indicates the artificial sequence GAASM left after cleavage.

### 4.4.3 Structured RNAs and Nop53 bind to distinct interfaces

As mentioned before the KOW domain binds tRNAs but the RNA elements required for recognition are unknown so far. The previously described 21 bp siRNA was used to titrate a dsRNA to Mtr4 KOW. Significant shifts occurred upon RNA addition but binding was weak

and saturation was not achieved even in 10-fold excess of RNA (Supplementary Figure 6.7 A). Chemical shift perturbations were calculated (Figure 4.30 A) and revealed two main regions to be involved in RNA binding: The N-terminal region of α-helix$_1$ which connects the KOW domain via the arches with the Mtr4 protein core and the flexible loop, especially residues 712-717. Note that a few additional peaks showed up in presence of RNA which are unassigned. In another titration Phe-tRNA from baker's yeast was used as a ligand (Supplementary Figure 6.7 B). In 7.5-fold excess shifts similarly to those from the dsRNA titration were observed in the spectra resulting in comparable CSPs (Figure 4.30 B). The residues mainly involved are histidine, serine, asparagine, aspartate and glutamate, all good H-bond donors and potential RNA binders. These data prove that the KOW domain uses a specific interface to bind structured RNAs via their double-stranded regions. For all titrations involving structured RNAs 1D traces of the imino region were recorded to prove the presence of dsRNA (Figure 4.30).

Mtr4 KOW binds Nop53$^{211}$, a tumor suppressor homologue that recruits Mtr4 to the ribosome. Sebastian Falk and Elena Conti determined the crystal structure of Mtr4 bound to the Nop53 peptide and using NMR I analyzed peptide binding in solution. Upon titrating the peptide to the KOW domain significant chemical shifts were observed (Supplementary Figure 6.7 C). Note that the protein was not saturated due to the low affinity even though a high excess (20-fold) of peptide was used. Chemical shift perturbations were calculated and plotted against the residue number (Figure 4.30 C). The bar plot indicates residues 772-781 and 797 as the core region that was mostly affected by peptide binding (namely leucine, isoleucine, methionine, asparagine and valine). These findings are in agreement with the crystal structure. The binding interface for Nop53 thus differs from that for dsRNA.

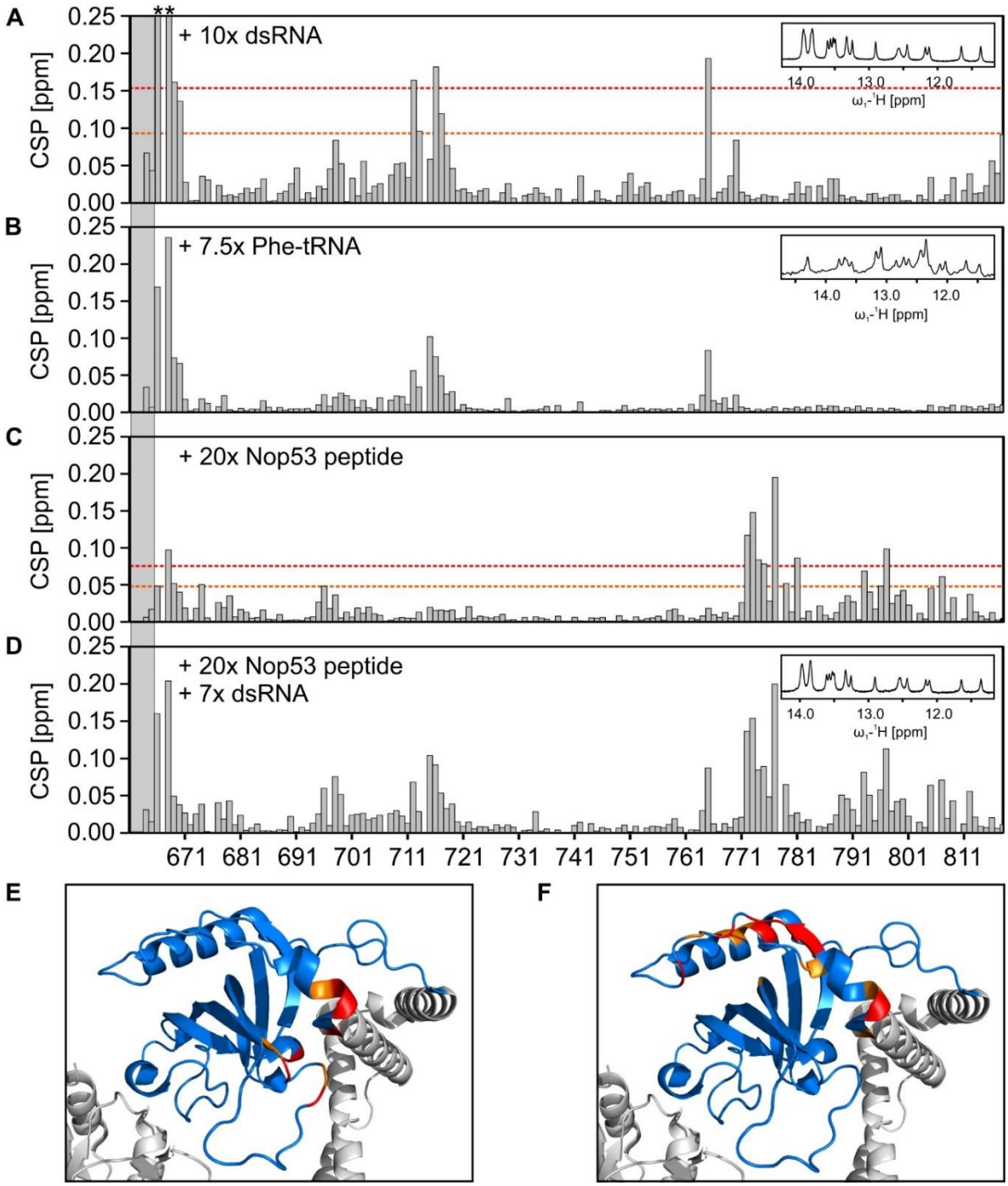



Figure 4.30: KOW titrations with different ligands. Chemical shift perturbations of the titrations with **A)** dsRNA, **B)** Phe-tRNA, **C)** Nop53 peptide and **D)** both Nop53 peptide and dsRNA plotted against the residue number. Residues His666 and Ala668 marked with asterisks in A) show CSP values of 0.38 and 0.51 respectively. For all RNA titrations $^{1}$H imino spectra are shown in small boxes which prove duplex integrity. For A) and C) two lines indicate the average + 1 x standard deviation (orange) and average + 2 x standard deviation (red). **E)** and **F)** show CSPs for dsRNA and Nop53 peptide titration respectively mapped onto the available crystal structure (PDB: 2XGJ). Only residues larger than the average + 1 or 2 standard deviations (see colored lines in A) and C)) are highlighted in orange and red.

The Nop53 bound Mtr4 KOW domain was further titrated with the dsRNA (Supplementary Figure 6.7 D) and again significant shifts could be observed, corresponding to those observed for the dsRNA titration. The CSPs plot referenced to free KOW showed changes in both peptide and dsRNA binding region (Figure 4.30 D) reflecting the sum of both individual titrations. When referring to the KOW-Nop53 complex (Figure 4.31 A) CSPs corresponding to RNA binding were observed. His666, Ala668, Asn669, Ala670 were affected by both peptide and dsRNA binding and shifted back to the RNA bound state (Figure 4.30 B). These results strongly suggest that Mtr4 KOW can bind both Nop53 and dsRNA simultaneously and any preference may be fine tuned by the respective concentrations and affinities for both ligands. Due to the close proximity of these residues to the N-terminus the effect in the full-length protein might be different.

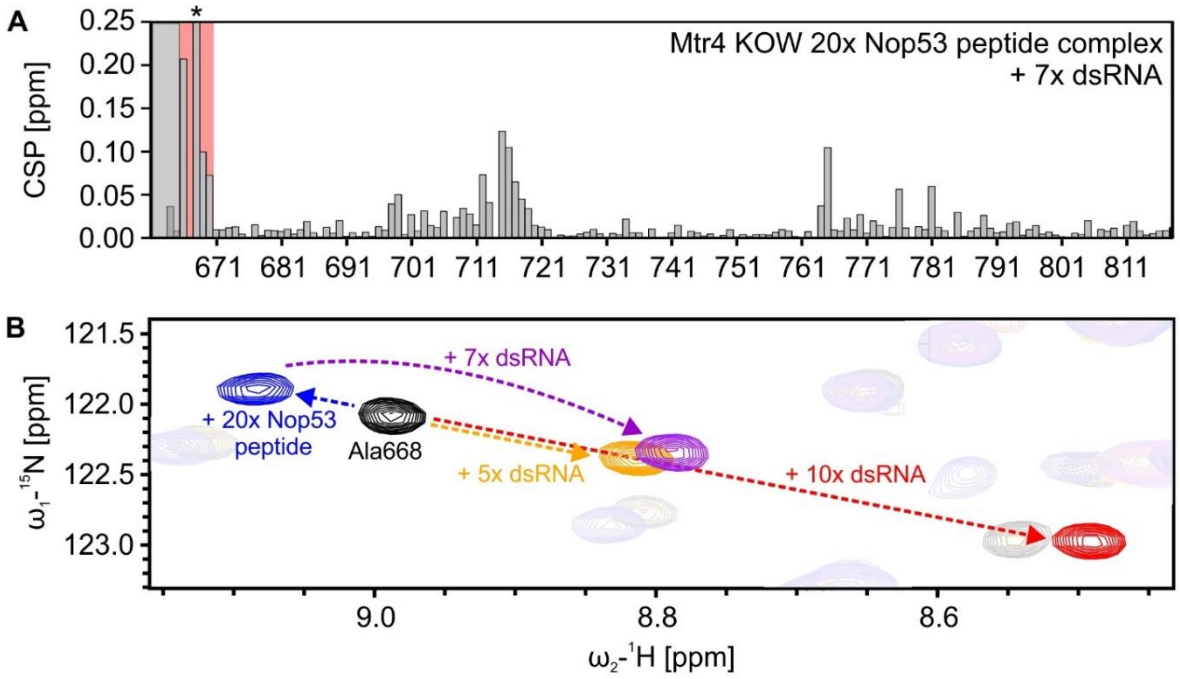



Figure 4.31: Analysis of simultanoues titration of dsRNA and Nop53 peptide. **A)** Chemical shift perturbations of the Mtr4 KOW-Nop53 peptide complex titration with 7x dsRNA plotted against residue number. The pattern corresponds to the difference of Figure 4.30 C and D and reflects the RNA binding pattern from Figure 4.30 A and B. The red box indicates residues shifting to the RNA bound form in presence of both ligands. Residue Ala668 marked with an asterisk shows a CSP value of 0.30 ppm. **B)** A zoom-in of the $^{1}H,^{15}N$-HSQC overlays of free Mtr4 (black), with 5 x dsRNA (orange), 10 x dsRNA (red), 20 x Nop53 peptide (blue) and both 20 x Nop53 peptide plus 7 x dsRNA (purple) shows the shifts of Ala668. Clearly RNA and peptide bound states can be distinguished and in presence of both ligands Ala668 shifts to the RNA bound state.

Note that in all titrations saturation was not reached due to the weak affinity and all spectra showed moderate line broadening in presence of the ligands. From NMR no reliable $K_D$s could be obtained. For both dsRNA and Nop53 peptide complexes heteronuclear NOE data were recorded but did not show any differences in local flexibility (Supplementary Figure 6.8).

### 4.4.4 KOW does not bind single-stranded RNA

To determine the RNA binding specificity a ssRNA was titrated to the Mtr4 KOW domain as well. Upon addition of high excess (7-fold) of single-stranded RNA (ssRNA; 5'-CCUUACCACACAACU-3') no shifts in the $^{1}H^{15}N$-HSQC could be observed (Figure 4.32 A). However, a moderate and uniform line broadening for the majority of signals was detected while several peaks only showed up in presence of ssRNA or gained in intensity (Figure 4.32 A).

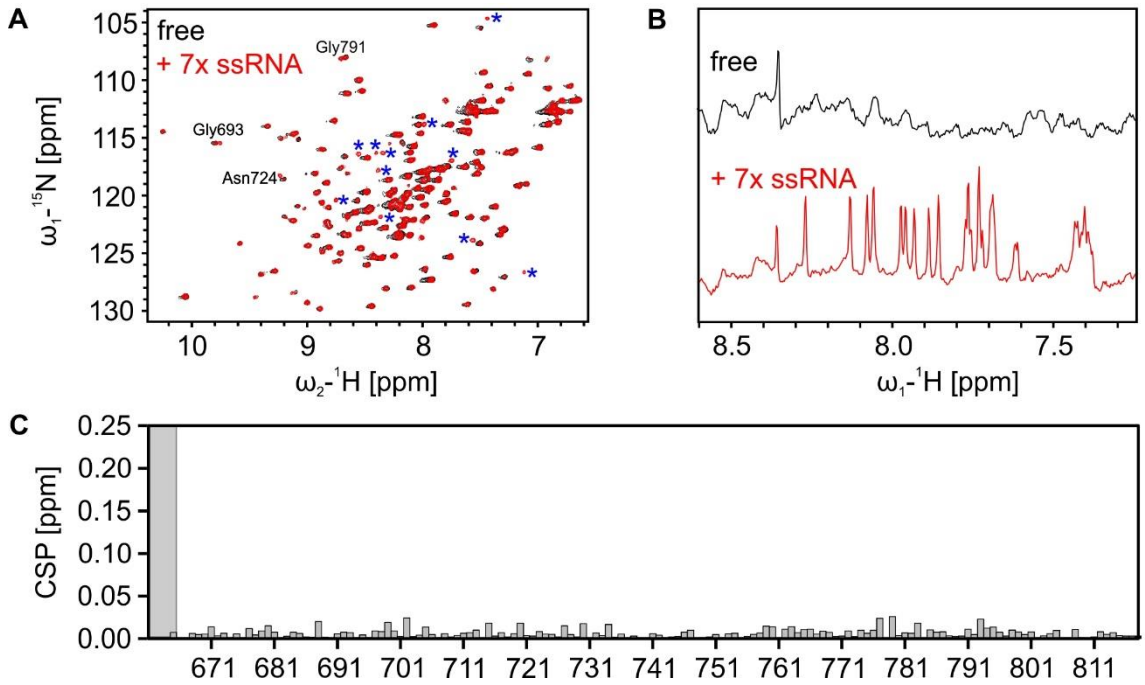



Figure 4.32: KOW titration with single-stranded RNA. **A)** $^{1}H^{15}N$-HSQC overlay of the free KOW domain (black) and in presence of 7-fold ssRNA (5'-CCUUACCACACAACU-3', red). The experimental setup, processing and plotting was identical for both spectra. A moderate loss in intensity can be observed for most signals. Residues marked with a blue asterisk gain in intensity in presence of ssRNA while Gly693, Asn724 and Gly791 show peak splitting. **B)** 1D $^{1}H$ spectra of amino region of free Mtr4 KOW and in presence of ssRNA. **C)** Chemical shift perturbations of the KOW titration with 7 x ssRNA plotted against residue number.

