THE ROLE OF DSRBDS IN THE MIRNA MATURATION PATHWAY

A Dissertation in
Chemistry
by
Christopher William Wostenberg

© 2012 Christopher William Wostenberg

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2012
The dissertation of Christopher William Wostenberg was reviewed and approved by the following:

Scott A. Showalter
Assistant Professor of Chemistry
Dissertation Adviser
Chair of Committee

Philip C. Bevilacqua
Professor of Chemistry

David Boehr
Assistant Professor of Chemistry

Andrey Krasilnikov
Associate Professor of Biochemistry and Molecular Biology

Barbara J. Garrison
Shapiro Professor of Chemistry
Head of the Chemistry Department

*Signatures are on file in the Graduate School.
ABSTRACT

Macromolecular interactions are involved in important biological processes like transcription, translation, signaling, and defense. In order to understand these biochemical processes it is vital to comprehend the mechanism of macromolecular interactions via two key characteristics: structure and dynamics. I chose to investigate fast dynamics on two systems which were already structurally characterized: the processing of microRNAs (miRNAs) precursors by dsRNA binding domains (dsRBDs) in humans and the intrinsically disordered protein FCP1 binding to RAP74, which is involved in termination of transcription and recycling of RNA polymerase II.

miRNAs, a large class of small ssRNAs, are involved in gene regulation by base-pairing with messenger RNA (mRNA). Maturation of miRNAs occurs in two independent and spatially separated steps. In the nucleus, the ssRNA tail of primary miRNA is cleaved by the "microprocessor", which is composed of a RNase III enzyme (Drosha) and its cofactor (DGCR8). Next, the precursor miRNA is exported to the cytosol, and the terminal loop is cleaved by another RNase III enzyme (Dicer), which is aided by a cofactor (either TRBP or PACT). Afterwards, the miRNA is loaded into the argonaute (Ago) protein generating the RNA-induced silencing complex (RISC). Maturation of miRNA is accomplished by total of ten dsRBDs over five proteins in the pathway. This dissertation specifically focuses on understanding the fast dynamics of four of them: DGCR8-dsRBD1, DGCR8-dsRBD2, Drosha-dsRBD and Dicer-dsRBD.

The dsRBD is one of the most common RNA-binding motifs, found in all organisms and in both the cytoplasm and the nucleus. The dsRBD forms an αββα
topology, where the two α-helices lie on one face of three anti-parallel β-strands, with the other surface of the sheet being solvent exposed. The preferred binding partner of dsRBDS is A-form double-helix RNA, where a minimum of 11 base pairs of dsRNA interact with a dsRBD generally without sequence specific RNA interactions.

A dynamic profile was obtained for the dsRBDS through MD simulations and NMR spin relaxation along with binding assays to determine functionality of the dsRBDS. To start, the two dsRBDS in tandem from DGCR8 were explored through MD simulations (chapter 2), motivated by the pre-organization observed in crystal and verified by preliminary NMR data. The results demonstrate that correlated motions impacting the conformation of the RNA-binding surface are mediated by a few interfacial interactions. Next, MD simulations in connection with NMR spin relaxation experiments were performed on Drosha-dsRBD, DGCR8-dsRBD1, and Dicer-dsRBD (chapter 3 and 4) revealing a dynamic profile where loop 2 is dynamic in all dsRBDS. Electrophoretic mobility shift assays (EMSAs) of the dsRBDS showed that isolated Drosha-dsRBD is the only dsRBD studied that does not bind dsRNA, which is attributed to Drosha-dsRBD having increased flexibility in loop 1. Together, the data illustrates that flexibility in loop 2 of the dsRBDS allows for binding of dsRNA, but flexibility in loop 1 hinders binding.

To further study macromolecular interactions via MD simulations, the intrinsically disordered C-terminal domain of FCP1 bound to the winged helix domain of RAP74 was investigated. Unlike ordered protein complexes, which involve polar interactions over a small fraction of the protein surface, disordered proteins complexes like the RAP74-FCP1 complex involve large hydrophobic interactions. Additionally, FCP1 in the complex retains significant flexibility throughout the simulation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBREVIATIONS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF EQUATIONS</td>
<td>xvi</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>xvii</td>
</tr>
<tr>
<td>Chapter 1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2 MD Simulations of the dsRBP DGCR8 Reveal Correlated Motions that May Aid pri-miRNA Binding</td>
<td>44</td>
</tr>
<tr>
<td>Chapter 3 Dynamic Origins of Differential RNA Binding Function in Two dsRBDs from the miRNA “Microprocessor” Complex</td>
<td>77</td>
</tr>
<tr>
<td>Chapter 4 The Role of the Dicer-dsRBD in Isolation on Processing Small Regulatory RNAs</td>
<td>115</td>
</tr>
<tr>
<td>Chapter 5 Future Directions for Studying dsRBDS in the miRNA Maturation Pathway</td>
<td>148</td>
</tr>
<tr>
<td>Chapter 6 Atomistic Simulations Reveal Structural Disorder in the RAP74-FCP1 Complex</td>
<td>162</td>
</tr>
</tbody>
</table>
ABBREVIATIONS

Ago: argonaute

ANM: anharmonic normal model

AUC: analytical ultracentrifugation

B#: β strand #

bp: base-pairs

CD: circular dichroism

centFCP1: central domain of FCP1

CK2: Casein Kinase 2

CPMG: Carr-Purcell-Meiboom-Gill

CSA: chemical shift anisotropy

cterFCP1: C-terminal domain of FCP1

DGCR8: DiGeorge critical region 8

DLS: dynamic light scattering

DNA: deoxyribonucleic acid

DOSY: diffusion-ordered 2D NMR spectroscopy

ds: double-stranded

dsRBDs: dsRNA binding domains

dsRBPs: dsRNA binding proteins

DUF: domain of unknown function

EMSAs: electrophoretic mobility shift assays
FCP1: TFIIF-associating CTD Phosphatase

ΔG: free energy

ΔΔG: change in free energy

ΔG_c: cooperative free energy

GST: glutathione S-transferase

H#: α helix #

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HSQC: heteronuclear single quantumcoherence

HTH: helix-turn-helix

IDE: isotropically distributed ensemble

IDPs: intrinsically disordered proteins

IDRs: intrinsically disordered regions

IDT: Integrated DNA Technologies

IPTG: isopropyl β-D-1-thiogalactopyranoside

iRED: isotropic reorientational eigenmode dynamics

ITC: isothermal titration calorimetry

K_d: dissociation constant

L#: loop #

MD: molecular dynamics

miRNA: microRNA

mRNA: messenger RNA
**MWCO:** molecular weight cutoff

**Ni-NTA:** nickel-nitriloacetic acid

**NMR:** nuclear magnetic resonance

**NOE:** Nuclear Overhauser Effect

**nt:** nucleotides

**NTPs:** nucleoside triphosphates

**PES:** polyethersulfone

**PCA:** principal component analysis

**PCR:** polymerase chain reaction

**PDB:** protein data bank (structure format)

**PKR:** protein kinase R

**PolII:** RNA polymerase II

**pre-miRNA:** precursor miRNA

**pri-miRNA:** primary miRNA

**R₁:** longitudinal relaxation rate \((R₁ = 1/T₁)\)

**R₂:** transverse relaxation rate \((R₂ = 1/T₂)\)

**R_{ex}:** exchange broadening relaxation rate

**RAP30:** RNA polymerase II-associated protein 30

**RAP74:** RNA polymerase II-associated protein 74

**RDE-4:** RNAi defective-4

**RISC:** RNA-induced silencing complex
**RMSD**: root-mean-square deviation

**RMSF**: root-mean-square fluctuations

**RNA**: ribonucleic acid

**RNAi**: RNA interference

**RRMs**: RNA recognition motif

**$S\#$**: β strand #

**$S^2$**: order parameter

**SASA**: solvent accessible surface area

**SH2**: Src homology 2

**siRNA**: small-interfering RNA

**ss**: single-stranded

**$\tau_e$**: internal motion correlation time

**$\tau_m$**: overall motion correlation time

**TBE**: Tris/Borate/EDTA

**TEN$_{250}$**: 10 mMTris, 1 mM EDTA, 250mM NaCl

**TF**: transcription factor

**TFE**: 2,2,2-trifluoroethanol

**TMAO**: trimethylamine N-oxide

**WH**: winged-helix
LISTS OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Canonical dsRBD Structure and Interactions with dsRNA</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Protein Alignment of Various dsRBDs</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>miRNA Maturation Pathway</td>
<td>13</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Cartoon Representation and Tertiary Structure of DGCR8-Core</td>
<td>47</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>RMSDs and PDBs of the dsRBDs from DGCR8</td>
<td>52</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Ribbon Diagram Representation of ANM Analysis of DGCR8-Core</td>
<td>56</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Cα Correlations of the dsRBDs from DGCR8</td>
<td>59</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>IDE Analysis of DGCR8-Core</td>
<td>62</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Ribbon Diagram Representation of IDE Analysis of DGCR8-Core</td>
<td>64</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Interfacial Interactions of DGCR-Core</td>
<td>66</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Alignment of a Select Few dsRBDs</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Cartoon Representation and Tertiary Structure of Drosha-dsRBD and DGCR8-dsRBD1</td>
<td>83</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>EMSA of Pri-miR-16-1 Binding to DGCR-dsRBD1</td>
<td>90</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>$^{15}$N Spin Relaxation of Drosha-dsRBD and DGCR8-dsRBD1</td>
<td>92</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Order Parameters of Drosha-dsRBD and DGCR8-dsRBD1</td>
<td>94</td>
</tr>
</tbody>
</table>
Figure 3.6: RMSD of Drosha-dsRBD and DGCR8-dsRBD1