For Gly693, Asn724 and Gly791 peak splitting was observed and the calculated chemical shift perturbations revealed no significant binding sites (Figure 4.32 C). Interestingly, the residues that show peak splitting are in close proximity to each other. Most of the upcoming signals did also gain in intensity when dsRNA or both dsRNA and Nop53 peptide were titrated, while for peptide only these signals were not observed. As a control 1D traces of the amino region were monitored prior and after addition of the RNA (Figure 4.32 B) where sharp peaks indicated the presence of the RNA. These findings suggest ssRNA might interact with a low populated conformation of KOW that is hardly detectable in the free form. It is also possible that the strong negative charge of any RNA has the described effect on the spectrum.

### 4.4.5 The arch domain interacts with KOW

Besides the KOW domain ($Mtr4_{666-818}$) another construct was used comprising the KOW domain and the first two helices of the stalk ($Mtr4_{638-842}$) to test for potential interactions within the protein. An overlay of the $^{1}H,^{15}N$-HSQCs of the free proteins revealed striking differences between the two spectra (Figure 4.33 A). Due to many shifts the assignment of the previously used KOW construct could only be partially transferred. Those residues assigned showed moderate to strong CSPs (Figure 4.33 C). For not assigned residues even stronger CSPs can be expected. These results suggest that the stalk might interact with the KOW domain.

$Mtr4_{638-842}$ was titrated with Phe-tRNA and dsRNA and showed only minor shifts (Figure 4.33 B). Nevertheless, several signals showed severe line broadening in presence of either RNA. Due to the partial assignment no proper analysis of CSPs for RNA titrations was possible. Interestingly, several of the line broadened residues shifted in the $Mtr4_{666-818}$ RNA titrations. These results suggest that the KOW domain with stalk also interacts with tRNA via a dsRNA binding interface. Furthermore, it seems like the binding site remains – at least partially – the same as for the KOW domain itself.

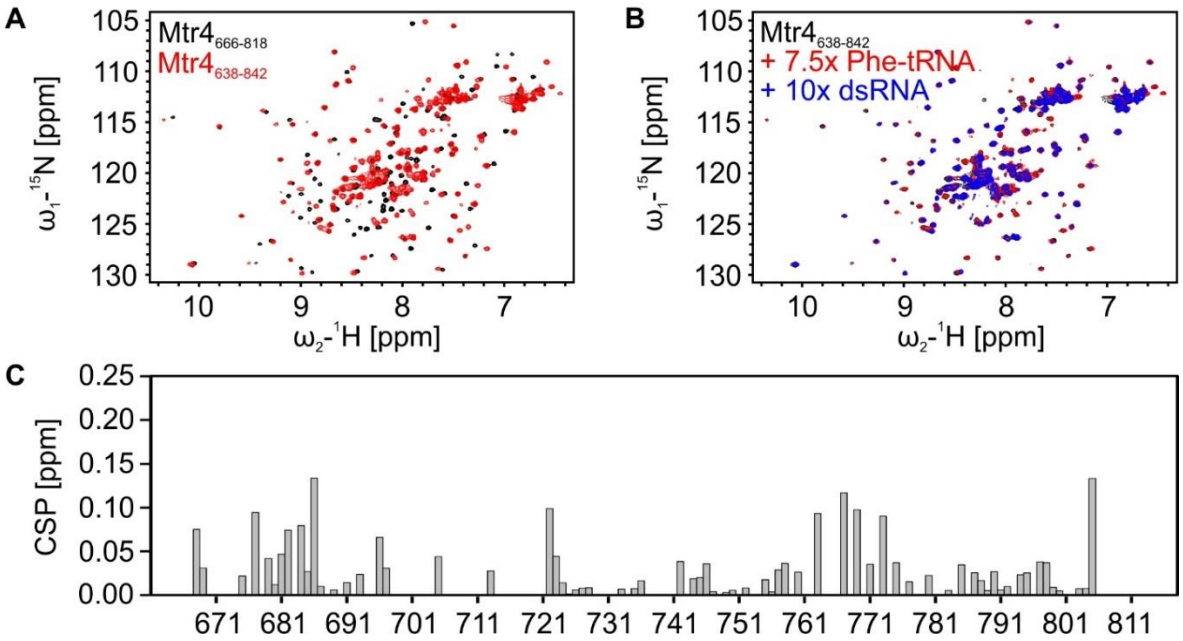



Figure 4.33: Comparison of KOW with and without the extended stalk. **A)** $^{1}H,^{15}N$-HSQC overlay of free $Mtr4_{666-818}$ (corresponds to KOW; black) and $Mtr4_{638-842}$ (red). **B)** $Mtr4_{638-842}$ titration with Phe-tRNA (red) and dsRNA (blue). **C)** CSPs for comparing free $Mtr4_{666-818}$ and $Mtr4_{638-842}$.

# 5 Discussion

## 5.1 Loqs-PD contribution to RISC formation and asymmetry sensing

In this work it was shown that Loqs-PD$_{\Delta NC}$ is monomeric and binds in a 1:1 ratio to its RNA substrate. The C-terminus, however, mediates dimerization of Loqs-PD (Figure 5.1 A) and formation of a heterodimer with Dicer-2 (Figure 5.1 B) as it was shown by crosslinking experiments from Stephanie Fesser and Klaus Förstemann. Whether full length Loqs-PD binds as a dimer to RNA was not tested in this thesis. Dimerization was shown recently for Loqs-PB[221] and thus seems to be conserved for all Loqs isoforms. Since Loqs-PB binds in a 1:1 ratio to Dicer-1[221] it seems plausible that both Loqs-PB and Loqs-PD dimerize only in the absence of the ribonuclease. Dimerization of a dsRBP could serve as a storage pool for Dicer binding partners that is immediately available when needed. This would allow a quick response to stimuli and changing conditions.

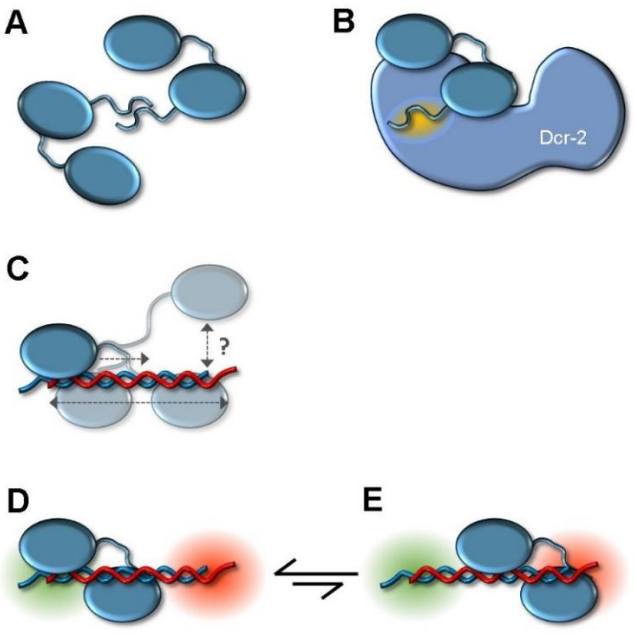



Figure 5.1: Binding properties of Loqs-PD and its interaction partners. **A)** The C-terminus of Loqs-PD mediates dimerization of the protein. **B)** Via the C-terminus Loqs-PD can interact with Dicer-2 (yellow spot). **C)** Loqs-PD's single and tandem domains bind dynamically to dsRNA which potentially involves sliding and dissociation. **D)** Loqs-PD binds to the entire RNA duplex but is often found at the termini. It seems to have a weak preference for the more stable 5' end (highlighted by the green sphere).

A comparison of the single and tandem domains shows that both domains are independent in both the free and RNA bound form (this thesis and previous work from Thomas Kern). Loqs-PD$_{\Delta NC}$ is saturated in presence of a low excess of RNA which confirms the above stated stoichiometry. This agrees also with a titration of a mixture of both single dsRBDs. Interestingly, saturation occurs when the RNA is added in excess of the total protein concentration which corresponds to the sum of both dsRBD concentrations. It suggests that also the individual dsRBDs bind in a 1:1 complex and confirms that the domains are independent. The linker therefor is required to assure that both dsRBDs bind the same RNA simultaneously. Without the linker probably the dynamics within the system (Figure 5.1 C) allow only one dsRBD to bind the RNA.

Overall the linker does not affect the individual domain structures and only influences domain RNA binding when a critical minimal length is undershot. The linker itself does not contribute to RNA binding. Stephanie Fesser and Klaus Förstemann could also confirm that the interaction with Dicer-2 is not affected by the linker length. Hence, the long linker between the two domains is required for simultaneous RNA binding but its length uncouples the binding modes and allows flexibility which seem to be required for the protein to sample different binding sites. When too short the linker probably restricts motions and optimal domain arrangement.

From cross-linking experiments (Stephanie Fesser, Klaus Förstemann) it is known that Loqs-PD binds preferentially to duplex RNA termini. However, SAXS and NMR data strongly indicate binding to be dynamic. The PRE experiments confirmed that binding occurs at the ends of the RNA but it seems to be transient because the overall PRE effect observed was very weak. Thus at a given time point only a small portion of Loqs-PD is found at either end. All data combined point at sliding of the protein on RNA like it was shown for human TRBP[76] and Staufen-1[75].

Despite the dynamic binding a weak preference of Loqs-PD to bind the more stable 5’ end of the RNA duplex could be detected (Figure 5.1 D). Interestingly, the cross-linking data of the ternary complex showed that Dicer-2 binds the less stable end while Loqs-PD occupies the stable terminus. An asymmetric complex forms (figure 5.2 B) which could set the basis for strand selection during siRNA maturation and RISC formation. Therefor, Loqs-PD seems to scan the RNA and its high affinity for stable RNA duplexes might direct the protein to more

stably paired binding sites. This means Loqs-PD and potentially other dsRBPs as well have an intrinsic ability to sense the thermodynamic asymmetry of an RNA duplex and contribute to strand selection. Different experiments already indicated other dsRBPs to be capable to sense thermodynamic stability as well[79,80]. In this work for the first time an NMR-based approach was used to visualize asymmetry sensing. Besides the different functions described before[65] the contribution of a dsRBP to strand selection seems to be a key feature in RISC maturation. The effect is weak though and needs to be amplified by Dicer-2. Data from cross-linking experiments with Loqs-PD$_{\Delta NC}$ show that no asymmetric complex forms and Dicer-2 alone distributes equally between both termini. The dsRBP is hence the driving force for strand selection and the protein-protein interaction of Loqs-PD and Dicer-2 is thus required to fix the ternary complex and switch from dynamic to static binding. To confirm this hypothesis further RNAs could be tested by both NMR and cross-linking where the thermodynamic difference between both termini is increased. Note that the RNA used in this work has the minimal difference as the termini differ only in one GC base pair (Figure 4.1 A).

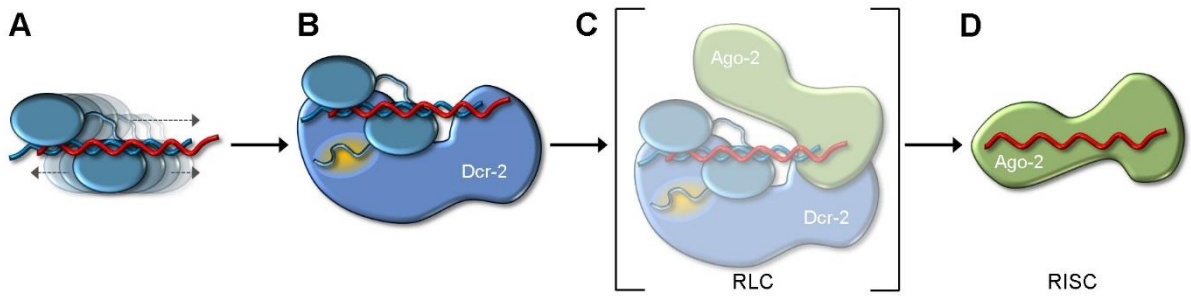



Figure 5.2: Possible function of Loqs-PD during RISC formation. **A)** Loqs-PD binds to stable RNA duplexes and slides along its substrate, possibly scanning for more stably paired regions. **B)** Dicer-2 binds to Loqs' C-terminus and fixes the conformation to form an asymmetric complex where Loqs-PD binds to the stable and Dicer-2 to the less stable end. **C)** Together with Ago-2 an alternative RISC loading complex is formed, the passenger strand is degraded and **D)** the mature RISC forms.

All data from this thesis, collaborators and literature combined can be used for a refined model of siRNA maturation and RISC formation in *Drosophila* (Figure 5.2). Loqs-PD binds dynamically to a precursor siRNA (Figure 5.2 A) and scans for stable regions. Upon engagement of Dicer-2 the two proteins interact via Loqs' C-terminus and form an asymmetric complex where Loqs binds the stable and Dicer the unstable terminus (Figure 5.2 B). Romy

Böttcher and Klaus Förstemann could show that Loqs-PD and Dicer-2 are sufficient to load siRNAs onto Ago-2 in cells (data not shown). Thus after recruiting Ago-2 to the ternary complex an alternative RISC loading complex forms (Figure 5.2 C). Here strand selection occurs as well as degradation of the passenger strand. The guide strand afterwards is loaded onto Ago to form the mature RISC (Figure 5.2 D). Besides its various other functions like the modulation of Dicer's substrate specificity and processivity of dsRNA cleavage Loqs-PD is a key factor to sense the thermodynamic stability of a duplex RNA. Thereby it sets the basis for strand selection and assures proper siRNA loading onto RISC.

## 5.2 A semi-quantitative look at dsRBD sliding on dsRNA

Different experiments presented in this thesis pointed at a sliding mechanism of Loqs-PD (see 5.1). Sliding was described for other dsRBPs as well[75,76] and seems plausible because most protein-RNA contacts occur via the phosphate backbone or the sugar moieties of the nucleic acids[68,72,73]. Specific contacts are thus unlikely except for RNAs that differ from the ideal A-type helix or show mismatches. This was for example shown for ADAR2 that binds sequence-specifically to stem loop RNAs with mismatches[74].

An interesting experiment was made by Loth et al.[218] who analyzed sliding of the DNA-binding lac repressor using NMR. In this thesis the approach was transferred to Loqs-PD for RNA binding analyses. I could successfully establish for the first time an NMR-based sliding analysis of a multidomain protein bound to RNA. In contrast to the lac repressor, Loqs-PD binds with a significantly higher affinity to its substrate which suggests the data to be more reliable since the majority of the protein should be bound in a stable complex.