Figure 3.7: Cα Correlations of Drosha-dsRBD and DGCR8-dsRBD1

Figure 4.1: Cartoon Representation, Tertiary Structure, and Primary Sequence of Dicer-dsRBD

Figure 4.2: EMSA of Pre-miR-16-1 and ds44 Binding by Dicer-dsRBD

Figure 4.3: EMSA of ds16-tetra and ds16-octa Binding by Dicer-dsRBD

Figure 4.4: $^{15}$N Spin Relaxation of Dicer-dsRBD

Figure 4.5: Order Parameters and RMSD of Dicer-dsRBD

Figure 4.6: Cα Correlations of Dicer-dsRBD

Figure 5.1: Thermodynamic Cycles

Figure 6.1: Sequence and Tertiary Complex of RAP74 and FCP1

Figure 6.2: RMSD, RMSF, and PDBS of apo-RAP74 and RAP74-FCP1

Figure 6.3: Difference in SASA of RAP74 and FCP1 When Bound

Figure 6.4: Dynamic Motion of Met-949

Figure 6.5: N-terminal turn of Helix 3 in apo-RAP74 Instability

Figure 6.6: Cα Correlation of apo-RAP74 and RAP74-FCP1

Figure 6.7: Radius of Gyration of FCP1
Figure A.1: EMSA of Pri-miR-16-1 with Drosha-dsRBD 207

Figure A.2: Different Fits of EMSA of Pri-miR-16-1 with DGCR8-dsRBD1 208

Figure A.3: $^{15}$N HSQC spectra of Drosha-dsRBD and DGCR8-dsRBD1 209

Figure B.1: Representation of Transcription Construct to Obtain
Pre-miR-16-1 216

Figure B.2: AUC data of Dicer-dsRBD binding ds16 217

Figure B.3: $^{15}$N HSQC spectrum of Dicer-dsRBD 218

Figure B.4: Dicer-dsRBD DLS data at NMR concentration 219

Figure B.5: NMR diffusion of Dicer-dsRBD 220

Figure C.1: Dynamic Motion of Leu-953 228

Figure C.2: Dynamic Motion of Leu-957 229

Figure C.3: Dynamic Motion of Leu-960 230

Figure C.4: Dynamic Motion of Met-961 231

Figure C.5: Changes in Solvent Accessible Surface Area (SASA) of RAP74 upon binding FCP1. 233

Figure C.6: Alignment of FCP1 in RAP74 Binding Pocket 234

Figure C.7: Characterization of RAP74 H2 in the apo and holo state 235
Figure C.8: Packing of RAP74 Hydrophobic Groove the apo and holo state 236

Figure C.9: Location and Fluctuations of H2.5 in RAP74 the apo and holo state 237

Figure C.10: Orientation of the Helices from RAP74 the apo and holo state 238

Figure C.11: Changes in H1 from RAP74 Structure and Orientation resulting from binding FCP1 239

Figure C.12: Orientation of secondary structures in RAP74 before and after 2 binding FCP1 240

Figure C.13: Angle of Successive Helix Directors Along N-terminal of FCP1 241

Figure C.14: Angle of Successive Helix Directors Along C-terminal of FCP1 242

Figure C.15: Distance Between Successive Turns Along N-terminal of FCP1 243

Figure C.16: Distance Between Successive Turns Along N-terminal of FCP1 244

Figure C.17: Dipole Moment for FCP1 Backbone 245
LISTS OF TABLES

Table 4.1: Binding Affinity of Dicer-dsRBD for Various RNA Constructs by EMSA

Table A.1: Chemical Shifts (ppm) from the Backbone Residues of Drosha-dsRBD

Table A.2: Chemical Shifts (ppm) from the Backbone Residues of DGCR8-dsRBD1

Table B.1: Chemical Shifts (ppm) from the Backbone Residues of Dicer-dsRBD

Table C.1: MD Average Distances for Key Intermolecular Interactions
LISTS OF EQUATIONS

Equation 1.1: General Hill model for fitting binding data  19
Equation 1.2: Fully cooperative fitting equation for binding data  19
Equation 1.3: N-independent and identical binding site model for fitting binding data  20
Equation 1.4: General form of the covariance matrix  28
Equation 1.5: IDE form of the covariance matrix  29
Equation 1.6: iRED form of the covariance matrix  30
Equation 1.7: Order parameter calculation from iRED covariance matrix  31
Equation 5.1: Relationship Between Free Energy and Dissociation Constant  150
Equation B.1: NMR diffusion fitting equation  215
ACKNOWLEDGEMENTS

This dissertation would not be possible without the help and support of so many people, so I would like to start with a general thank you to everyone and I apologize if I fail to mention anyone specifically. This general thank you extends to all the teachers I have had throughout my academic career.

I would like to start by thanking my Ph.D. advisor Scott Showalter, who was kind enough to take a chance and accept a very lost and discouraged organic chemistry student as his first Ph.D. student. Scott helped to rekindle my passion for science while teaching me to be a better all around chemist. In the process, I have learned much about the ins and outs of being a professor and through this process has fostered my desire to become a professor in the near future. Also, I would like to thank all the present and former lab members of the Showalter lab who have been like an extended family to me as they have helped me learn and grow as a chemist while putting up with all my idiosyncrasies: Alain Bonny, Chad Lawrence, Kaycee Quarles, Rico Acevedo, Monique Bastidas, Brittney Nagle, Debasish Sahu, Ellen Forsyth, and Eric Gibbs.

Next, I would like to acknowledge my thesis committee, Phil Bevilacqua, David Boehr, and Andrey Krasilnikov, who have provided guidance in the form of comments and suggestions throughout my graduate career. Also, I would like to thank the labs of Phil Bevilacqua, David Boehr, Craig Cameron, Will Noid, and Booker Squire who at one time or another have provided helpful advice or assistance with experiments. A special thanks goes out to the office staff in the chemistry department, notably, Dana Coval-Dinant, Tina Wills, Tasha Marasco, and Sabrina Glasgow, for making sure I got all my
paperwork in on time and answering any administrative questions in a friendly and
courteous manner.

Next, I would like to thank my undergraduate advisor, Eric Marinez, and the
Marinez lab, who are responsible for my motivation in attending graduate school and
getting a Ph.D. in chemistry. Also, I would like to thank the entire faculty at CSULB,
especially the departments of Chemistry and Biochemistry and Mathematics, who
prepared me for graduate school.

I would be remiss if I failed to mention my friends at Penn State who have kept
me sane throughout the experience and provided a great divergence from science at
times. This includes, but is not limited to Jenay Robert, Ben Smith, Dan Dewey, Phil
Peterman, Kristin Cederquist, Rico and Jess Acevedo, Ran Tu, and Allen Chao. Also
thank you for keeping me well fed through all the dinner parties. And special thanks to
my fiancée, Darlene Biziak, who has definitely kept me sane and going these last few
years of graduate school. Also, she has been around when I needed a good laugh or
someone to complain to without asking for anything in return.

Last but not least, I would like to end by thanking my family, who have been
supportive and encouraging even though they are a country apart. They have always
been there when I needed them and were understanding of my busy schedule that has
made me unavailable at times.

Again I apologize if I have forgotten to specifically acknowledge someone.
Thanks again to everyone who has helped me along the way.

“A mind is like a parachute. It doesn’t work if it is not open.” Frank Zappa
Chapter 1

Introduction

1.1 Overview of Macromolecular Interactions:

Biochemistry is dominated by macromolecular interactions. A macromolecular interaction produces a complex that can be between two like species (e.g., protein-protein interactions) as exemplified by the transcription pre-initiation complex; or between two different classes of macromolecules (e.g., protein-nucleic acid) as exemplified by a RNase enzyme interacting with its target RNA. Incorrect interactions can lead to disease states like Alzheimer’s disease, caused by the aggregation of amyloid, and autoimmune diseases, caused by the misidentification of self cells. Therefore, in order to understand vital biochemical processes and disease states, it is critical to understand the underlying macromolecular interactions that comprise them.

Macromolecular interactions can be understood by studying two key characteristics, structure and dynamics. Structure is determined by knowing the average location of each atom and from this the forces between atoms is inferred. Dynamics describe how atoms move in time, which is separated into two regimes: the slow timescale and the fast timescale. In general, structural information is a prerequisite for obtaining dynamic information on a system, thus both together are critical for understanding macromolecular interactions. With that said, this thesis focuses on understanding the dynamics on the fast timescale for a few systems with structural information already determined. Therefore, the introduction will cover an overview of dynamics as they relate to biological systems. Then, the two model
systems for understanding macromolecular interactions will be introduced. Finally, the methods used for studying the dynamics of these systems will be discussed.

1.2 Overview of Dynamics:

Originally macromolecules, specifically proteins, were thought to be rigid bodies leading to the structure dictates function paradigm together with the lock-and-key model for macromolecular interactions; therefore, many structures were determined. With the advent of molecular dynamics (MD) simulations and NMR relaxation experiments it was determined that macromolecules are more flexible than originally thought. Due to the intrinsic flexibility of some macromolecules, the induced fit model was proposed, where macromolecules go through conformational changes upon binding, leading to a focus to not only understand structure, but dynamics as well.