The line broadening analysis (see 4.2) confirms the data from 4.1 and proves that the individual dsRBDs (Figure 5.3 C) as well as tandem dsRBDs (Figure 5.3 A) of Loqs-PD slide on dsRNA. This is in agreement with other studies[76]. The RNA length seems to be a key factor to modulate sliding: For a short RNA of 16bp only minimal line broadening was observed which corresponds to no sliding. The RNA is too short to enable sliding and even not long enough to allow full affinity binding[67,72]. For increasing RNA lengths sliding gets more severe. The RNA sequence on the other hand has hardly any influence on sliding (Figure 5.3 B) and chemical shift perturbations. Both findings confirm that binding (and also sliding) is sequence-independent. Different stabilities of RNA duplexes as well did not show any effect. Interpretation of this data is critical though since all experiments were done at 25°C where all RNAs are a stable duplex according to CD melting point analyses. However, local fluctuations or ‘breathing’ of the duplex RNA could be possible.

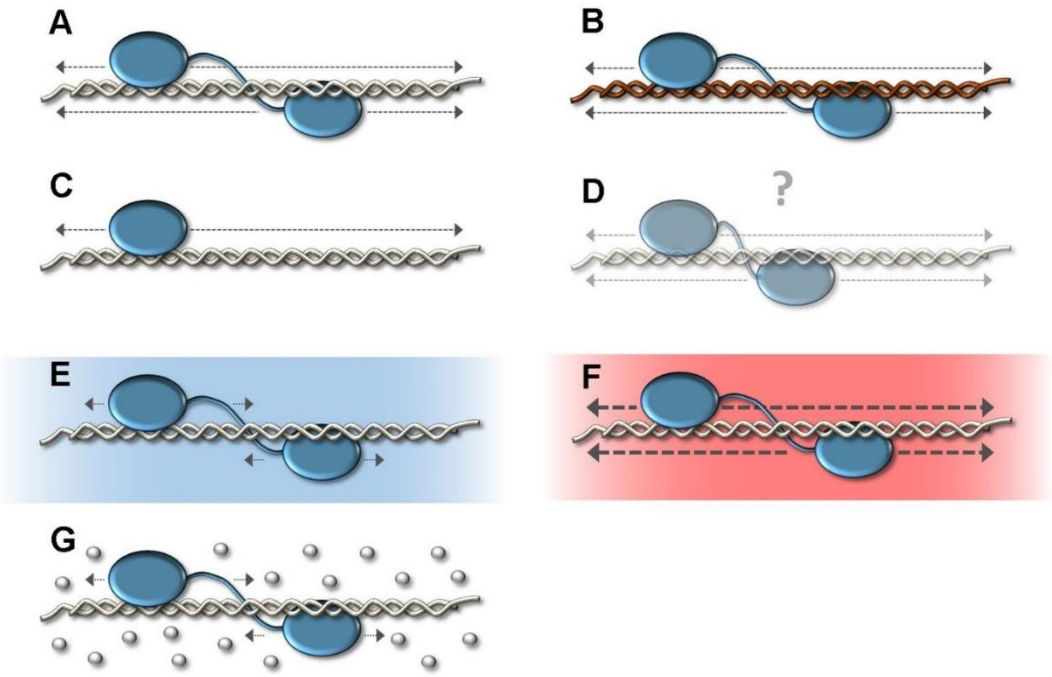



Figure 5.3: Effects on sliding of Loqs-PD tandem dsRBDs. **A)** The tandem dsRBDs of Loqs bind RNA simultaneously and exhibit sliding on the dsRNA. **B)** Sliding seems to be independent of the RNA sequence. Like tandem domains **C)** each individual dsRBD slides as well. **D)** The importance of the linker for sliding remains elusive, however, it seems to play a minor role. **E)** Lower temperatures and **G)** increased viscosity slow down sliding. **F)** For elevated temperatures sliding is enhanced.

Single molecule experiments and MD simulations showed that the linker between two dsRBDs is crucial for sliding activity[75]: A short linker reduces sliding. The linker mutants of Loqs tested in this work did not show a significant effect on sliding (Figure 5.3 D). It is possible that, similarly for the binding mode discussed in 5.1, once a critical linker length is exceeded it does not affect sliding any more. Only a very short linker might limit dynamics or even enhance dynamics since the binding mode seems to be suboptimal for those mutants. Besides that Wang et al.[75] compared different proteins which might have a different sliding propensity by default. Therefor, the importance of the linker for sliding cannot be finally assessed. It is obvious, however, that the linker role is far less important than RNA length or temperature.

In the NMR experiment all parameters are defined and an artificial system is analyzed. Hence I wanted to estimate whether sliding could also occur within the cell and is potentially relevant. To mimic cellular conditions different temperatures and viscosities were tested in line broadening experiments. Overall viscosity limits sliding (Figure 5.3 G) while increased

temperatures enhance motions (Figure 5.3 E, F) as expected. In an approximation (see 4.2.6 and 4.2.7) in cellular cytoplasm at elevated temperatures (1-2 cP viscosity, 30-37°C) the two opposing effects of viscosity and temperature could cancel each other out. Sliding in the cell would thus be comparable to the NMR experiments and our data reflect a realistic scenario. For sure other parameters *in vivo* will influence sliding, like cofactors and binding partners.

Overall Loqs-PD slides on duplex RNAs though sliding could be affected by potential dimerization (see 5.1). As known from other systems binding and sliding is sequence-independent and strongly influenced by external parameters that influence thermodynamics, e.g. temperature and viscosity. Sliding is thus not only a feature already known from DNA binding proteins but is rather common for (double-stranded) RNA binding proteins as well. It could facilitate scanning of RNA duplexes for thermodynamic stability. The complex of TRBP and Dicer slides as well[76] and sliding possibly enhances cleavage efficiency of Dicer when bound to si- or miRNA precursors. Other translocation mechanisms known from DNA binding proteins are intersegment transfer and hopping. It would be interesting to extend this study to also test for these processes, albeit both mechanisms are harder to address.

## 5.3 Staufen-2 dsRBD1 and 2 bind RNA localization signal

So far the function of dsRBD1 and 2 of mouse Staufen-2 were poorly characterized. Also their contribution to RNA binding remained elusive. In the NMR experiments presented (4.3) both dsRBDs bind to the double-stranded stem loop RNA alone and in tandem (Figure 5.4 B, C). Like for Loqs-PD the two domains seem to be independent of each other (Figure 5.4 A). In contrast to that SPR measurements indicate cooperative binding of both dsRBDs. Thus besides dsRBD3 and 4 dsRBD1 and 2 of Staufen-2 also contribute to RNA binding. Since the RNA used in this thesis is probably part of a larger composition of secondary structure elements in a 3'-UTR it is possible that the four dsRBDs bind to different stem loops to recognize specific shapes.

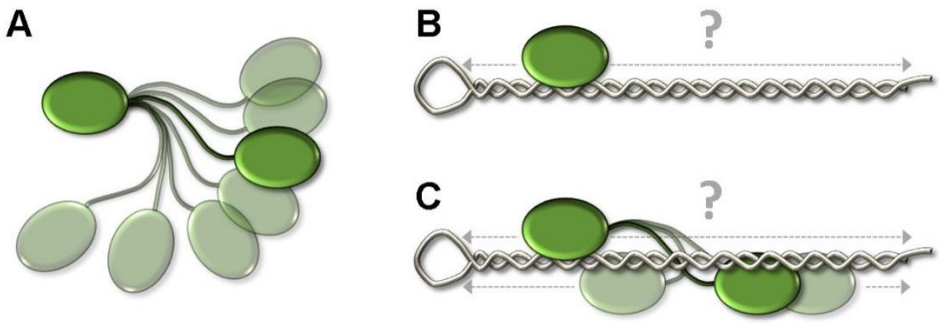



Figure 5.4: Staufen-2 tandem dsRBDs 1-2 are independent and bind RNA dynamically. **A)** The two Staufen 2 dsRBD1 and 2, connected by a linker, are independent in the free form. **B)** Each single domain binds to stem loop RNA (dsRNA respectively). Binding is likely to be dynamic and possibly binding occurs at different time scales for the two domains. **C)** In the tandem construct both domains bind to RNA via the binding interface shown in the single domain titrations. Considerable dynamics during binding are present as well.

For all titrations line broadening was observed for both the protein as well as the RNA. This points at dynamics which could include dissociation and re-association or sliding. The affinity was determined by Simone Heber and Dierk Niessing by SPR experiments: Both dsRBDs bind RNA with low nanomolar affinity. Therefor, dissociation seems unlikely. Furthermore, line broadening is different for the two single dsRBDs which suggests that binding involves different dynamics for the two domains and motions occur at another time scale.

The started assignment revealed some RNA binding residues for dsRBD1. These are conserved and mutants showed reduced binding in SPR experiments. The structure models suggest the conserved αβββα-fold for both dsRBDs. According to the models the assigned RNA binding residues are located in the loop connecting the first two β-sheets. However, dsRBD2 has an extended loop in the structure model which agrees with the sequence alignment. NMR and SAXS data confirm that at least dsRBD1 oligomerizes or exists in multiple conformations. It is known that Staufen-1 dimerizes via its dsRBD5[222], but a similar function has not been described so far for dsRBD1.

DsRBD1 and 2 thus both bind dsRNA and binding is likely to be dynamic as it was observed for Loqs (see 5.2). Both Loqs-PD and Staufen-2 show characteristic line broadening in protein and RNA NMR spectra. Binding of dsRBD1 furthermore could occur via the conserved binding mode though further data needs to be analyzed. Another shared feature is the independence of both domains. Here, further information are required since the lacking assignment does not allow to determine the linker length between the two Staufen dsRBDs.

## 5.4 The Mtr4 KOW domain contributes to tRNA binding and recruitment to the ribosome

Based on the experiments shown before it seems unlikely that Mtr4 KOW interacts with ssRNA which agrees with literature (Figure 5.5 A)[194]. Most peaks that show up in presence of ssRNA can also be found in presence of dsRNA with or without the Nop53 peptide. A possible explanation could be the existence of a low populated conformation of KOW that interacts (maybe unspecifically) with any kind of RNA. Additionally, the negative charge of the RNA could interact unspecifically with parts of the protein.

The NMR data presented in this thesis revealed that structured RNAs and Nop53 bind to the Mtr4 KOW domain via distinct interfaces (Figure 5.5 C, D). The solution data for the peptide complex is in agreement with the crystal structure from our collaborator Sebastian Falk (data not shown). Among the residues affected are leucine, isoleucine and valine which could bind the AIM of Nop53 via hydrophobic interactions and the crystal structure indeed confirms that these residues interact with residues 59-69 of Nop53 (ALFHVDVEGDE) which comprise a small β-strand that packs antiparallel against β-strand 5 of KOW.

The RNA titrations show that KOW interacts with dsRNA and tRNA via the same binding site suggesting recognition of tRNAs (or otherwise structured RNAs) is depending on double-stranded regions within the RNA. This site comprises residues that are good h-bond donors (histidine, serine, asparagine, aspartate and glutamate) which makes them suitable RNA binding residues.

The different binding sites allow simultaneous binding of both Nop53 and RNA substrates, however, based on the NMR titrations the binding sites overlap partially. The residues involved switch to the RNA bound form when both ligands are present (Figure 5.5 E). In presence of 20-fold Nop53 peptide a 7-fold access of RNA was needed to achieve similar shifts as for 5-fold access of RNA only. This indicates that the two ligands influence each other and might modulate affinities of KOW. Individual affinities and ligand concentrations will fine tune complex formation. A simultaneous binding mode seems important for the biological function of Mtr4 and the exosome: The RNA binding site points to the inside of Mtr4 and the ssRNA binding channel of the core while Nop53 binds to the top of the domain. This

arrangement enables Mtr4 to be recruited to its target via Nop53/Utp18 while binding its RNA substrate. Depending on the affinities in the natural complex (means with all cofactors) it could even be possible that upon ribosome/substrate binding Nop53/Utp18 is released from Mtr4.

Overall KOW has a low affinity (based on the NMR titrations where no saturation was reached) for all tested ligands[199], which agrees with reports that mutations in the KOW domain reduce RNA binding and unwinding but do not abolish it completely[210]. Since both the Mtr4 core (Rec domains) as well as Air1/2 bind the RNA substrate, the KOW domain probably assists in RNA binding and assures correct orientation of the substrate to enable proper unwinding and access to the Mtr4 channel. Furthermore, it serves as a platform to mediate interactions to its ribosome targets via Nop53 and Utp18.

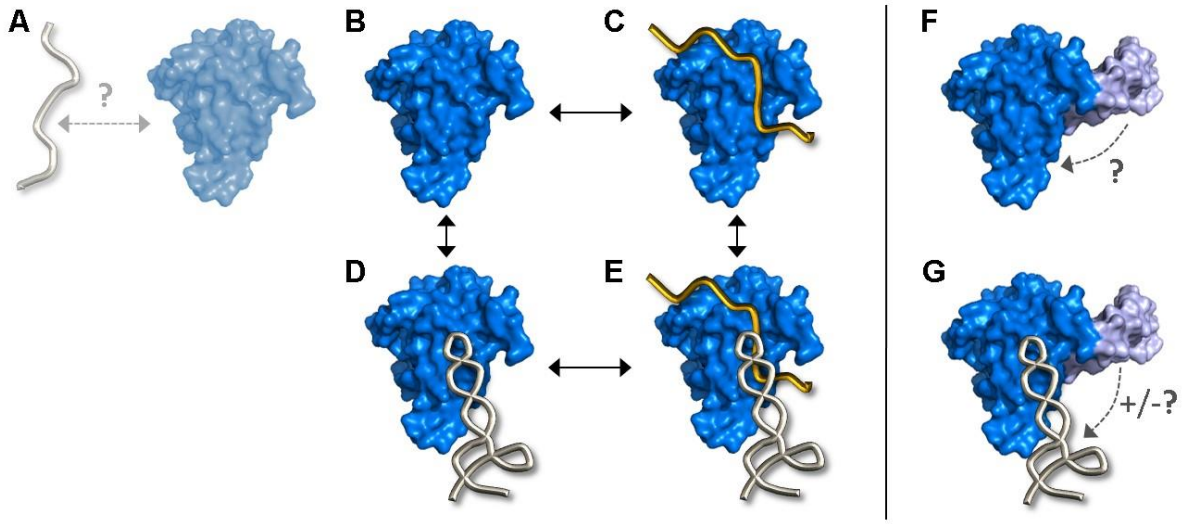



Figure 5.5: Interaction modes of Mtr4 KOW domain with different ligands. **A)** ssRNA might interact with a minor population of KOW (blue, adapted from PDB: 2XGJ) though direct proof is missing. **B, C, D, E)** Two potential sequential binding schemes for Nop53 and tRNA binding are shown. The binding site for D structured RNAs (grey) differs from the C Nop53 (gold) binding interface. **F)** The stalk (lilac) interacts with the KOW domain and **G)** might contribute to RNA binding or modulate it.