1.2.1 Slow Dynamics:

Slow dynamics, lifetimes slower than a microsecond, are considered on the biological timescale because important biological functions like folding/unfolding of a macromolecule, diffusion of a macromolecule, ligand binding and catalysis, happen in the microsecond to second range. Slow dynamics reflect fluctuations between kinetically distinct states that are separated by high energy barriers (greater than the product of the Boltzmann constant and the absolute temperature)\(^1\) and correspond with conformational changes of the macromolecule. The transitions between the states are rare, leading to long lifetimes which allow for both information on the structure of the state and the rate of interconversion between states to be determined experimentally. At the current time, slow dynamics are outside the practical limits of MD simulations.
1.2.2 Fast Dynamics:

In the previous section, slow dynamics were equated with biological function, but it is also important to understand faster dynamics that have sub-microsecond characteristic lifetimes as they can mediate the slower processes. Typically, fast dynamics represent local flexibility of a macromolecule, thus the states are structurally similar with only fluctuations of certain regions of the macromolecule, like loops and side chains of proteins. Common motions on the fast timescale include bond vibrations, rotation, molecular tumbling and proton transfer. The energy barrier between the states are low (less than the product of the Boltzmann constant and the absolute temperature),\(^1\) allowing for a large number of sub-states to be sampled.

Fast dynamics being in the sub-microsecond regime makes it impossible to obtain structures of the sub-states because the timescale is faster than the detection rate. Instead, it is possible to know what atoms or regions of a macromolecule are moving and how fast compared with other atoms or regions. One way is through NMR spin relaxation, which provides information on the amplitude of atomic fluctuations, via the order parameter \((S^2)\), and the timescale, via the internal correlation time (for more details see Section 1.5.2).\(^1\) It is important to note that the order parameter is derived without reference to a specific structure and is on a normalized scale (with 1 being a completely rigid site and 0 being a freely rotating site) making it a measure of flexibility that is transferrable between compared macromolecules – a property that will be utilized extensively in this thesis. Computationally, MD simulations report on fast dynamics and they can reproduce experimental data, like order parameters,\(^2\) thus providing a complement with experiments. All atom MD simulations allow for structural information
on the sub-states; therefore, information on the exact motions of the atoms can be visualized. In addition, MD simulations also provide information on correlated motions of distal atoms, which is also difficult to obtain experimentally. It is these strengths of MD simulations that will be heavily utilized throughout this thesis.

1.2.3 Correlation Between Dynamics and Function:

1.2.3.1 Trp Repressor Regulating Different Operons:

The degree of flexibility is critical for the ability of a protein to bind various targets, as illustrated in the trp repressor from E. coli. The trp repressor regulates several different operator sequences, requiring it to be a versatile DNA binder, and is itself regulated by the concentration of its cofactor, L-Trp, which causes dimerization. NMR spin relaxation of two mutants have revealed that the degree of flexibility in the helix-turn-helix (HTH) DNA binding motif is critical for the ability of the trp repressor to bind various targets. First, an A77V mutant has decreased flexibility in the HTH motif for the apo-state as compared with the wild-type repressor. Upon binding L-Trp, the flexibility of the HTH motif in the wild-type repressor is reduced, but no reduction in flexibility is observed in the A77V mutant. Functionally, this makes the A77V mutant a super repressor of the trp operator but a poor repressor of other operators. Next, the L75F mutant was shown to have increased flexibility in the HTH motif of the apo-state leading to a decrease in binding affinity for the cofactor, L-Trp. Therefore, increased flexibility reduces the trp repressor from binding any operator, while decreased dynamics reduces its ability from binding multiple operators.
1.2.3.2 Differences Between SH2 Domain Dynamics:

Even through the structure of macromolecules might be the same, this does entail that the dynamics are. The Forman-Kay group investigated the dynamic profiles of the two structurally similar SH2 domains from the Syp tyrosine phosphatase and phospholipase Cγ1, which have similar binding affinity for their respective targets. Syp tyrosine phosphatase has slightly lower than average order parameters for residues involved in peptide binding, which increase (become more rigid) upon binding of its target peptide. In contrast, phospholipase Cγ1 has a highly dynamic interface that does not change when binding. The similarity in binding affinities together with the differences in the dynamics between the structurally similar domains imply that a different mechanism is employed to regulate their function.

1.2.3.3 U1A Binding ssRNA:

Finally, MD simulations together with NMR spectroscopy can be utilized to determine correlated motions. The N-terminal RNA recognition motif (RRM) of the U1A protein, which is a component of the spliceosome, has been shown to have a large network of correlated motions that facilitate an induced fit mechanism. It has been shown that loop 3, which directly interacts with RNA, is quite flexible in the free state, while it is suggested that its motions become restricted in the bound state. MD simulations revealed a network of correlated motions between loop 1 and loop 6 with loop 3 mediated through Gly53, which when mutated decreases the binding ability of U1A for RNA illustrating the importance of these correlated motions. The mutagenesis study led to the hypothesis that the correlated motions observed in MD simulations are involved in facilitating a conformational change upon RNA binding. All
of the studies mentioned so far were on the apo-state of U1A, but the hypothesis of correlated motions facilitating binding was reinforced via MD simulations of the complex.\textsuperscript{13}

1.3 Protein-RNA Interactions:

RNA-protein interactions are vital for many cellular processes including transcription, translation, gene regulation and viral defense. RNA is a long polyanionic structure that is highly flexible compared with structured proteins. Normally RNA is single-stranded, but often adopts secondary structural elements and when doing so forms an A-form helix. The A-form helix has a deep, narrow major groove that makes the bases inaccessible for interactions with proteins. Therefore, most protein interactions with dsRNA are non-sequence specific, but instead are rely on electrostatic interactions with the RNA backbone. Several protein domain folds as well as amino acid clusters have been shown to interact with either ssRNA or dsRNA,\textsuperscript{14} this dissertation focuses on one of them, the dsRNA binding domain (dsRBD).

1.3.1 Double-Stranded RNA Binding Domains (dsRBDs):

dsRBDs are found in proteins throughout the cell as well as in all forms of life,\textsuperscript{15} making them one of the most common RNA-binding motifs. They can exist as a single copy or as multiple copies, usually in tandem, within a polypeptide chain. Quite a few structures exist of various dsRBDs in either the free or bound state. In the minimal form, the dsRBD consists of roughly 67 amino acid residues,\textsuperscript{16-18} with a αββα canonical fold, where the two α-helices on are the same side of three antiparallel β-sheets (Figure 1.1A).\textsuperscript{17} The RNA binds along the α helical face of the domain leaving the other face of the β-sheet solvent exposed (Figure 1.1B).\textsuperscript{16-18}
Figure 1.1: Canonical dsRBD structure and interactions with dsRNA. A ribbon diagram depicting the tertiary structure of a canonical dsRBD (Staufen-dsRBD3 bound to a dsRNA PDB: 1EKZ) showing the αβββα topology with the secondary elements labeled (A) and the binding interactions with a model RNA (B). For clarity, the residues that interact with the RNA have been color coded as follows: glutamate – yellow, histidine – red and lysine – green.
As the name implies, dsRBDs preferentially bind dsRNA over ssRNA, dsDNA, and DNA-RNA hybrids, but it has also been shown to be involved in protein-protein interactions. As little as one turn of canonical A-form RNA (~11 base pairs) has been shown to interact with one dsRBD, but in most cases the dsRBD spans 16 base-pairs along one face of the dsRNA. The interaction of dsRBDs with dsRNA tend to be non-sequence specific (a counter example being ADAR2), but some evidence exists that RNA structure, most notably internal loops/bulges and terminal loops, might play an important part in the dsRBD recognition of its target RNA. The interactions involve helix 1 and loop 2 interacting with the minor groove, through hydrogen bonds with the 2’ hydroxyl groups on the RNA, and helix 2 interacting with the major groove, through either direct or water mediated contacts with the phosphate backbone of the RNA. From an alignment study, three well conserved lysine or arginine residues in helix 2 (Figure 1.2 residues highlighted green) are responsible for the interactions with the phosphate backbone and have been shown to be critical for binding dsRNA. Structural data along with the alignment, suggest that the His residue in loop 2 and the His/Gln/Asn/Ser followed by the Glu sequence of helix 1 (Figure 1.2 residues highlighted red) are important for minor groove interactions, but no mutagenesis work exists to support this claim. A region of high conservation, outside of the binding interface, is an aromatic residue (Figure 1.2 residue highlighted yellow) hypothesized to maintain the distance between loop 2 and loop 4/beginning of helix 2 to match the spacing between the major and minor groove of A-form RNA. The structural information of both the apo- and holo-state of dsRBDs along with biochemical assays have provided a basic understanding on how dsRBDs interact with
Figure 1.2: Protein alignment of various dsRBDs. Alignment using ClustalW2 for the ten dsRBDs from the miRNA maturation pathway (below blue line) with five canonical dsRBDs from the literature (above blue line) reveal conserved amino acids. The conserved aromatic residue is highlighted in yellow. The conserved lysine/arginine residues in helix 2 that interacts with phosphate backbone are highlighted green. The conserved residues in helix 1 and loop2 that interact with the 2’ hydroxyl groups on the RNA are highlighted in red. The rest of the shaded positions moving from gray to black indicate increasing sequence conservation. Approximate locations of secondary structural elements within the sequence are depicted above the alignment.
dsRNA, but these data fail at contributing dynamic information to the mechanism of dsRBD binding. As mentioned earlier, one way to experimentally determine fast timescale dynamics is through NMR spin relaxation, which can provide flexibility on a per-residue basis. Only a couple of cases exist of NMR spin relaxation being utilized to determine dynamic information on dsRBDs. In both cases, loop 2 – the largest loop – is the most dynamic region of the domain, even in the bound state. Order parameters from the apo-state of the two dsRBDs in the antiviral protein, PKR, demonstrates that decreased flexibility in loop 2 correlates with decreased binding ability. With only dynamic information available on three dsRBDs – PKR-dsRBD1, PKR-dsRBD2 and Staufen-dsRBD3 – it is difficult to generalize the exact roles dynamics has on the binding mechanism of dsRBDs.