The NMR experiments for the extended KOW domain (including parts of the stalk) point at an interaction of the stalk with the KOW domain (Figure 5.5 F) and a possible participation of the stalk in RNA binding or at least modulation of RNA binding (Figure 5.5 G). A scenario where the stalk modulates accessibility and/or affinity of KOW for RNA seems thus plausible. Due to large shifts of the free extended form compared to the KOW domain a new

backbone assignment would be necessary to allow a proper analysis and extended model. The influence of the stalk could also be an artefact that could be absent in the context of full length Mtr4 or in presence of cofactors. Here further structural experiments would be necessary, e.g. SAXS/SANS to determine the overall shape.

Based on this work and literature data a possible model for pre-ribosomal RNA processing involves initial binding of Nop53/Utp18 via its AIM to KOW (Figure 5.6 A, B). Thereby the respective factor recruits the TRAMP-exosome complex to the pre-ribosome and KOW binds the ribosomal RNA precursor (Figure 5.6 C). This binding is accompanied by a conformational change within the AIM of Nop53/Utp18. Here the cofactor might get released from the complex. When the RNA is properly bound by the Mtr4 core and Air1/2 it gets unwound through the central channel of Mtr4 and degraded by the connected exosome (Figure 5.6 D).

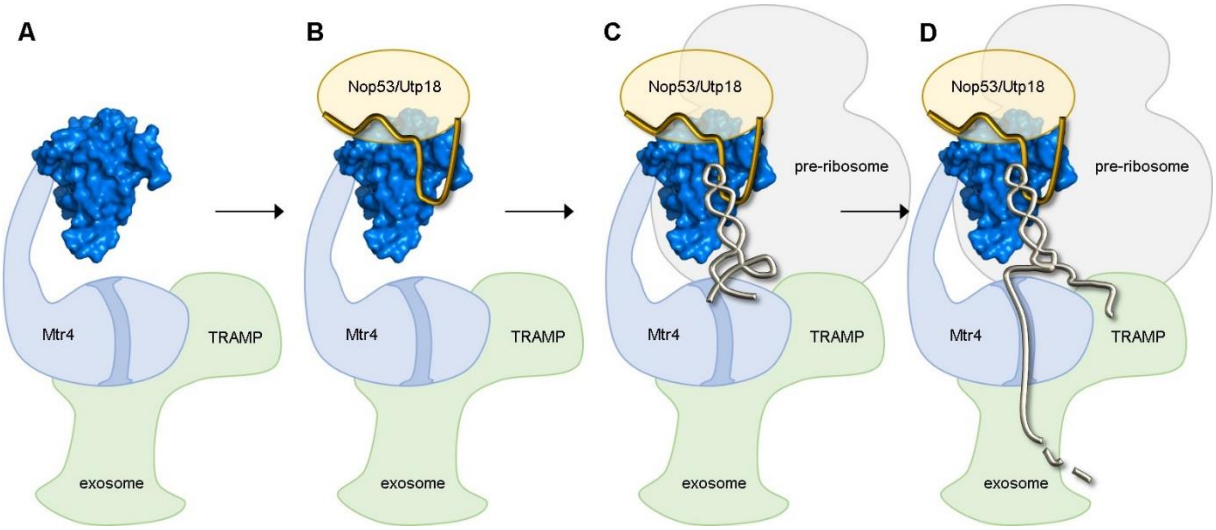



Figure 5.6: A model for ribosome precursor-RNA processing via the TRAMP-Exosome complex. **A)** Mtr4 (blue, adapted from PDB: 2XGJ) as part of the TRAMP complex (green) binds the exosome (green). **B)** Nop53//Utp18 bind via their AIM to the KOW domain of Mtr4 and recruit the entire complex to **C)** the pre-ribosome. Upon binding of ribosomal precursor-RNAs the AIM of Nop53/Utp18 performs a conformational switch and partially releases Mtr4. **D)** The RNA gets processed and unwound by Mtr4 (TRAMP) and subsequently degraded by the exosome.

# 6 Appendix

## 6.1 Loqs siRNA binding

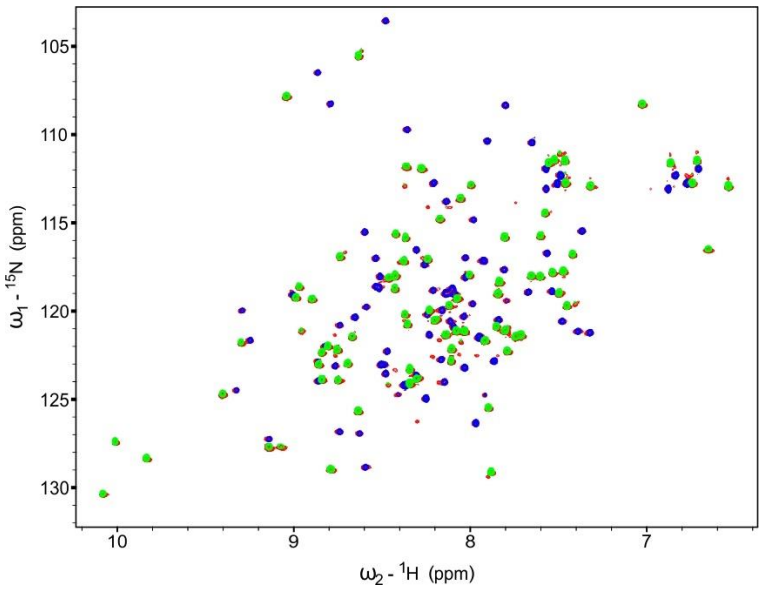
Supplementary Table 6.1: Dali alignments for dsRBD1 and dsRBD2.

**dsRBD1**

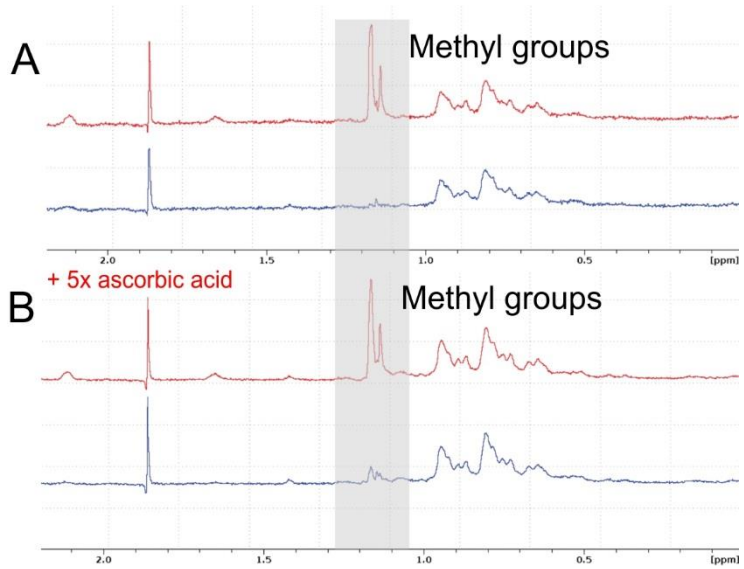
| No | Chain | Z | rmsd | lali | nres | %id | PDB Description |
|---|---|---|---|---|---|---|---|
| **1** | 1rc7-A | 7.8 | 2.2 | 60 | 220 | 33 | Ribonuclease III |
| **2** | 3llh-B | 7.7 | 2.3 | 59 | 66 | 42 | RISC-Loading Complex Subunit TARBP2 |
| **3** | 3adl-A | 7.6 | 2.0 | 59 | 76 | 27 | RISC-Loading Complex Subunit TARBP2 |
| **4** | 1di2-A | 7.6 | 2.1 | 59 | 69 | 29 | Double Stranded RNA Binding Protein A |
| **5** | 2nuf-A | 7.6 | 2.2 | 61 | 219 | 33 | 28-mer |
| **6** | 1yz9-A | 7.6 | 2.2 | 60 | 220 | 33 | Ribonuclease III |
| **7** | 2nuf-B | 7.5 | 2.3 | 61 | 219 | 33 | 28-mer |
| **8** | 2nug-B | 7.5 | 2.2 | 61 | 218 | 34 | 5’-R(P*AP*AP*GP*GP*UP*CP*AP*UP*UP*CP*G)-3’ |
| **9** | 1yz9-B | 7.5 | 2.5 | 61 | 218 | 33 | Ribonuclease III |
| **10** | 1yyk-A | 7.4 | 2.2 | 60 | 221 | 33 | Ribonuclease III |

**dsRBD2**

| No | Chain | Z | rmsd | lali | nres | %id | PDB Description |
|---|---|---|---|---|---|---|---|
| **1** | 1di2-A | 11.1 | 1.6 | 66 | 69 | 41 | Double Stranded RNA Binding Protein A |
| **2** | 3adl-A | 10.6 | 1.6 | 66 | 76 | 41 | RISC-Loading Complex Subunit TARBP2 |
| **3** | 1rc7-A | 10.1 | 1.9 | 66 | 220 | 35 | Ribonuclease III |
| **4** | 2nug-B | 10.1 | 2.0 | 66 | 218 | 35 | 5’-R(P*AP*AP*GP*GP*UP*CP*AP*UP*UP*CP*G)-3’ |
| **5** | 1yyk-A | 10.0 | 1.8 | 66 | 221 | 35 | Ribonuclease III |
| **6** | 2nue-B | 10.0 | 1.7 | 66 | 220 | 35 | 46-mer |
| **7** | 1yyw-D | 10.0 | 1.6 | 66 | 219 | 35 | Ribonuclease III |
| **8** | 1yyw-A | 10.0 | 1.7 | 66 | 219 | 35 | Ribonuclease III |
| **9** | 2nuf-A | 9.9 | 2.0 | 66 | 219 | 35 | 28-mer |
| **10** | 1yz9-A | 9.9 | 1.9 | 66 | 220 | 35 | Ribonuclease III |

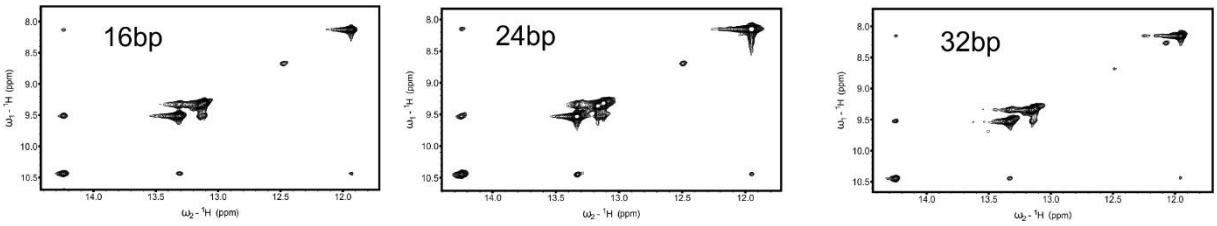



Supplementary Figure 6.1: $^1$H,$^{15}$N-HSQC overlay of single dsRBDs and mixed.

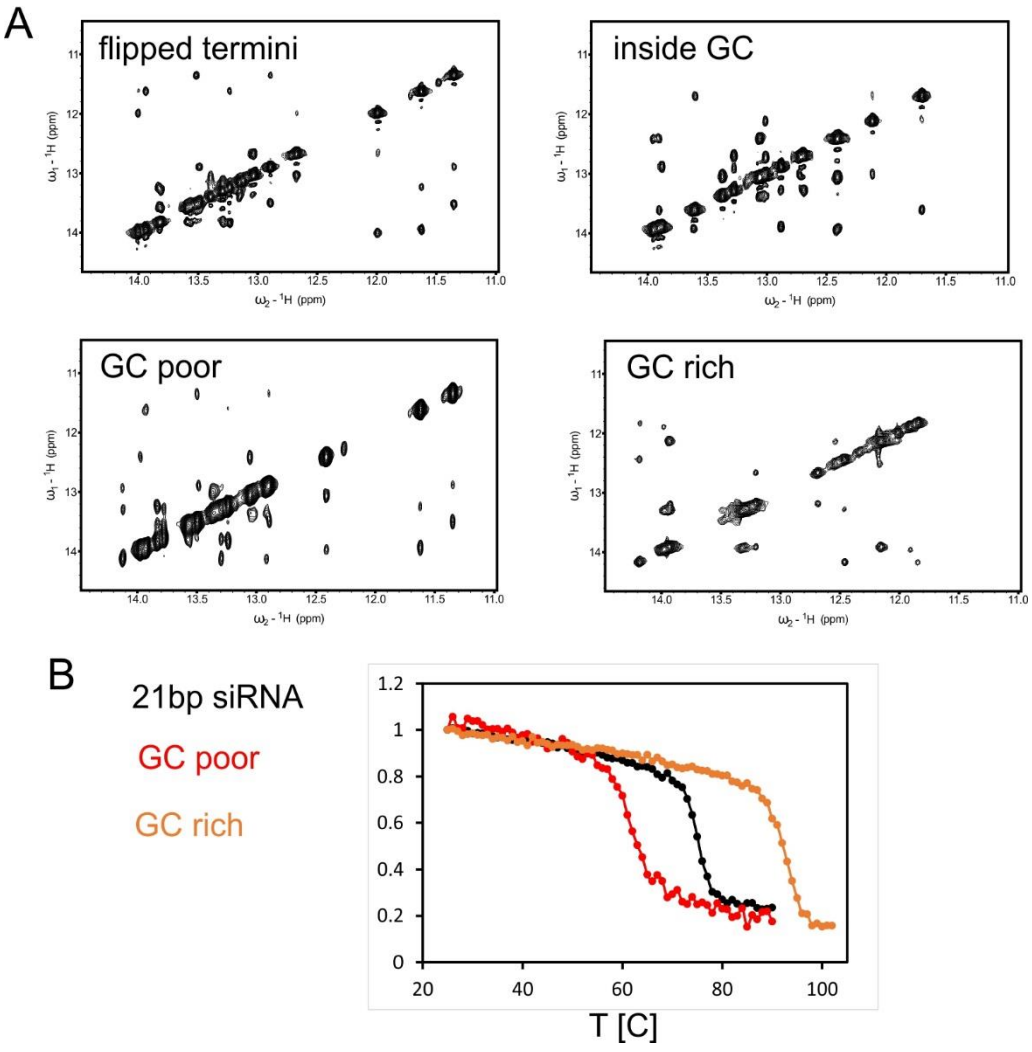



Supplementary Figure 6.2: Methyl groups confirm successful reduction of TEMPO spin label. Upon reduction (red) with ascrobic acid the characteristic signals corresponding to the methyl groups of TEMPO show up (grey box). **A)** TEMPO spin label attached to the stable 5' end and **B)** to the less stable end.