Even with all the structures determined of dsRBDs, both with and without RNA bound, very few atomistic MD simulations exist, hindering our understanding of the dynamics of dsRBDs. One notable example is the MD simulations of *Drosophila* Staufen-dsRBD3 with and without dsRNA bound. In this case, loop 2 and loop 4 (both involved in RNA recognition) are highly dynamic in the apo-state and the holo-state, which agreed with heteronuclear NOE data. While this validation of the MD simulations is critical, without further MD simulations of dsRBDs to compare it with, it is difficult to build a comprehensive understanding of dsRBDs.

Overall, the biggest gap in our understanding of dsRBD function is the roles dynamics play in the binding mechanism. This is evident when comparing the number of structures determined versus the amount of dynamic information reported to date, which is in stark contrast to the RNA recognition motif (RRM) that binds ssRNA, as
exemplified by U1A. In order to understand the mechanism of RNA recognition by dsRBDs, both dynamic and structural information, together with binding affinities, need to be gathered on the same dsRBDs. As a follow up, the same information from various dsRBDs must be compared in order to formulate a predictive model for the role dynamics have on dsRBD binding to dsRNA.

1.3.2 Model System: The miRNA Maturation Pathway:

To understand RNA binding by dsRBDs, I have investigated a multitude of human proteins containing dsRBDs that contribute to one specific metabolic task: maturation of microRNA (miRNA). The miRNA maturation pathway is an especially effective choice for this study because it contains ten dsRBDs from five different proteins, so within one pathway these various dsRBDs must have evolved different binding abilities and functions. This enables each dsRBD to carry out a specific role within the pathway while discriminating from other dsRNAs in the cell.

miRNAs are a large class of RNAs that affect gene-regulation through the RNA interference (RNAi) mechanism. To date, over 1500 mature human miRNAs have been deposited on the miRNA database (www.mirbase.org) and more than 60 percent of the human genes are regulated post-transcriptionally by at least one miRNA. Roughly 22 nucleotides in length, miRNAs contribute to cellular homeostasis, while misregulation leads to disease states such as cancer, Alzheimer’s disease, heart disease, and autoimmune diseases. Current research aims at utilizing miRNAs in regulating cellular protein levels, in detecting biomarkers for disease, and in “gain-of-function” or “loss-of-function” treatments.
Most miRNAs are transcribed \emph{in vivo} by RNA polymerase II (PolII) as a long RNA with a well conserved secondary structure that consists of a single-stranded (ss) tail region, some imperfections (internal loops and/or bulges) along the A-form helix and a terminal loop (Figure 1.3). After transcription, the maturation of miRNAs is carried out in a two-step process that is independently and spatially separate from each other (Figure 1.3A). First, the ssRNA tail region is cleaved off by the “microprocessor” in the nucleus.\(^{45}\) The microprocessor is composed of a RNase III enzyme (Drosha) with its cofactor (DGCR8), and together they are responsible for cleaving pri-miRNA roughly 11 base-pairs from the ss-ds junction to form a 60-70 nt in length with a characteristic 2 nt 3′ overhang pre-miRNA (Figure 1.3B).\(^{45}\) Without DGCR8, Drosha will cleave randomly along the dsRNA part of pri-miRNA.\(^{41}\) Next, the pre-miRNA is exported to the cytosol by exportin-5 in a RanGTP dependent manner, where the second step in the maturation occurs by another RNase III enzyme, Dicer, removing the terminal loop (Figure 1.3B).\(^{45}\) Like Drosha, Dicer requires a dsRBP (either TRBP or PACT) to properly carry out its role in the miRNA maturation pathway.\(^{45}\) The mature miRNA is then loaded into an Argonaut (Ago) family protein forming the RNA-induced silencing complex (RISC).\(^{45}\)

The maturation steps for non-intronic miRNAs are nearly universal, even though the miRNAs themselves are quite diverse, suggesting that the secondary structure elements (i.e., the ss-ds junction, terminal loop and internal loops/bulges) of the intermediates (pri-miRNA and pre-miRNA) is critical for their recognition by the various proteins in the pathway. This hypothesis is supported by crystallographic data for the exportation of the pre-miRNA to the cytoplasm\(^{46}\) and Dicer processing in the cytosol,\(^{45}\) but little is known on the early, nuclear maturation mechanism. The secondary
Figure 1.3: miRNA Maturation Pathway. (A) Representation of the stages in miRNA maturation and function. (B) Representation of the two cleavage steps in the miRNA maturation process. miR-16-1 is used as an example miRNA in the representation, with the mature miRNA shown in red, the pre-miRNA depicted in capital letters, and the portion of the pri-miRNA that is cleaved in the first step shown in grey and depicted in lower case letters. The location in the cell that the process occurs is to the right of the vertical arrow representing the cleavage.
structures of many pri-miRNAs have been computationally predicted, but none have been verified by crystallography or NMR spectroscopy prior to the start of my thesis work. During my thesis work, a crystal structure of pre-miR-30a bound to a exportin-5:RanGTP complex was reported. The diffraction data were only sufficient to resolve the A-form helical section of the pre-miRNA, leaving the terminal loop and bulge undefined. The base-pairing register of the stem is consistent with computational structures posted on the miRNA database and predicted by mfold.

A majority of the proteins mentioned in the miRNA maturation pathway contain at least one dsRBD, with many of the proteins containing multiple dsRBDs in tandem. As stated earlier, ten dsRBDs are spread between five proteins in this pathway: one from each RNase III enzyme, two from DGCR8 and three each from TRBP and PACT. Prior to the start of my dissertation work, only the crystal structure of the two dsRBDs from human DGCR8 and the crystal structure of the dsRBD from mouse Dicer (one hundred percent identical to human) with the RNaseIIIb domain were reported. Since then, the NMR solution structure of the dsRBD from Drosha and the crystal structure of the first two dsRBDs from TRBP have been reported. These recent structure determinations illustrate the interest in the field for understanding these domains within the context of the miRNA maturation pathway. While all the structural information is useful, it alone fails at providing a complete understanding of the mechanism dsRBDs implement in binding miRNA precursors and discriminating from other dsRNAs in the cell. Therefore, the focus of this dissertation has been to expand upon static structure determination by understanding the role dynamics, especially on the fast timescale, has on dsRBD binding of dsRNA. This work focuses specifically on four out the ten dsRBDs.
in the miRNA maturation pathway: the two from DGCR8, one Drosha and one from Dicer.

1.4 Protein Interactions of Intrinsically Disordered Proteins (IDPs):

The dogmatic role of RNA as a messenger has recently been expanded with the emergence of functional RNAs, like miRNAs. Likewise, the field of protein chemistry has seen recent challenges to the structure-function paradigm with the discovery of proteins or regions of proteins that have no structure, but still have well-defined functions in the cell, such as signal transduction, transcription and translation. In a sense, intrinsically disordered proteins (IDPs) are very similar to RNAs because both are highly flexible in solution and both have a preponderance of hydrophilic groups making them soluble in an aqueous media in an extended conformation. Thus, studying an IDP system provides a good complement to understanding protein-RNA interactions.

1.4.1 Intrinsically Disordered Proteins (IDPs):

IDPs are defined as lacking a unique three-dimensional structure. Based on current data at least 50% of proteins in the human genome contain disordered stretches of at least 30 residues. IDPs are enriched in charged and polar amino acid residues (e.g., Arg, Lys, Glu, Pro, and Ser), while depleted in hydrophobic amino acid residues (e.g., Cys, Trp, Tyr, Ile, and Val), which leads to low sequence complexity. The unique physical and sequence properties of IDPs suggest a biological advantage for interactions with IDPs such as many-to-one signaling, high-specificity low-affinity binding, and rapid binding kinetics. Progression of diseases like cancer, Parkinson’s and Alzheimer’s have been linked with IDPs.
Many IDPs fold upon binding to a stable structure when interacting with the appropriate partner, but the resulting complex is quite different from the complexes between two well ordered proteins.\textsuperscript{59-61} First, ordered protein complexes are generally stabilized by polar interactions of residues, while disordered protein complexes are more frequently stabilized by hydrophobic interactions.\textsuperscript{60, 62} Second, the residues involved in the binding surface are often distant in the protein sequence for ordered protein complexes, but are usually close in the protein sequence for disordered protein complexes.\textsuperscript{60, 62} Third, in general, IDPs utilize more interaction surface area to obtain highly specific protein-protein interactions than well ordered protein complexes and thus bury a much larger surface when bound.\textsuperscript{62} These observation lead to the hypothesis that IDPs are evolutionary favored for protein-protein interactions because they can accomplish their roles with shorter polypeptide sequences which leads to a saving in cellular space and genomic storage.\textsuperscript{60} The difference in the interaction made by IDPs with their partners from well ordered systems leads to qualitative difference in the resulting complex. In well ordered complexes, the resulting complex has an energy minimum, but this is not necessarily true for disordered complexes where an energy minimum might not exist leading to the colloquial expression in the literature of a “fuzzy complex”.\textsuperscript{61, 63}

Much like miRNAs, a lot of work on IDPs has been done over the last few decades. Research in this field has focused at understanding the binding mechanism and characteristics of disordered complexes versus well-ordered complexes. While limiting models have been developed and examples exist to support the different models, more questions regarding the general characteristics of IDP binding still exist.
Again, in order to fully understand IDP functions in the cell, both a structural and dynamic description of IDP interactions with their partners need to be elucidated.

### 1.4.2 Model System: the C-terminal Domain of FCP1:

For this section of the dissertation, I have chosen to focus on the intrinsically disordered C-terminal region of FCP1 and its binding partner the winged-helix domain of RAP74, together they are responsible for termination of transcription and PolII recycling. Much like the dsRBDS from the miRNA maturation pathway, the motivation for studying this system stems from the lack of dynamic information – either experimental or computational – accompanying the available structural and functional information. Experimental studies of FCP1 have been carried out by a fellow graduate student in the Showalter group, Chad Lawrence. In discussion, it became clear that computational studies would provide strong synergy with Chad’s ongoing studies; therefore, dynamics of the RAP74-FCP1 complex \textit{in silico} were performed.

The phosphorylation state of the PolII C-terminal repeat domain (CTD) regulates progression through the mRNA biogenesis cycle. Together, the general transcription factor II\(\text{F}\) (TFIIF) and the TFIIF-associating CTD phosphatase (FCP1) are responsible for the dephosphorylation of the CTD allowing for transcription to terminate and PolII to be recycled elsewhere on the DNA strand. FCP1, the only known CTD specific phosphatase for PolII in humans, activity is promoted through interaction of its intrinsically disordered C-terminal acidic domain (residues 879-961) with the C-terminal winged-helix (WH) domain of the RAP74 chain of TFIIF.
The C-terminal domain of RAP74 is composed of three consecutive α-helices, connected by two loops (4 and 10 residues in length, respectively) and two antiparallel β-strands, forming a WH domain. The WH domain of RAP74 has been shown to be both necessary and sufficient for FCP1 activation.67,68 The structure of the FCP1 acidic region bound to the WH domain RAP74 has been determined by crystallography69 and NMR spectroscopy70 revealing an α-helical conformation for residues 944-960 of FCP1 that inserts into a hydrophobic groove formed between helix 2 and helix 3 of RAP74. Additionally, this data shows that most of the contacts between RAP74 and FCP1 are van der Waals interactions, with only a few salt bridges and hydrogen bonds.69-71

Previously reported structures for apo-RAP74 and the RAP74-FCP1 complex provided insight into molecular interactions between the proteins, but do not provide a quantitative understanding of the spatio-temporal dynamics involved in binding and in maintaining the complex; no reports of experimental (e.g. NMR spin relaxation) or computational investigations into the dynamics of this system appear in the literature. To elucidate the interactions stabilizing the complex, their effect on dynamics, and the mode of binding of FCP1 by RAP74, atomistically detailed MD simulations of the apo-RAP74 structure and the RAP74-FCP1 complex were performed.

1.5 Methods for Studying Macromolecular Interactions:

1.5.1 Electrophoretic Mobility Shift Assay (EMSA):

The major focus of this dissertation is on understanding the role dynamics has on function; therefore, functionality must be determined in connection with dynamic studies. A dsRBD is defined as functional if it is able to bind dsRNA, regardless of its role in the
cell. While quite a few methods exist for determining the binding affinity for protein-RNA interactions, I chose to use the EMSA technique because it is widely used in the literature, especially in the miRNA maturation pathway.\textsuperscript{27, 72-74}

The basic idea is that free nucleic acid migrates further on a non-denaturing gel than the protein-nucleic acid complex, due to an increase in the molecular weight of the complex, giving raise to distinct bands.\textsuperscript{75} By running multiple binding reactions, where the protein concentration is varied while keeping nucleic acid concentration the same, on the same gel the binding affinity can be determined. As the protein concentration is increased the free nucleic acid band should decrease in intensity while the band representing the bound state should increase. A plot of the fraction of RNA bound versus protein concentration should give a sigmodial curve, which can be fit to a general Hill equation:

\[
\text{Fraction bound} = \frac{[\text{protein}]^n}{[\text{protein}]^n + K_d^n} \quad (1.1)
\]

where $n$ is the Hill coefficient and $K_d$ is the dissociation constant.\textsuperscript{76} The Hill coefficient is used to describe the cooperativity of binding with a $n$ value greater than one meaning positive cooperativity, a $n$ value less than one meaning negative cooperativity and a $n$ value equal to one meaning non-cooperative binding. In the general Hill equation the stoichiometry of the binding is not known \textit{a priori}, but if the stoichiometry is known it is possible to use the more rigorous fully cooperative model:

\[
\text{Fraction bound} = \frac{(K_d[\text{protein}])^N}{1+(K_d[\text{protein}])^N} \quad (1.2)
\]
where \( N \) is the stoichiometry and is not allowed to be varied in the fitting algorithm.\(^{76}\) Likewise the sigmodial curve can also be fit to a general model assuming \( N \)-independent and identical binding sites:

\[
\text{Fraction bound} = \frac{NK_a[protein]}{1+K_a[protein]} \quad (1.3)
\]

Throughout the dissertation the \( N \)-independent and identical binding site model fit the EMSA data poorly as demonstrated in Appendix A Figure A.2C. Therefore, the fully cooperative model where \( N \) is allowed to float (equivalent to the general Hill equation) was utilized to fit the data since the stoichiometry of the system is not known. Note, that the fitting methods described above are not specific to the EMSA method, but could be used in any experiment where the concentration of the protein is varied and the fraction bound can be monitored.

The EMSA technique is one of the most widely used techniques for determining the binding affinity of protein-nucleic acid interaction because it is robust and accommodates a wide range of binding conditions including sub-nanomolar concentrations,\(^{77}\) but it is not without limitations. First, the samples are not at chemical equilibrium during the gel electrophoresis.\(^{77}\) Therefore, depending on the dissociation rate, the complex might go undetected (rapid dissociation) or cause an underestimation of the binding density (slow dissociation).\(^{77}\) Second, the mobility of protein-nucleic acid complexes depends on many factors other than the size of the complex and thus does not provide straightforward information on the binding stoichiometry.\(^{77}\) Without knowing the stoichiometry, only a general model (like the general Hill model) can be used to fit the data and therefore not provide the true binding affinity. Stoichiometry can be
determined through various other methods including analytical ultracentrifugation (AUC, which was used in collaboration with Prof. Jim Cole for one specific complex discussed in Chapter 4. Finally, EMSAs do not provide any direct information on the binding site on the nucleic acid that the protein recognizes. When this information is important for a given study, methods like nuclease and chemical footprinting assays must be performed independently of EMSAs.

1.5.2 NMR Spin Relaxation:

The principle motivation of my dissertation work has been understanding the correlation between fast dynamics and function in proteins. Fast dynamics of a macromolecule can be experimentally determined through NMR spin relaxation. This section will cover how NMR spin relaxation measurements are collected and fitted using the Lipari-Szabo Model-free formalism to obtain information on the flexibility of a protein on a per residue basis. Throughout this dissertation backbone amide spin relaxation measurements have been collected in order to obtain fast timescale dynamic information, therefore the discussion that follows pertains to this type of relaxation.

1.5.2.1 NMR Spin Relaxation Measurements:

NMR spin relaxation is the process by which non-equilibrium magnetization returns to its equilibrium state (or transforms to another non-equilibrium state). For spin ½ nuclei, such as 15N, the two major mechanisms for magnetic relaxation in solution that are influenced by molecular motions on the fast timescale are dipole-dipole interactions and chemical shift anisotropy (CSA). First, the dipole-dipole mechanism arises due to the fluctuating coupling between two magnetic nuclei as the vector connecting them rotates relative to the permanent magnetic field. The rotation of the
bond vector could be due to either the molecular tumbling or internal motions.\textsuperscript{78} The strength of the dipolar coupling is dependent on the distance between the two nuclei, thus for backbone amides this mechanism is dominated by contributions from the attached proton.\textsuperscript{78} Next, the CSA mechanism arises due to the fluctuating magnetic field surrounding the nuclei caused by variations in shielding of the nuclei.\textsuperscript{78} Both molecular tumbling and oscillations of the amide group relative to the rest of the protein causes changes in the orientation of the shielding tensor of the nitrogen nucleus relative to the permanent magnetic field, thus stimulating relaxation.\textsuperscript{78}

Typically, backbone NH bond vector dynamics are measured via three different experiments (all measured through the 2D $^{15}$N-HSQC): $^{15}$N longitudinal relaxation ($R_1$), $^{15}$N in-phase transverse relaxation ($R_2$), and the heteronuclear Overhauser effect ($^{1}$H-$^{15}$N NOE).\textsuperscript{78} First, the longitudinal or spin-lattice relaxation is the measurement of the recovery of Z-magnetization to its thermodynamic equilibrium state with the surroundings (i.e., the lattice).\textsuperscript{79} This is experimentally measured by transferring magnetization to the Z-axis on the nitrogen, but having the equilibrium polarization of the proton, and allowing the magnetization to relax by varying the time interval (generally between 50 and 1000 ms) before transferring magnetization back to the proton for detection. During this period the proton is decoupled to suppress cross-relaxation. Multiple spectra are collected in an interleaved manner where the only difference is the time interval for relaxation. At longer time intervals more relaxation occurs causing a lower intensity in the $^{15}$N-HSQC peaks as compared with shorter time intervals. The intensity of the peaks in the $^{15}$N-HSQC spectrum from a given residue can be plotted against the time interval and fitted to an exponential decay function in
order to obtain \( R_1 \) (or \( T_1 \), which is \( 1/R_1 \)) for each residue. High \( T_1 \) values correspond with dynamic residues in the protein. Longitudinal relaxation is field dependent with lower fields giving lower \( T_1 \) values.