## 6.2 Loqs sliding




Supplementary Figure 6.3: $^{1}$H,$^{1}$H-NOESYs of palindromic RNAs.

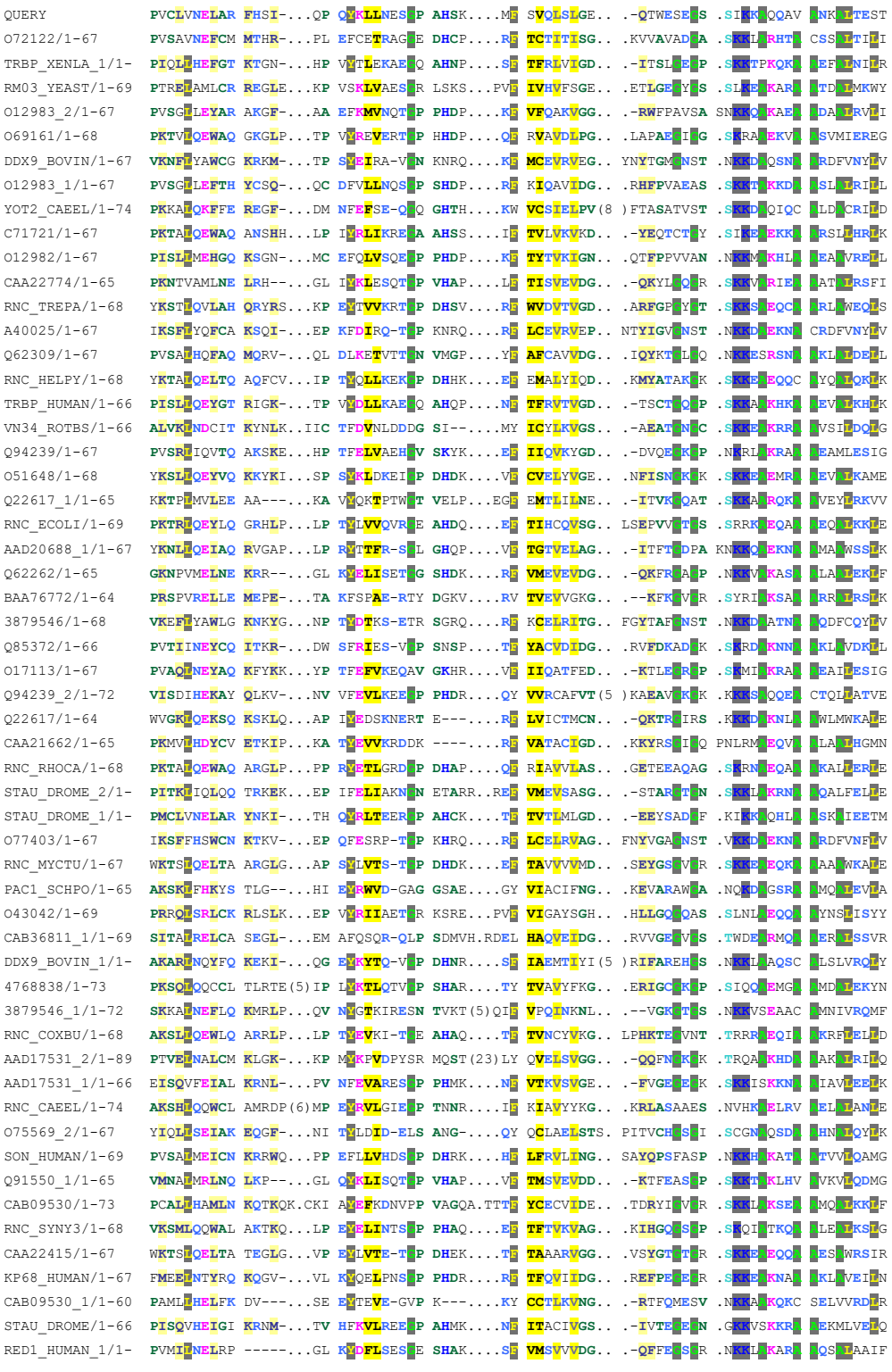


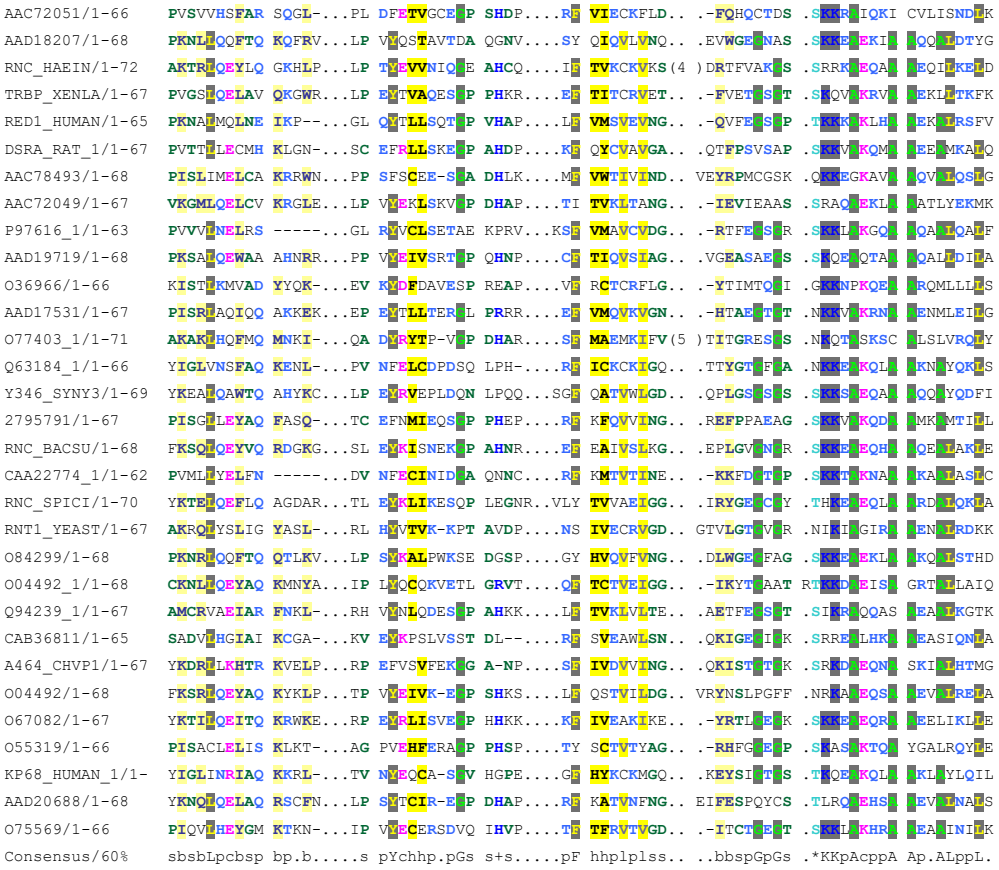

Supplementary Figure 6.4: $^{1}$H,$^{1}$H-NOESYs and CD spectra of different siRNAs.**A)** Imino NOESY spectra of different siRNAs show the connectivities between neighbouring residues. **B)** CD spectra of selected siRNAs. The GC content strongly influences the melting temperature.

## 6.3 Staufen 2

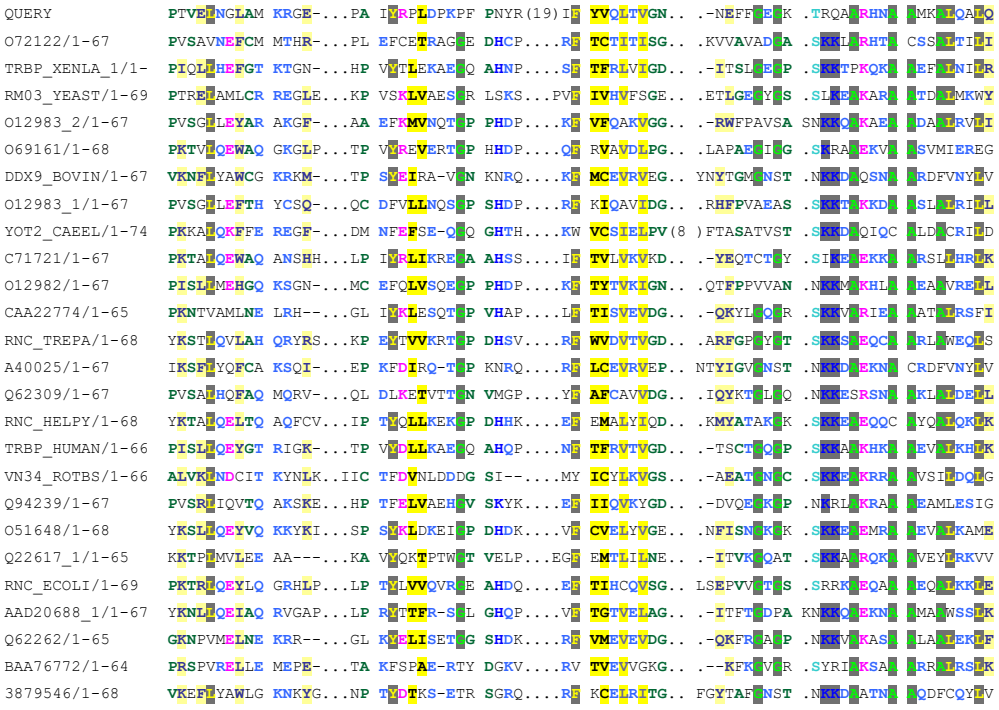
### Supplementary Table 6.2: Sequence alignment of Staufen-2 dsRBD1 and 2.

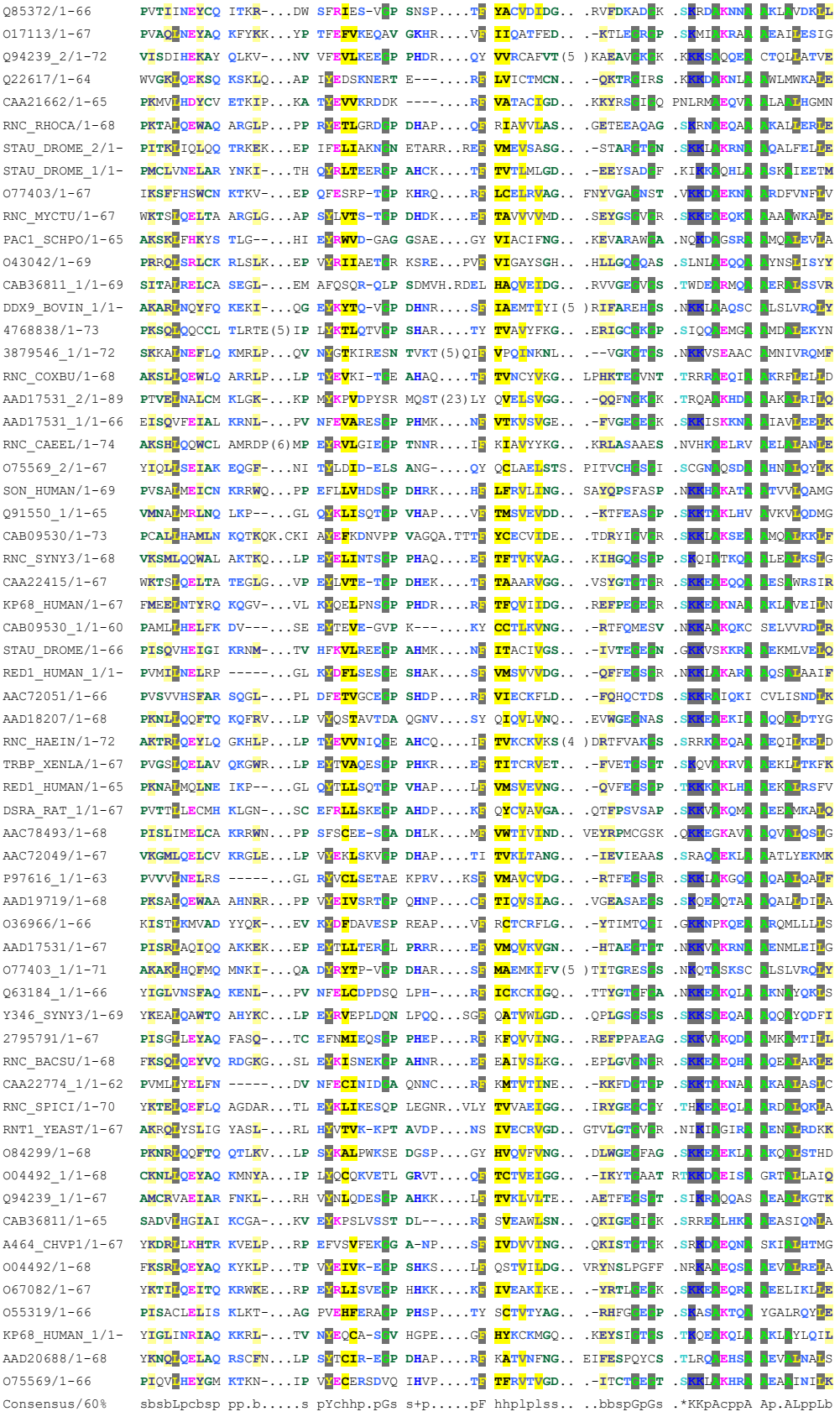
Sequence alignment of Staufen dsRBD1 (residues 9-74) from SMART webserver (http://smart.embl-heidelberg.de/).