Next, the transverse or spin-spin relaxation is the measurement by which the transverse component of the magnetization returns to its equilibrium value of zero.\(^79\) This is experimentally measured by obtaining transverse magnetization on the nitrogen in phase with respect to the scalar-coupled proton. This magnetization is allowed to evolve according to chemical shift for a variable time interval (\( \tau \), generally varies between 15 and 200 ms). Next, a 180 degree pulse is applied to invert the magnetization and allowed to refocus during a time interval equal to \( \tau \). As described, this is the Hahn Spin-Echo experiment (a single spin echo experiment); in practice the Carr-Purcell-Meiboom-Gill (CPMG) sequence, a double spin echo experiment, is utilized to suppress errors in inaccurate 180° pulses and chemical exchange. Like longitudinal relaxation, multiple transverse relaxation spectra are collected in an interleaved manner where the only difference is the time interval for relaxation. Also like \( T_1 \), the \( T_2 \) for each residue is obtained by fitting the intensity of peaks in the \(^{15}\text{N}-\text{HSQC} \) to an exponential decay with time as the independent variable, with a high \( T_2 \) value corresponding with a residue that is more dynamic. Unlike longitudinal relaxation, transverse relaxation is not heavily field dependent.

Finally, the \( \{^{1}\text{H}\}-^{15}\text{N} \) NOE is a measurement of relaxation through dipole-dipole coupling (i.e., through space interactions) from a saturated spin state to the measured spin state.\(^78\) This is accomplished by saturation of the proton for a long time causing a change in the steady state of the nitrogen when the \(^{15}\text{N}-\text{HSQC} \) is started. The
experiment is then repeated without the pre-saturation in order to measure the equilibrium state of the nitrogen as a reference. Owing to low signal-to-noise ratio in heteronuclear NOE experiment, the experiment is repeated more than once per field strength to reduce errors in the measurements. The NOE value is calculated by taking the ratio of the intensity from each peak in the proton saturation experiment to the reference experiment. In the extreme narrowing limit, when the Larmor frequency times the correlation time is much less than one (so for small molecular weight molecules or very rapidly fluctuating sites in proteins), the NOE equals the order parameter, and thus, sometimes, fast timescale dynamics for residues are inferred from the NOE values alone and not the complete spin relaxation set.

Once all of the NMR spin relaxation data is collected, then estimation for the global rotational diffusion properties of the macromolecule is performed in a two step process. To accomplish this only the internally rigid residues are utilized by determining the ratio between the transverse to the longitudinal relaxation ($R_2/R_1$). Most residues will have an intermediate $R_2/R_1$ reflecting the global tumbling parameters and the angle(s) between the bond vector of interest and the global diffusion axes (the latter is trivial in the isotropic case). $R_2/R_1$ values higher than the expected value represent sites that undergo chemical exchange on a slow timescale, while low $R_2/R_1$ values reflect sites with very fast internal dynamics. Also, sites with very fast internal dynamics are observed to have low NOE values. Both residues with high and low $R_2/R_1$ values, as well as low NOE values, are excluded for the estimation calculation of global rotational diffusion properties but these residues are retained for future Lipari-Szabo Model-free fitting. The residues that are not excluded are then utilized to determine the
optimal rotational diffusion tensor to use for model-free analysis (following section) through $F$-statistical comparisons of the different models (isotropic, axially symmetric and anisotropic).\textsuperscript{78}

\textbf{1.5.2.2 Lipari-Szabo Model-Free Fitting:}

Once the spin relaxation measurements have been collected and an estimate of the molecular rotational diffusion tensor has been determined, then relaxation data can be analyzed utilizing the Lipari-Szabo Model-free formalism to obtain the normalized order parameter ($S^2$) and the apparent internal motion correlation time ($\tau_e$).\textsuperscript{80} While other methods exist for analyzing the spin relaxation data, the Lipari-Szabo Model-free formalism is preferred because as the name implies it does not invoke a specific model for internal motion and therefore it is transferrable from one biological system to another.\textsuperscript{80} The major assumption the formalism makes is that the internal dynamics can be separated from the global motions.\textsuperscript{80} Initially the fitting equations were for isotropic rotation (i.e., spherical molecules) but they have since been expanded to allow for axially symmetric and completely anisotropic molecules.\textsuperscript{78} Additionally, a modification has been made to account for proteins with internal motions on two timescales, only one of which is fast relative to overall tumbling but the values obtained from this method are not necessarily realistic.\textsuperscript{78} Finally, in some cases conformational exchange in the microsecond to millisecond occur, thus inflating the transverse relaxation rate constant relative to the value defined by dipole-dipole and CSA mechanisms.\textsuperscript{78} In this case, the transverse relaxation rate is the summation of the dipolar, CSA and exchange broadening (designated $R_{\text{ex}}$) contributions, where $R_{\text{ex}}$ is an adjustable parameter during the fitting.\textsuperscript{78}
Throughout this dissertation the spin relaxation data fitting was performed utilizing the ModelFree program (A. G. Palmer III, Columbia University), specifically model 2 and occasionally model 3 of the software package, which represent the original model-free formalism and the correction for conformational exchange broadening by addition of the $R_{ex}$, respectively. Due to this fact the following section will go over the original fitting method for NMR spin relaxation data along with brief comments about corrections for the $R_{ex}$ term. The overall motions are assumed to have a correlation function that decays to zero exponentially with a typically slow characteristic timescale ($\tau_m$), while the internal motions give a correlation function that decays to a finite value with a typically fast characteristic timescale ($\tau_e$). Therefore, the resulting total correlation function is a double exponential with slow and fast phases, representing the global and internal motions, respectively. The order parameter ($S^2$) is defined as the amplitude of the internal motion, which represents the degree of rotational restriction for the internal motions. The order parameter is a normalized value, where $S^2 = 1$ is completely restricted motion and $S^2 = 0$ is completely unrestricted motion. In the original form, two values are fitted ($\tau_e$ and $S^2$) per each bond vector with a third global value ($\tau_m$) fitted utilizing all residues. Because the global value is the same for all n bond vectors then 3n (per field) experimental data points are used to fit 2n + 1 parameters, so the system is over-determined. As mentioned earlier, $R_2$ is affected by chemical exchange on the slow timescale, which is minimized by using the CPMG sequence, but is not always fully suppressed. In this case, the $R_2$ value is used to fit an additional term ($R_{ex}$), while the $\tau_e$ and $S^2$ are determined solely by the measured $R_1$ and NOE values (model 3).
To recap, the Lipari-Szabo model-free fitting takes in $R_1$, $R_2$, and NOE data and provides information on the global motion, internal motion and the order parameter, a normalized value on the degree of flexibility on a per residue basis. As mentioned the fitting is not based on knowing anything about the chemistry of the system and the outputted order parameter is normalized, making it possible to quantitatively compare values from different systems. It is this transferability from one system to another that will be utilized to correlate dynamics and function from one protein to another.

1.5.3 Molecular-Dynamics (MD) Simulations:

A symbiotic relationship exists between NMR experimental data and MD simulations since both report on fast timescale dynamics and MD simulations have been shown to reproduce NMR data.$^2$, $^{82-84}$ The basic idea behind MD simulations is that dynamics of a biological system can be modeled by integrating Newton’s laws of motion, thus gaining information on the trajectory and position of the atoms in the system over time.$^{85}$ This section aims to focus on only a few aspects of MD simulations as they relate to rest of the dissertation; notably the derivation of correlated motions and the derivation of computationally calculated order parameters.

1.5.3.1 Correlated Motions:

One of the major reasons for running MD simulations in this dissertation is the ability of them to provide information on correlated dynamics, especially over long distances, which are difficult to achieve experimentally. Correlated motions give information on how atoms move in relationship to each other. The motions can be positively or negatively correlated. Positive correlations mean that the pair of atoms move in the same direction at the same time, therefore the distance between them
remains the same, which might imply a through space interaction, e.g., hydrogen bond or salt bridge. On the other hand negative correlations mean that the pair of atoms move in the opposite direction at the same time, therefore the atoms get closer or further apart from each other over time.