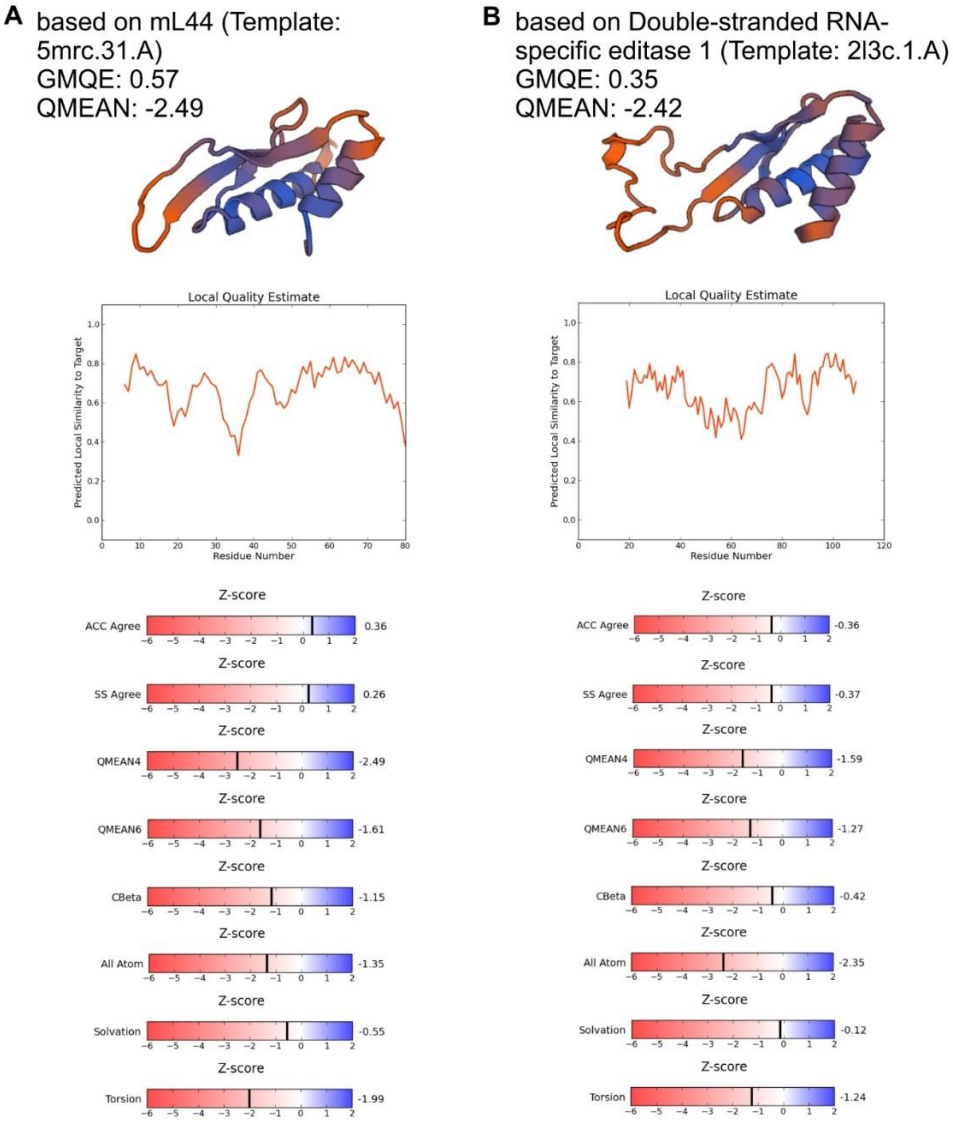


Sequence alignment of Staufen dsRBD2 (residues 96-180) from SMART webserver (http://smart.embl-heidelberg.de/).

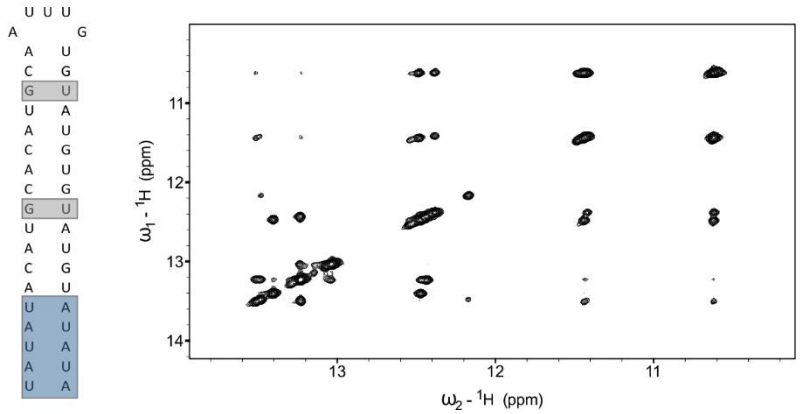



Q85372/1-66
O17113/1-67
Q94239_2/1-72
Q22617/1-64
CAA21662/1-65
RNC_RHOCA/1-68
STAU_DROME_2/1-
STAU_DROME_1/1-
O77403/1-67
RNC_MYCTU/1-67
PAC1_SCHPO/1-65
O43042/1-69
CAB36811_1/1-69
DDX9_BOVIN_1/1-
4768838/1-73
3879546_1/1-72
RNC_COXBU/1-68
AAD17531_2/1-89
AAD17531_1/1-66
RNC_CAEEL/1-74
O75569_2/1-67
SON_HUMAN/1-69
Q91550_1/1-65
CAB09530/1-73
RNC_SYNY3/1-68
CAA22415/1-67
KP68_HUMAN/1-67
CAB09530_1/1-60
STAU_DROME/1-66
RED1_HUMAN_1/1-
AAC72051/1-66
AAD18207/1-68
RNC_HAEIN/1-72
TRBP_XENLA/1-67
RED1_HUMAN/1-65
DSRA_RAT_1/1-67
AAC78493/1-68
AAC72049/1-67
P97616_1/1-63
AAD19719/1-68
O36966/1-66
AAD17531/1-67
O77403_1/1-71
Q63184_1/1-66
Y346_SYNY3/1-69
2795791/1-67
RNC_BACSU/1-68
CAA22774_1/1-62
RNC_SPICI/1-70
RNT1_YEAST/1-67
O84299/1-68
O04492_1/1-68
Q94239_1/1-67
CAB36811/1-65
A464_CHVP1/1-67
O04492/1-68
O67082/1-67
O55319/1-66
KP68_HUMAN_1/1-
AAD20688/1-68
O75569/1-66
Consensus/60%

| Group name | Amino acids | Displayed as |
|---|---|---|
| Default | X | . |
| Single | X | - |
| Alanine | A | A |
| Cysteine | C | C |
| Aspartic Acid | D | D |
| Glutamic Acid | E | E |
| Phenylalanine | F | F |
| Glycine | G | G |
| Histidine | H | H |
| Isoleucine | I | I |
| Lysine | K | K |
| Leucine | L | L |
| Methionine | M | M |
| Asparagine | N | N |
| Proline | P | P |
| Glutamine | Q | Q |
| Arginine | R | R |
| Serine | S | S |
| Threonine | T | T |
| Valine | V | V |
| Tryptophan | W | W |
| Tyrosine | Y | Y |
| Negative | D,E | - |
| Ser/Thr | S,T | * |
| Aliphatic | I,L,V | l |
| Positive | H,K,R | + |
| Tiny | A,G,S | t |
| Aromatic | F,H,W,Y | a |
| Charged | D,E,H,K,R | c |
| Small | A,C,D,G,N,P,S,T,V | s |
| Polar | C,D,E,H,K,N,Q,R,S,T | p |
| Big | E,F,H,I,K,L,M,Q,R,W,Y | b |
| Hydrophobic | A,C,F,G,H,I,L,M,T,V,W,Y | H |

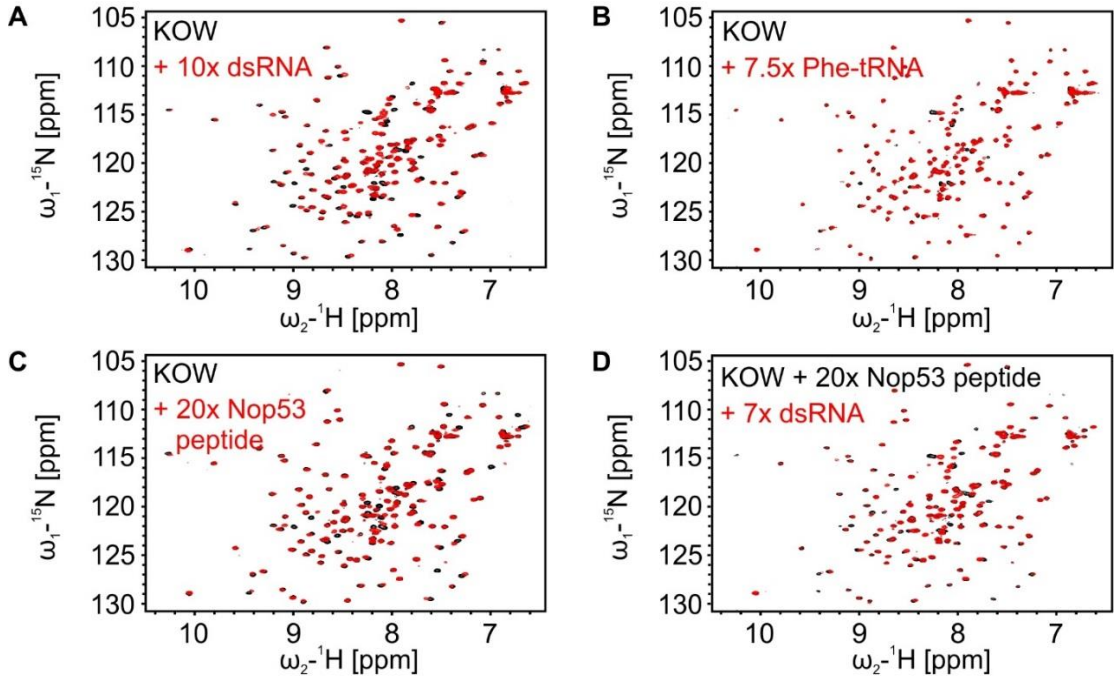



Supplementary Figure 6.5: Extended data for structure models of Staufen-2 dsRBD1 and 2.

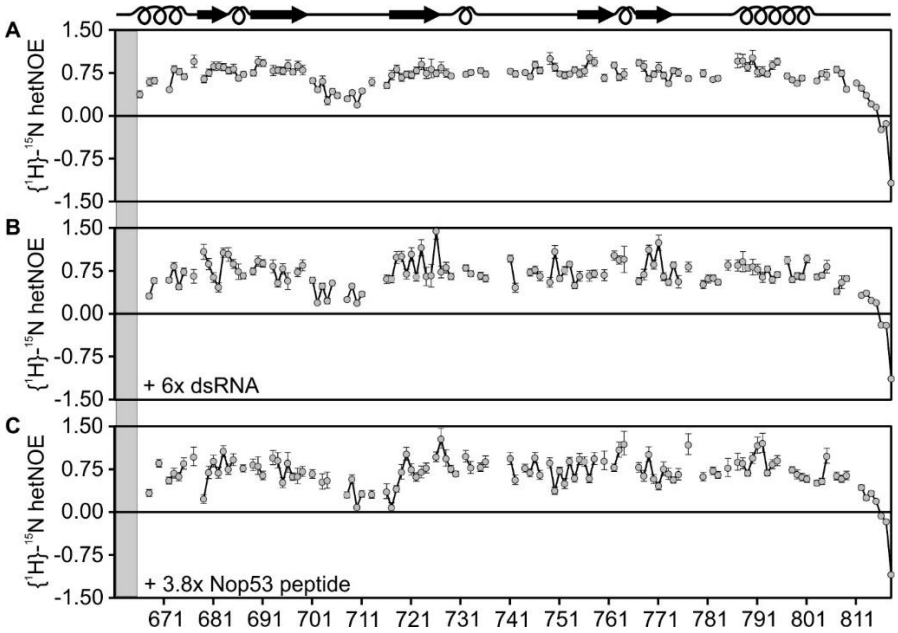



Supplementary Figure 6.6: Sequence of Rgs4 SRS2 RNA and $^{1}$H,$^{1}$H-NOESY. The RNA comprises an additional sequence to stabilize the duplex (blue box). The grey boxes indicate mismatches.

## 6.4 Mtr4 KOW




Supplementary Figure 6.7: Titrations of KOW domain with different ligands. $^{1}H^{15}N$-HSQC overlay of the free KOW domain (black) and in presence of the respective ligand (red). **A)** Titration with 10-fold dsRNA. **B)** Titration with 7.5-fold Phe-tRNA. **C)** Titration with 20-fold Nop53 peptide and **D)** with both 20-fold Nop53 peptide and 7-fold dsRNA.




Supplementary Figure 6.8: hetNOE data for KOW in complex with ligands. **A)** Free KOW domain and in presence of **B)** 6-fold dsRNA and **C)** 3.8-fold Nop53 peptide. No significant changes in local flexibility can be observed in presence of either ligand.

# 7 References


1. Wilson, R.C. & Doudna, J.A. Molecular mechanisms of RNA interference. *Annu Rev Biophys* **42**, 217-39 (2013).
2. Lee, R.C., Feinbaum, R.L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54 (1993).
3. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-11 (1998).
4. Egecioglu, D.E., Henras, A.K. & Chanfreau, G.F. Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. *RNA* **12**, 26-32 (2006).
5. Grivna, S.T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20**, 1709-14 (2006).
6. Carthew, R.W. & Sontheimer, E.J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642-55 (2009).
7. Hutvagner, G. & Simard, M.J. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**, 22-32 (2008).
8. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-60 (2004).
9. Borchert, G.M., Lanier, W. & Davidson, B.L. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**, 1097-101 (2006).
10. Lee, Y., Jeon, K., Lee, J.T., Kim, S. & Kim, V.N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-70 (2002).
11. Hayashita, Y. et al. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* **65**, 9628-32 (2005).
12. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. & Hannon, G.J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231-5 (2004).
13. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
14. Sohn, S.Y. et al. Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* **14**, 847-53 (2007).
15. Rolando, C. et al. Multipotency of Adult Hippocampal NSCs In Vivo Is Restricted by Drosha/NFIB. *Cell Stem Cell* **19**, 653-662 (2016).
16. Macias, S., Cordiner, R.A., Gautier, P., Plass, M. & Caceres, J.F. DGCR8 Acts as an Adaptor for the Exosome Complex to Degrade Double-Stranded Structured RNAs. *Mol Cell* **60**, 873-85 (2015).
17. Yeom, K.H., Lee, Y., Han, J., Suh, M.R. & Kim, V.N. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res* **34**, 4622-9 (2006).
18. Han, J. et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887-901 (2006).
19. Yi, R., Qin, Y., Macara, I.G. & Cullen, B.R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-6 (2003).

20. Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**, 447-59 (2013).
21. MacRae, I.J., Ma, E., Zhou, M., Robinson, C.V. & Doudna, J.A. In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci U S A* **105**, 512-7 (2008).
22. Chendrimada, T.P. et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-4 (2005).
23. Liang, C. et al. TAF11 Assembles the RISC Loading Complex to Enhance RNAi Efficiency. *Mol Cell* **59**, 807-18 (2015).
24. Nowek, K. et al. Expression of a passenger miR-9* predicts favorable outcome in adults with acute myeloid leukemia less than 60 years of age. *Leukemia* **30**, 303-9 (2016).
25. Bang, C. et al. Cardiac fibroblast-derived microRNA passenger strand-enriched exosomes mediate cardiomyocyte hypertrophy. *J Clin Invest* **124**, 2136-46 (2014).
26. Wen, J., Ladewig, E., Shenker, S., Mohammed, J. & Lai, E.C. Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates. *PLoS Comput Biol* **11**, e1004441 (2015).
27. Agranat-Tamir, L., Shomron, N., Sperling, J. & Sperling, R. Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. *Nucleic Acids Res* **42**, 4640-51 (2014).
28. Skalsky, R.L. & Cullen, B.R. Viruses, microRNAs, and host interactions. *Annu Rev Microbiol* **64**, 123-41 (2010).
29. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-33 (2009).
30. Chendrimada, T.P. et al. MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447**, 823-8 (2007).
31. Pillai, R.S. et al. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**, 1573-6 (2005).
32. Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* **20**, 515-24 (2006).
33. Ameres, S.L. & Zamore, P.D. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* **14**, 475-88 (2013).
34. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol* **3**, e85 (2005).
35. Yekta, S., Shih, I.H. & Bartel, D.P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594-6 (2004).
36. Hartig, J.V. & Forstemann, K. Loqs-PD and R2D2 define independent pathways for RISC generation in Drosophila. *Nucleic Acids Res* **39**, 3836-51 (2011).
37. Okamura, K. & Lai, E.C. Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* **9**, 673-8 (2008).
38. Ding, S.W. & Lu, R. Virus-derived siRNAs and piRNAs in immunity and pathogenesis. *Curr Opin Virol* **1**, 533-44 (2011).
39. McCue, A.D. et al. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J* **34**, 20-35 (2015).
40. Golden, D.E., Gerbasi, V.R. & Sontheimer, E.J. An inside job for siRNAs. *Mol Cell* **31**, 309-12 (2008).
41. Leuschner, P.J., Ameres, S.L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep* **7**, 314-20 (2006).

42. Schwarz, D.S. et al. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).
43. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J* **20**, 6877-88 (2001).
44. Yao, C., Sasaki, H.M., Ueda, T., Tomari, Y. & Tadakuma, H. Single-Molecule Analysis of the Target Cleavage Reaction by the Drosophila RNAi Enzyme Complex. *Mol Cell* **59**, 125-32 (2015).
45. Jacque, J.M., Triques, K. & Stevenson, M. Modulation of HIV-1 replication by RNA interference. *Nature* **418**, 435-8 (2002).
46. Wang, X.H. et al. RNA interference directs innate immunity against viruses in adult Drosophila. *Science* **312**, 452-4 (2006).
47. Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-103 (2007).
48. Aravin, A.A., Hannon, G.J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761-4 (2007).
49. Vagin, V.V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320-4 (2006).
50. Kitamura, S. et al. Characterization of RNase HII substrate recognition using RNase HII-argonaute chimaeric enzymes from Pyrococcus furiosus. *Biochem J* **426**, 337-44 (2010).
51. Nicholson, A.W. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA* **5**, 31-48 (2014).
52. Lee, Y.S. et al. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69-81 (2004).
53. Lau, P.W. et al. The molecular architecture of human Dicer. *Nat Struct Mol Biol* **19**, 436-40 (2012).
54. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11**, 576-7 (2004).
55. Macrae, I.J., Li, F., Zhou, K., Cande, W.Z. & Doudna, J.A. Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb Symp Quant Biol* **71**, 73-80 (2006).
56. Wostenberg, C. et al. The role of human Dicer-dsRBD in processing small regulatory RNAs. *PLoS One* **7**, e51829 (2012).
57. Ye, R. et al. A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis. *Mol Cell* **61**, 222-35 (2016).
58. Parker, J.S., Roe, S.M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J* **23**, 4727-37 (2004).
59. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**, 465-9 (2003).
60. Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E. & Weichenrieder, O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* **11**, 522-7 (2010).
61. Fu, Q. & Yuan, Y.A. Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). *Nucleic Acids Res* **41**, 3457-70 (2013).

62. Yamashita, S. et al. Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci* **20**, 118-30 (2011).
63. Chiliveri, S.C. & Deshmukh, M.V. Structure of RDE-4 dsRBDs and mutational studies provide insights into dsRNA recognition in the Caenorhabditis elegans RNAi pathway. *Biochem J* **458**, 119-30 (2014).
64. Tian, B., Bevilacqua, P.C., Diegelman-Parente, A. & Mathews, M.B. The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol* **5**, 1013-23 (2004).
65. Saunders, L.R. & Barber, G.N. The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J* **17**, 961-83 (2003).
66. Bycroft, M., Grunert, S., Murzin, A.G., Proctor, M. & St Johnston, D. NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J* **14**, 3563-71 (1995).
67. Ryter, J.M. & Schultz, S.C. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* **17**, 7505-13 (1998).
68. Donovan, J., Dufner, M. & Korennykh, A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc Natl Acad Sci U S A* **110**, 1652-7 (2013).
69. Barraud, P. et al. An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *EMBO J* **30**, 4223-35 (2011).
70. Wostenberg, C., Quarles, K.A. & Showalter, S.A. Dynamic origins of differential RNA binding function in two dsRBDs from the miRNA "microprocessor" complex. *Biochemistry* **49**, 10728-36 (2010).
71. Jammi, N.V. & Beal, P.A. Phosphorylation of the RNA-dependent protein kinase regulates its RNA-binding activity. *Nucleic Acids Res* **29**, 3020-9 (2001).
72. Bevilacqua, P.C. & Cech, T.R. Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry* **35**, 9983-94 (1996).
73. Masliah, G., Barraud, P. & Allain, F.H. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci* **70**, 1875-95 (2013).
74. Stefl, R. et al. The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**, 225-37 (2010).
75. Wang, X., Vukovic, L., Koh, H.R., Schulten, K. & Myong, S. Dynamic profiling of double-stranded RNA binding proteins. *Nucleic Acids Res* **43**, 7566-76 (2015).
76. Koh, H.R., Kidwell, M.A., Ragunathan, K., Doudna, J.A. & Myong, S. ATP-independent diffusion of double-stranded RNA binding proteins. *Proc Natl Acad Sci U S A* **110**, 151-6 (2013).
77. Liu, Q. et al. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**, 1921-5 (2003).
78. Ye, X., Paroo, Z. & Liu, Q. Functional anatomy of the Drosophila microRNA-generating enzyme. *J Biol Chem* **282**, 28373-8 (2007).
79. Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P.D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377-80 (2004).

80. Gredell, J.A., Dittmer, M.J., Wu, M., Chan, C. & Walton, S.P. Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry* **49**, 3148-55 (2010).
81. Sinha, N.K., Trettin, K.D., Aruscavage, P.J. & Bass, B.L. Drosophila dicer-2 cleavage is mediated by helicase- and dsRNA termini-dependent states that are modulated by Loquacious-PD. *Mol Cell* **58**, 406-17 (2015).
82. Zheng, X. & Bevilacqua, P.C. Straightening of bulged RNA by the double-stranded RNA-binding domain from the protein kinase PKR. *Proc Natl Acad Sci U S A* **97**, 14162-7 (2000).
83. Lim, M.Y. et al. The Drosophila Dicer-1 Partner Loquacious Enhances miRNA Processing from Hairpins with Unstable Structures at the Dicing Site. *Cell Rep* **15**, 1795-808 (2016).
84. Benoit, M.P. et al. The RNA-binding region of human TRBP interacts with microRNA precursors through two independent domains. *Nucleic Acids Res* **41**, 4241-52 (2013).
85. Wilson, R.C. et al. Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Mol Cell* **57**, 397-407 (2015).
86. Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H. & Siomi, M.C. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. *RNA* **16**, 506-15 (2010).
87. Saito, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol* **3**, e235 (2005).
88. Liu, X. et al. Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* **13**, 2324-9 (2007).
89. Czech, B. et al. An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**, 798-802 (2008).
90. Hartig, J.V., Esslinger, S., Bottcher, R., Saito, K. & Forstemann, K. Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J* **28**, 2932-44 (2009).
91. Zhou, R. et al. Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**, 1886-95 (2009).
92. Marques, J.T. et al. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. *Nat Struct Mol Biol* **17**, 24-30 (2010).
93. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**, 1655-66 (2004).
94. Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y. & Zamore, P.D. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**, 287-97 (2007).
95. Yigit, E. et al. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**, 747-57 (2006).
96. Iwasaki, S. et al. Defining fundamental steps in the assembly of the Drosophila RNAi enzyme complex. *Nature* **521**, 533-6 (2015).
97. Iwasaki, S. et al. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39**, 292-9 (2010).
98. Martin, K.C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719-30 (2009).
99. Berleth, T. et al. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *EMBO J* **7**, 1749-56 (1988).

100. Kugler, J.M. & Lasko, P. Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. *Fly (Austin)* **3**, 15-28 (2009).
101. Barr, J., Yakovlev, K.V., Shidlovskii, Y. & Schedl, P. Establishing and maintaining cell polarity with mRNA localization in Drosophila. *Bioessays* **38**, 244-53 (2016).
102. Jung, H., Yoon, B.C. & Holt, C.E. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat Rev Neurosci* **13**, 308-24 (2012).
103. Cody, N.A., Iampietro, C. & Lecuyer, E. The many functions of mRNA localization during normal development and disease: from pillar to post. *Wiley Interdiscip Rev Dev Biol* **2**, 781-96 (2013).
104. McDermott, S.M., Meignin, C., Rappsilber, J. & Davis, I. Drosophila Syncrip binds the gurken mRNA localisation signal and regulates localised transcripts during axis specification. *Biol Open* **1**, 488-97 (2012).
105. Chang, C.W. et al. Anterior-posterior axis specification in Drosophila oocytes: identification of novel bicoid and oskar mRNA localization factors. *Genetics* **188**, 883-96 (2011).
106. Ding, D., Parkhurst, S.M., Halsell, S.R. & Lipshitz, H.D. Dynamic Hsp83 RNA localization during Drosophila oogenesis and embryogenesis. *Mol Cell Biol* **13**, 3773-81 (1993).
107. Semotok, J.L. et al. Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. *Curr Biol* **15**, 284-94 (2005).
108. St Johnston, D. Moving messages: the intracellular localization of mRNAs. *Nat Rev Mol Cell Biol* **6**, 363-75 (2005).
109. Muslimov, I.A. et al. RNA transport in dendrites: a cis-acting targeting element is contained within neuronal BC1 RNA. *J Neurosci* **17**, 4722-33 (1997).
110. Jambhekar, A. & Derisi, J.L. Cis-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA* **13**, 625-42 (2007).
111. Tolino, M., Kohrmann, M. & Kiebler, M.A. RNA-binding proteins involved in RNA localization and their implications in neuronal diseases. *Eur J Neurosci* **35**, 1818-36 (2012).
112. Gagnon, J.A. & Mowry, K.L. Molecular motors: directing traffic during RNA localization. *Crit Rev Biochem Mol Biol* **46**, 229-39 (2011).
113. Marx, A., Hoenger, A. & Mandelkow, E. Structures of kinesin motor proteins. *Cell Motil Cytoskeleton* **66**, 958-66 (2009).
114. Heinrich, B. & Deshler, J.O. RNA localization to the Balbiani body in Xenopus oocytes is regulated by the energy state of the cell and is facilitated by kinesin II. *RNA* **15**, 524-36 (2009).
115. Shen, Z., Paquin, N., Forget, A. & Chartrand, P. Nuclear shuttling of She2p couples ASH1 mRNA localization to its translational repression by recruiting Loc1p and Puf6p. *Mol Biol Cell* **20**, 2265-75 (2009).
116. Bohl, F., Kruse, C., Frank, A., Ferring, D. & Jansen, R.P. She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. *EMBO J* **19**, 5514-24 (2000).

117. Deshler, J.O., Highett, M.I., Abramson, T. & Schnapp, B.J. A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. *Curr Biol* **8**, 489-96 (1998).
118. Irie, K. et al. The Khd1 protein, which has three KH RNA-binding motifs, is required for proper localization of ASH1 mRNA in yeast. *EMBO J* **21**, 1158-67 (2002).
119. Messias, A.C. & Sattler, M. Structural basis of single-stranded RNA recognition. *Acc Chem Res* **37**, 279-87 (2004).
120. Niessing, D., Huttelmaier, S., Zenklusen, D., Singer, R.H. & Burley, S.K. She2p is a novel RNA binding protein with a basic helical hairpin motif. *Cell* **119**, 491-502 (2004).
121. Kim, M., Bellini, M. & Ceman, S. Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. *Mol Cell Biol* **29**, 214-28 (2009).
122. Wolozin, B. Regulated protein aggregation: stress granules and neurodegeneration. *Mol Neurodegener* **7**, 56 (2012).
123. Liu-Yesucevitz, L. et al. Local RNA translation at the synapse and in disease. *J Neurosci* **31**, 16086-93 (2011).
124. Marion, R.M., Fortes, P., Beloso, A., Dotti, C. & Ortin, J. A human sequence homologue of Staufen is an RNA-binding protein that is associated with polysomes and localizes to the rough endoplasmic reticulum. *Mol Cell Biol* **19**, 2212-9 (1999).
125. Heraud-Farlow, J.E. & Kiebler, M.A. The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. *Trends Neurosci* **37**, 470-9 (2014).
126. Mallardo, M. et al. Isolation and characterization of Staufen-containing ribonucleoprotein particles from rat brain. *Proc Natl Acad Sci U S A* **100**, 2100-5 (2003).
127. Dubnau, J. et al. The staufen/pumilio pathway is involved in Drosophila long-term memory. *Curr Biol* **13**, 286-96 (2003).
128. Lebeau, G. et al. Staufen 2 regulates mGluR long-term depression and Map1b mRNA distribution in hippocampal neurons. *Learn Mem* **18**, 314-26 (2011).
129. Cockburn, D.M. et al. The double-stranded RNA-binding protein Staufen 2 regulates eye size. *Mol Cell Neurosci* **51**, 101-11 (2012).
130. Wickham, L., Duchaine, T., Luo, M., Nabi, I.R. & DesGroseillers, L. Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum. *Mol Cell Biol* **19**, 2220-30 (1999).
131. Monshausen, M. et al. Two rat brain staufen isoforms differentially bind RNA. *J Neurochem* **76**, 155-65 (2001).
132. Micklem, D.R., Adams, J., Grunert, S. & St Johnston, D. Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. *EMBO J* **19**, 1366-77 (2000).
133. Ramos, A. et al. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* **19**, 997-1009 (2000).
134. Fritzsche, R. et al. Interactome of two diverse RNA granules links mRNA localization to translational repression in neurons. *Cell Rep* **5**, 1749-62 (2013).
135. Ricci, E.P. et al. Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol* **21**, 26-35 (2014).
136. Laver, J.D. et al. Genome-wide analysis of Staufen-associated mRNAs identifies secondary structures that confer target specificity. *Nucleic Acids Res* **41**, 9438-60 (2013).

137. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-76 (2009).
138. Parker, R. & Song, H. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* **11**, 121-7 (2004).
139. Parker, R. RNA degradation in Saccharomyces cerevisae. *Genetics* **191**, 671-702 (2012).
140. Li, W.M., Barnes, T. & Lee, C.H. Endoribonucleases--enzymes gaining spotlight in mRNA metabolism. *FEBS J* **277**, 627-41 (2010).
141. Tomecki, R. & Dziembowski, A. Novel endoribonucleases as central players in various pathways of eukaryotic RNA metabolism. *RNA* **16**, 1692-724 (2010).
142. Johanson, T.M., Lew, A.M. & Chong, M.M. MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. *Open Biol* **3**, 130144 (2013).
143. Vasudevan, S. & Steitz, J.A. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* **128**, 1105-18 (2007).
144. She, M. et al. Structural basis of dcp2 recognition and activation by dcp1. *Mol Cell* **29**, 337-49 (2008).
145. Fromm, S.A. et al. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J* **31**, 279-90 (2012).
146. Meyer, S., Temme, C. & Wahle, E. Messenger RNA turnover in eukaryotes: pathways and enzymes. *Crit Rev Biochem Mol Biol* **39**, 197-216 (2004).
147. Chlebowski, A., Lubas, M., Jensen, T.H. & Dziembowski, A. RNA decay machines: the exosome. *Biochim Biophys Acta* **1829**, 552-60 (2013).
148. Wang, Z. & Kiledjian, M. Functional link between the mammalian exosome and mRNA decapping. *Cell* **107**, 751-62 (2001).
149. Liu, Q., Greimann, J.C. & Lima, C.D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223-37 (2006).
150. Schaeffer, D. et al. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol* **16**, 56-62 (2009).
151. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* **91**, 457-66 (1997).
152. Wasmuth, E.V. & Lima, C.D. Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Mol Cell* **48**, 133-44 (2012).
153. Coy, S., Volanakis, A., Shah, S. & Vasiljeva, L. The Sm complex is required for the processing of non-coding RNAs by the exosome. *PLoS One* **8**, e65606 (2013).
154. Rege, M. et al. Chromatin Dynamics and the RNA Exosome Function in Concert to Regulate Transcriptional Homeostasis. *Cell Rep* **13**, 1610-22 (2015).
155. Moraes, K.C. RNA surveillance: molecular approaches in transcript quality control and their implications in clinical diseases. *Mol Med* **16**, 53-68 (2010).
156. Houseley, J., LaCava, J. & Tollervey, D. RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* **7**, 529-39 (2006).
157. Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. & Jensen, T.H. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* **413**, 538-42 (2001).

158. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* **17**, 227-39 (2016).
159. Nag, A. & Steitz, J.A. Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing. *RNA Biol* **9**, 334-42 (2012).
160. Pan, K., Lee, J.T., Huang, Z. & Wong, C.M. Coupling and coordination in gene expression processes with pre-mRNA splicing. *Int J Mol Sci* **16**, 5682-96 (2015).
161. Bernard, P., Drogat, J., Dheur, S., Genier, S. & Javerzat, J.P. Splicing factor Spf30 assists exosome-mediated gene silencing in fission yeast. *Mol Cell Biol* **30**, 1145-57 (2010).
162. Lee, N.N. et al. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* **155**, 1061-74 (2013).
163. Marin-Vicente, C., Domingo-Prim, J., Eberle, A.B. & Visa, N. RRP6/EXOSC10 is required for the repair of DNA double-strand breaks by homologous recombination. *J Cell Sci* **128**, 1097-107 (2015).
164. Pefanis, E. & Basu, U. RNA Exosome Regulates AID DNA Mutator Activity in the B Cell Genome. *Adv Immunol* **127**, 257-308 (2015).
165. Pefanis, E. et al. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161**, 774-89 (2015).
166. Wyers, F. et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725-37 (2005).
167. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851-4 (2008).
168. Lubas, M. et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* **43**, 624-37 (2011).
169. Li, W., Notani, D. & Rosenfeld, M.G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**, 207-23 (2016).
170. van Hoof, A., Frischmeyer, P.A., Dietz, H.C. & Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262-4 (2002).
171. Mitchell, P. & Tollervey, D. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. *Mol Cell* **11**, 1405-13 (2003).
172. Lejeune, F., Li, X. & Maquat, L.E. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol Cell* **12**, 675-87 (2003).
173. Houseley, J., Kotovic, K., El Hage, A. & Tollervey, D. Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO J* **26**, 4996-5006 (2007).
174. Das, B., Butler, J.S. & Sherman, F. Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae. *Mol Cell Biol* **23**, 5502-15 (2003).
175. Wu, G.C. et al. Emerging role of long noncoding RNAs in autoimmune diseases. *Autoimmun Rev* **14**, 798-805 (2015).
176. Rigby, R.E. & Rehwinkel, J. RNA degradation in antiviral immunity and autoimmunity. *Trends Immunol* **36**, 179-88 (2015).
177. Hardwick, S.W. & Luisi, B.F. Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control. *RNA Biol* **10**, 56-70 (2013).

178. Py, B., Higgins, C.F., Krisch, H.M. & Carpousis, A.J. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **381**, 169-72 (1996).
179. Lebreton, A., Tomecki, R., Dziembowski, A. & Seraphin, B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* **456**, 993-6 (2008).
180. Han, J. & van Hoof, A. The RNA Exosome Channeling and Direct Access Conformations Have Distinct In Vivo Functions. *Cell Rep* **16**, 3348-58 (2016).
181. Liu, J.J. et al. Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM. *Nat Struct Mol Biol* **21**, 95-102 (2014).
182. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A. & Conti, E. Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol Cell* **29**, 717-28 (2008).
183. Synowsky, S.A., van den Heuvel, R.H., Mohammed, S., Pijnappel, P.W. & Heck, A.J. Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol Cell Proteomics* **5**, 1581-92 (2006).
184. Matoulkova, E., Michalova, E., Vojtesek, B. & Hrstka, R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol* **9**, 563-76 (2012).
185. Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* **40**, 4727-41 (2012).
186. Chen, C.Y. & Shyu, A.B. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci* **20**, 465-70 (1995).
187. Peng, S.S., Chen, C.Y., Xu, N. & Shyu, A.B. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J* **17**, 3461-70 (1998).
188. Apponi, L.H. et al. An interaction between two RNA binding proteins, Nab2 and Pub1, links mRNA processing/export and mRNA stability. *Mol Cell Biol* **27**, 6569-79 (2007).
189. Nicewonger, J., Suck, G., Bloch, D. & Swaminathan, S. Epstein-Barr virus (EBV) SM protein induces and recruits cellular Sp110b to stabilize mRNAs and enhance EBV lytic gene expression. *J Virol* **78**, 9412-22 (2004).
190. Bergman, N. et al. Lsm proteins bind and stabilize RNAs containing 5' poly(A) tracts. *Nat Struct Mol Biol* **14**, 824-31 (2007).
191. Kufel, J., Allmang, C., Verdone, L., Beggs, J.D. & Tollervey, D. Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Mol Cell Biol* **22**, 5248-56 (2002).
192. Tran, H., Schilling, M., Wirbelauer, C., Hess, D. & Nagamine, Y. Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol Cell* **13**, 101-11 (2004).
193. Zonta, E. et al. The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Res* **41**, 554-64 (2013).
194. Halbach, F., Rode, M. & Conti, E. The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA* **18**, 124-34 (2012).
195. Halbach, F., Reichelt, P., Rode, M. & Conti, E. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* **154**, 814-26 (2013).

196. Vanacova, S. et al. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* **3**, e189 (2005).
197. Holub, P. et al. Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Res* **40**, 5679-93 (2012).
198. Hamill, S., Wolin, S.L. & Reinisch, K.M. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc Natl Acad Sci U S A* **107**, 15045-50 (2010).
199. Weir, J.R., Bonneau, F., Hentschel, J. & Conti, E. Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc Natl Acad Sci U S A* **107**, 12139-44 (2010).
200. Falk, S. et al. The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol Cell* **55**, 856-67 (2014).
201. Houseley, J. & Tollervey, D. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep* **7**, 205-11 (2006).
202. Jia, H., Wang, X., Anderson, J.T. & Jankowsky, E. RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc Natl Acad Sci U S A* **109**, 7292-7 (2012).
203. Taylor, L.L. et al. The Mtr4 ratchet helix and arch domain both function to promote RNA unwinding. *Nucleic Acids Res* **42**, 13861-72 (2014).
204. Jia, H. et al. The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* **145**, 890-901 (2011).
205. Wasmuth, E.V., Januszyk, K. & Lima, C.D. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* **511**, 435-9 (2014).
206. Dickinson, H., Tretbar, S., Betat, H. & Morl, M. The TRAMP complex shows tRNA editing activity in S. cerevisiae. *Mol Biol Evol* **29**, 1451-9 (2012).
207. de la Cruz, J., Kressler, D., Tollervey, D. & Linder, P. Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. *EMBO J* **17**, 1128-40 (1998).
208. Zinder, J.C. & Lima, C.D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31**, 88-100 (2017).
209. Kyrpides, N.C., Woese, C.R. & Ouzounis, C.A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem Sci* **21**, 425-6 (1996).
210. Li, Y., Burclaff, J. & Anderson, J.T. Mutations in Mtr4 Structural Domains Reveal Their Important Role in Regulating tRNAiMet Turnover in Saccharomyces cerevisiae and Mtr4p Enzymatic Activities In Vitro. *PLoS One* **11**, e0148090 (2016).
211. Thoms, M. et al. The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* **162**, 1029-38 (2015).
212. Sasaki, M. et al. Regulation of the MDM2-P53 pathway and tumor growth by PICT1 via nucleolar RPL11. *Nat Med* **17**, 944-51 (2011).
213. Schuch, B. et al. The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J* **33**, 2829-46 (2014).
214. Cavanagh, J., Skelton, N.J., Fairbrother, W.J., Rance, M. & Palmer, A.G.I. *Protein NMR Spectroscopy*, 912 (Elsevier, 2006).
215. Milligan, J.F., Groebe, D.R., Witherell, G.W. & Uhlenbeck, O.C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783-98 (1987).

216. Dingley, A.J., Nisius, L., Cordier, F. & Grzesiek, S. Direct detection of N-H[...]N hydrogen bonds in biomolecules by NMR spectroscopy. *Nat Protoc* **3**, 242-8 (2008).
217. Clore, G.M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108-39 (2009).
218. Loth, K., Gnida, M., Romanuka, J., Kaptein, R. & Boelens, R. Sliding and target location of DNA-binding proteins: an NMR view of the lac repressor system. *J Biomol NMR* **56**, 41-9 (2013).
219. Jagla, B. et al. Sequence characteristics of functional siRNAs. *RNA* **11**, 864-72 (2005).
220. Segur, J.B. & Oberstar, H.E. Viscosity of glycerol and its aqueous solutions. *Industrial & Engineering Chemistry* **43**, 2117-2120 (1951).
221. Jakob, L. et al. Structural and functional insights into the fly microRNA biogenesis factor Loquacious. *RNA* **22**, 383-96 (2016).
222. Gleghorn, M.L., Gong, C., Kielkopf, C.L. & Maquat, L.E. Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat Struct Mol Biol* **20**, 515-24 (2013).

# 8 Publications and talks

## 8.1 Publications

During this thesis parts of this work have been published in the following papers and reviews:

Schlundt, A., Tants, J-N., Sattler, M. (2017) Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods*, 118-119: p. 119-136.

Tants, J-N.*, Fesser, S.*, Hartlmüller, C., Kern, T., Stehle, R., Geerlof, A., Wunderlich, C., Böttcher, R., Kunzelmann, S., Kreutz, C., Förstemann, K., Sattler, M. (2017) Molecular basis for asymmetry sensing of siRNAs by the *Drosophila* Loqs-PD / Dcr-2 complex in RNA interference. *Nucleic Acids Research*, 45: p. 12536-12559.

Falk, S., Tants, J-N., Hurt, E., Sattler, M., Conti, E. (2017) Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the pre-ribosomal protein Nop53. *RNA*, 23: p. 1780-1787.

*contributed equally to this work

## 8.2 Talks

Parts of the work have been presented by me in talks at the following conferences:

2016 RNA structure meets function, EMBO workshop, Stockholm, Sweden

2015 RNA, RNA Society, Madison, USA

2014 NMR in structural biology of RNPs, Parpan, Switzerland

# Acknowledgements

Hereby I want to thank the following people, who either contributed scientifically to this work, made the last years memorable or supported me in any other way:

- Prof. Dr. Michael Sattler who gave me the opportunity to do my PhD in his lab and provided the interesting and challenging projects. I liked the scientific discussion that were always respectful, productive and gave me the feeling of being taken serious even though some of my ideas and proposals might have been science fiction. He always encouraged me to dig a bit deeper and pushed hard to improve also my skills outside science. Thanks to him I was able to participate in different conferences and present my results. Furthermore, he gave me a lot of freedom to pursue the projects and develop my own ideas.
- Thesis Committee and Thesis Advisory Committee members Prof. Dr. Michael Sattler, Prof. Dr. Klaus Förstemann and Prof. Dr. Kathrin Lang who always gave helpful and constructive feedback. The atmosphere during all our meetings was always nice and relaxed and allowed fruitful discussions.
- My collaborators Prof. Dr. Klaus Förstemann, Dr. Stephanie Fesser and Romy Böttcher (Loqs-PD), Prof. Dr. Christoph Kreutz and Dr. Christoph Wunderlich (spin labelled RNA), Prof. Dr. Dierk Niessing and Simone Heber (Staufen-2), Prof. Dr. Elena Conti and Dr. Sebastian Falk (Mtr4 KOW) for the scientific exchange, good collaboration, patience and opportunity to join these really interesting projects.
- Dr. Arie Geerlof and Dr. Ralf Stehle for SLS and SAXS measurements and analysis.
- My former supervisor Dr. Thomas Kern who handed over the Loqs-PD project to me and introduced me to both NMR and the group; Dr. Andreas Schlundt, Dr. Leonidas Emmanouilidis and Johannes Günther for scientific discussions and help whenever needed, be it at the spectrometer, in the lab or in theory.
- PD Dr. Gerd Gemmecker and Dr. Rainer Haessner for answering NMR or spectrometer related questions and for keeping both spectrometer and IT running; the secretary Waltraud Wolfson for support concerning all organisatorial issues and Gülden Yilmaz for technical support in the lab.
- All internship students I ever supervised and who contributed to lab work.

- Andreas Schlundt, Leonidas Emmanouilidis, Johannes Günther, Gülden Yilmaz (and several other Sattler group members) for fun time in both the lab and office.
- My parents and brother for support throughout the years.
- And especially Sarah, who supported me morally and kept my back free whenever needed; for her patience and understanding. And: For just being there.