In order to obtain correlated motions a covariance matrix must be calculated, which provides information about fluctuations on pairs of atoms.86,87 A prerequisite for acquiring the covariance matrix is that the internal motions must be separated from the global motions. In general, this is accomplished by removing the translational by centering each structure in the simulation about its center of mass. Removal of the rotational component of the global motion is more challenging and generally is reduced by superimposing each structure in the simulation on a reference point. Afterwards, the covariance matrix (M) can be computed from the atomic coordinates as follows:

\[ M_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \]  

(1.4)

where \( q_1, \ldots, q_{3N} \) are the mass-weighted Cartesian coordinates and the \( \langle \ldots \rangle \) represent the average over all sampled conformations. Therefore, if there are N atoms of interest in the system, then the dimension of the corresponding covariance matrix is 3N X 3N. The eigenvectors and eigenvalues of the covariance matrix yield information on the collective motional modes and the amplitudes of these modes, respectively. This represents the principle component analysis (PCA) method, which is a standard method and utilized in the AmberTools kit.88

The main drawback of PCA is that it requires that the structures be superimposed in order to remove rotational motions. In some cases the structures cannot be readily superimposed due to flexibility in the macromolecule, thus another
method like isotropically distributed ensemble (IDE), where a covariance matrix is derived without first removing the rotational motions must be utilized. In IDE the covariance matrix is calculated as followed:

\[ M_{ij} = \frac{1}{3}\langle r_i \cdot r_j \rangle \]  

(1.5)

where \( r_i \) represents the \( i_{th} \) atomic position vector with respect to the center of mass. In this analysis the covariance matrix only depends on the length and relative orientation of the vectors and not the exact Cartesian coordinate of the atoms as in PCA. This methodology implies that each structure is in an isotropic solution undergoing anisotropic or isotropic rotational tumbling; therefore, an isotropic orientation distribution of each structure is created. This creates a three-fold degeneracy of each eigenmode, so only an \( N \times N \) matrix needs to be diagonalized and the \( 3N \) eigenvectors can be reconstructed by the direct product of the eigenmodes with a 3D basis set. The first three eigenmodes correspond with overall rotational motion which can be separated from the rest of the eigenmodes allowing for the reconstruction of a covariance matrix of the internal motions that has a dimension of \( N-3 \times N-3 \). Due to this fact the correlation between globular domains in a protein or correlation of more flexible proteins, like IDPs, can be calculated from MD simulations without first superimposing the structures. Unless specifically mentioned correlations are determined through the standard AmberTools (PCA) method throughout this dissertation and the more rigorous IDE analysis is reserved for systems where I show that the structures cannot be superimposed.
1.5.3.2 MD-Derived Order Parameters:

As MD simulations report on fast timescale dynamics a good experimental validation tool is to compare order parameters derived from MD simulations with experimentally obtained NMR order parameters. To acquire order parameters from MD simulations a covariance matrix and a set of correlation times is needed. Again, a rigorous method for determining the covariance matrix is required but an additional condition that each internal conformation is represented by a large set of snapshots with an isotropic orientational distributed ensemble is needed.\(^9\) The IDE method for determining the covariance matrix above cannot be utilized for calculating the order parameters because the reorientational tumbling correlation time is within the nanosecond range, which is insufficiently sampled using current all atom MD simulations.\(^9\) Additionally, the IDE matrix tracks the fluctuations of any given atom relative to the coordinate origin and not specific vectors, e.g., the N-H vector, directly. Therefore, order parameters are best calculated utilizing the isotropically reorientational eigenmode dynamics (iRED) method, which calculates the covariance matrix as follows:

\[
M_{ij} = \frac{1}{2}(3(e_i - e_j)^2 - 1)
\]  

where \( e_i \) and \( e_j \) are the normalized bond vectors, taken from the same snapshot.\(^2\) The covariance matrix above is only composed of inner products making it valid in an arbitrary frame with no assumptions about the separability between internal and overall tumbling motions.\(^2\) In this representation, the first five eigenmodes represent global motions and can be separated from the rest of the eigenmodes similar to the IDE analysis. From the covariance matrix the order parameter values can be computed using the equation below:
\[ s^2 = 1 - \sum_{m=1}^{N-5} \lambda_m |m\rangle^2 \]  

where N is the number of interaction vectors \( e_m \) and \(|m\rangle\) is the internal modes with eigenvalues \( \lambda_m \).

1.6 Dissertation Outline:

The overall theme of this thesis is to understand how fast dynamics of proteins correlate with their function. This was accomplished by studying two model systems: dsRBDs within the miRNA maturation pathway (Chapters 2 through 5) and WH domain of RAP74 binding the C-terminal IDR of FCP1 (Chapters 6 through 7).

In Chapter 2, MD simulations were utilized to investigate correlated motions between the two tandem dsRBDs in DGCR8-Core as suggested by the packing of the domains against the C-terminal \( \alpha \)-helix and well-organized linker. The pre-organization of the two dsRBDs in the absence of RNA allows for correlated motions that are important for fine-tuning the distance separating the two dsRNA-binding surfaces. Additionally, MD simulations imply that these motions are mediated through two key interfacial interactions between the two dsRBDs – a hydrogen bond and a salt bridge – this can be furthered explored experimentally through thermodynamic cycles (Chapter 5).

Work in Chapter 3 and 4 investigate dynamics and its contribution to function through the individual and isolated dsRBDs from DGCR8, Drosha, and Dicer. Initial EMSAs revealed that while DGCR8-dsRBD1 binds pri-16-1, Drosha-dsRBD does not (Chapter 3). Dynamic data on the two dsRBDs in their apo-state reveal that the lack in
Drosha-dsRBD binding could be partly due to the increased flexibility in loop 1, a negatively charged extension of the canonical dsRBD loop 1 (Chapter 3). Purposed future work aims at validating this hypothesis through constructing chimeras that swap loop 1 of various dsRBDs (Chapter 5). In Chapter 4, EMSA data illustrate that not only is Dicer-dsRBD able to bind pre-miR-16-1 but also perfect duplex RNAs of various lengths (12 to 44 base-pairs) albeit with a minimal decrease in binding affinity for the smaller RNAs, which is recovered by the addition of a terminal loop or ss tail. The dynamic profile for apo-state Dicer-dsRBD is similar to DGCR8-dsRBD1 thus a dynamic profile for apo-state dsRBDs where loop 2 is generally the most flexible region of the domain is being established.

In Chapter 6 and chapter 7 the mechanism of IDP interactions through MD simulations utilizing the C-terminal IDR of FCP1 binding to the WH domain of RAP74 as a model system due to the extent of experimental data available is investigated. MD simulations of the complex and apo-RAP74 further illuminates on the significance of hydrophobic contacts for stabilizing disordered protein complexes, and how they allow for increased dynamics in the complex (chapter 6). Future computational studies of the RAP74-FCP1 complex on the adaptability and binding mechanism of RAP74 could be done by comparing MD simulations of RAP74 with the central domain and the C-terminal domain of FCP1 (Chapter 7). Also, MD simulations of the RAP74-FCP1 complex with crowding agents and osmolytes introduced can be performed to further elucidate the cellular mechanism of IDP interactions (Chapter 7).
1.7 References:


5. Tripet, B. P.; Goel, A.; Copie, V., Internal dynamics of the tryptophan repressor (TrpR) and two functionally distinct TrpR variants, L75F-TrpR and A77V-TrpR, in their L-Trp-bound forms. *Biochemistry* 2011, 50, 5140-5153.


43. Ai, J. Z., Rong; Li, Yue; Pu, Jielin; Lu, Yanjie; Jiao, Jundong; Li, Kang; Yu, Bo; Li, Zhuqin; Wang, Rongrong; Wang, Lihong; Li, Qiang; Wang, Ning; Shan, Hongli; Li, Zhongyu; Yang, Baofeng, Circulating microRNA-1 as a potential novel biomarker for acute myocardial infarction. *Biochem. Biophys. Res. Commun.* 2010, **391**, 73-77.


Chapter 2

MD Simulations of the dsRBP DGCR8 Reveal

Correlated Motions that May Aid pri-miRNA Binding


W. G. Noid made Figure 2.6.

2.1 Abstract:

Over the past decade, microRNAs (miRNAs) have been shown to affect gene regulation by base-pairing with messenger RNA, and their misregulation has been directly linked with cancer. DGCR8, a protein containing two dsRNA binding domains in tandem, is vital for nuclear maturation of primary miRNAs (pri-miRNAs) in connection with the RNase III enzyme, Drosha. The crystal structure of the DGCR8-Core (493-720) shows a unique well ordered structure of the linker region between the two dsRBDs, which differs from the flexible linker connecting the two dsRBDs in the antiviral response protein, PKR. To better understand the interfacial interactions between the two dsRBDs, we ran extensive MD simulations of the isolated dsRBDs (505-583 and 614-691) and the Core. The simulations reveal correlated reorientations of the two
domains relative to one another, with the well ordered linker and C-terminus serving as a pivot. The results demonstrate that motions at the domain interface dynamically impact the conformation of the RNA binding surface and may provide an adaptive separation distance necessary to interact with a variety of different pri-miRNAs with heterogeneous structures. These results thus provide an entry point for further in vitro studies of the potentially unique RNA-binding mode of DGCR8.

2.2 Introduction:

Mature microRNAs (miRNAs) are small single-stranded RNAs (ssRNAs), 21 to 25 nucleotides in length, that affect gene translation by base-pairing with messenger RNA. A majority of human genes (at least 60%) are regulated post-transcriptionally by one or more miRNAs. Recent studies have demonstrated that miRNAs contribute to the control of cellular homeostasis in multicellular organisms by regulating such biologically important processes as apoptosis, cell cycle progression, and cell type differentiation, as well as disease states leading to cancer, Alzheimer’s disease and autoimmune diseases. Maturation of miRNA occurs in two independent and spatially separated steps. The first step occurs in the nucleus where the single-stranded tail of primary microRNA (pri-miRNA) is removed by the “microprocessor,” which contains the RNase III enzyme Drosha and the double-stranded RNA (dsRNA)-binding protein (dsRBP) DGCR8 (known as Pasha in Drosophila). The second maturation step occurs in the cytosol, where the terminal loop of pre-miRNA is cleaved by Dicer, another RNase III enzyme.

Originally, the dsRNA binding domain (dsRBD) containing protein DGCR8 was identified in humans as a protein encoded in the region of chromosome 22 that is
deleted in patients with DiGeorge syndrome.\textsuperscript{9} Since then, DGCR8 has been shown to play a critical role in processing pri-miRNA into pre-miRNA by binding to the double-stranded region of the pri-miRNA.\textsuperscript{9} Throughout the maturation process, dsRBPs are required for recruitment and cleavage of the nascent miRNA. dsRBPs are seen in all forms of life, from viral-encoded to prokaryotic and eukaryotic products, making them one of the most common RNA-binding motifs.\textsuperscript{10} Also, dsRBPs are found in both the cytoplasm and the nucleus, as evidenced by the compartmentalization of the first and second stages of miRNA processing in the nucleus and cytosol respectively.\textsuperscript{7, 11, 12} The dsRBD, which is evolutionary conserved, consists of \textasciitilde 65 to 68 amino acids forming an $\alpha\beta\beta\alpha$ secondary structure.\textsuperscript{13-15} Some eukaryotic dsRBPs contain up to five dsRBDs, while others contain only one.\textsuperscript{16} As few as 11 base-pairs of dsRNA (coinciding with one turn of canonical A-form helix) has been shown to interact with a single dsRBD.\textsuperscript{17} A-form double helix RNA is the preferred binding partner of dsRBDs, which show little or no affinity for ssRNA, dsDNA and DNA-RNA hybrids.\textsuperscript{18, 19}

DGCR8 is a 773 amino acid residue protein that contains two dsRBDs at the C-terminal end of the protein, separated by a \textasciitilde 50 amino acid linker (Figure 2.1A). Except for a predicted WW domain in the middle of the sequence, the rest of the protein is intrinsically disordered and without known function. Even though DGCR8 does not contain known enzymatic activity, its role as a “molecular anchor” directing Drosha to cleave the pri-miRNA \textasciitilde 11 base-pairs from the ssRNA-dsRNA junction is vital for specificity in the miRNA maturation process.\textsuperscript{20}

Recently, Sohn \textit{et al.}\textsuperscript{21} determined the crystal structure of the Core region of DGCR8 (residues 493 to 720 of the human sequence), which contains both dsRBDs
**Figure 2.1**: Cartoon representation and tertiary structure of DGCR8-Core. (A) Schematic representation of the primary sequence of DGCR8. (B) A ribbon diagram representing the crystal structure of DGCR8 Core (PDB 2YT4, residues 505-701) shows a well structured linker (H3 and H4) and a C-terminal helix (H5, black). The nomenclature used is the same as the crystal structure nomenclature (20). dsRBD1 is shown in red and dsRBD2 is shown in green (in dsRBD2 helices and strands are denoted with primes to indicate equivalence with secondary structures to dsRBD1). The arrows indicate the proposed dsRNA binding sites.
and the intervening ~50 amino acid residue linker. The crystallographic data revealed that the two dsRBDs are arranged in a pseudo two-fold symmetry and packed against a well-defined secondary structure formed from the linker and the C-terminal tail of the construct.\textsuperscript{21} Notably, a key component of the interface is an \(\alpha\)-helix formed by the tail region at the C-terminal end of dsRBD2 (Figure 2.1B). This helix is tightly packed against the two dsRBDs and forms hydrophobic interactions with \(\alpha\)-helix H2 and \(\beta\)-strand S5 of both dsRBDs, with additional interactions from helix H1' of dsRBD2.\textsuperscript{21} Lastly, Sohn et al.\textsuperscript{21} noted several hydrogen bonds between both dsRBDs and the C-terminal \(\alpha\)-helix H5. This well organized arrangement of the dsRBDs in the absence of RNA is in contrast with the NMR structure of the antiviral response protein PKR, which shows a flexible linker and structural independence between its tandem dsRBDs in the apo-state.\textsuperscript{22}

Both dsRBDs of DGCR8 are essential for high affinity binding of dsRNA, \(K_D \approx 2.0 \mu M\) for pri-miR-16-1.\textsuperscript{21, 23} A mutational analysis shows Lys-561, Lys-562, and Lys-565 in dsRBD1 (found in \(\alpha\)-helix H2) and Lys-669, Arg-670, and Lys-673 in dsRBD2 (found in \(\alpha\)-helix H2') contribute favorably to the energetics of pri-miRNA binding.\textsuperscript{21} To date, no RNA bound structure of DGCR8 has been reported; however, Ryter and Schultz\textsuperscript{14} and Sohn et al.\textsuperscript{21} have proposed a binding mode based on mutational studies and homology modeling with the co-crystal structure of Xlrbpa bound to dsRNA. This model suggests that the pri-miRNA would have to bend upon binding, or that DGCR8 goes through a conformational change upon binding, or both. Recent studies demonstrate that, under native conditions, proteins routinely sample conformations in the unbound state that are
essential for bound-state function,\textsuperscript{24} indicating the likelihood that quantifying the apo-
state dynamics of DGCR8 will yield insight into the miRNA binding mechanism.

This motivates us to quantify the stability of the interactions at the interface and
determine their impact on structure and function. In an initial analysis of structure-
function relationships in DGCR8, Sohn \textit{et al.}\textsuperscript{21} performed a mutational analysis that
focused on the proposed RNA binding site, leaving the novel and intriguing domain
interface revealed by the structural work largely uninvestigated. In this study, we
provide an integrative look at the interface between the two dsRBDs and the impact of
its dynamics on the proposed interface for dsRNA binding using molecular-dynamics
(MD) simulations. Despite the ubiquitous presence of dsRBDs in nature, and the
availability of several atomic-resolution structures of dsRBD containing proteins,\textsuperscript{14, 22, 25-}
27 relatively few molecular dynamics simulations of these domains have been calculated
compared to other RNA-binding motifs and nucleic acid-protein complexes.\textsuperscript{28} In one
notable example, Castrignanò \textit{et al.}\textsuperscript{29} simulated the \textit{Drosophila} Staufen dsRBD3 free
and bound to dsRNA and showed a high degree of flexibility, even in the complex, of
the RNA recognition loops (loops 2 and 4). To date, the RNA binding protein most
extensively studied by MD simulation methods is the splicosomal protein U1A from the
U1 snRNP; however, this single-strand RNA binding protein utilizes a fundamentally
different binding mechanism mediated by ssRNA sequence.\textsuperscript{30-32} Our \textit{in silco} studies of
DGCR8 provide a starting point for understanding what appears from known atomic
structures of dsRBDS to be the unique binding mode of dsRNA by DGCR8, and lay a
foundation for future \textit{in vitro} and \textit{in vivo} experiments.
2.3 Materials and Methods:

Preliminary anharmonic normal mode (ANM) analysis on the crystal structure was done using the anistropic network model web server. The suggested parameters from the web server were used to run the simulation, which included a 15 Å interaction cutoff and a distance weight of 2.5 for the interactions between Ca atoms. Although the full set of modes reported by the server was analyzed, only the first mode is reported here.

MD trajectories were run in the AMBER 10.0 software package using the ff99SB force field. Simulations were carried out in explicit solvent represented by the SPC water model under particle mesh Ewald periodic boundary conditions. Three initial protein lengths were generated from the crystal structure of RNA-free DGCR8 (PDB code 2YT4): DGCR8-Core (505-701), DGCR8-dsRBD1 (505-583), and DGCR8-dsRBD2 (614-691). The crystal coordinates of DGCR8 are missing several residues that were built back into the starting conformations for the Core and dsRBD2 simulations. The loop formed by residues 643-648 was created in both starting conformations by restoring the residues VVPGKN using the Coot software package. For the Core simulation, residues 584-591 were also added through modeling of the sequence SEEKPKD in Coot. A number of chloride counter-ions sufficient to neutralize the net positive charge on the proteins were added, and the resulting systems were solvated such that no solute atom was within 10 Å of a box edge. This required 21,936 water molecules for the Core, 8777 for dsRBD1, and 9232 for dsRBD2. The starting configurations were energy minimized and equilibrated as previously reported. After the initial equilibration period, 250 ns of dynamics were run in an isothermal-isobaric
(NPT) simulation for each construct. Snapshots from each trajectory were stored to disk every 1.0 ps. The analysis of the trajectories was done in AMBER using the ptraj program. Molecular graphics images were created using the UCSF Chimera package. Additional analysis and visualization was accomplished in Matlab (MathWorks Natick, Massachusetts).

2.4 Results and Discussion:

The pre-organization of DGCR8’s dsRBDs seen crystallographically in the absence of RNA is in contrast to the only other previously determined structure of a protein with tandem dsRBDs, PKR, which shows a flexible linker. The well-defined structure in the linker and C-terminal of DGCR8 produces a pseudo two-fold symmetry in the Core that defines the relative orientation and spacing of the two dsRNA binding surfaces (Figure 2.1B). The novel interaction between the dsRBDs of DGCR8 seen in the crystal structure supports a new mechanism for dsRNA binding to dsRBDs and requires a quantitative understanding of the temporal and spatial dynamics in the linker, as they will directly impact the stability of the predefined spacing and orientation between the dsRNA-binding surfaces.

2.4.1 Root Mean-Square Deviation of DGCR8:

The interfacial interactions of DGCR8 were studied using MD simulations of three different constructs derived from the crystal structure of the RNA-free Core (PDB 2YT4): DGCR8-dsRBD1 (505-583), DGCR8-dsRBD2 (614-691), and DGCR-Core (505-701). Two loops are absent in the crystal structure due to low electron density and were therefore modeled back into the structure before running the simulations.
Figure 2.2: RMSDs and PDBs of the dsRBDs from DGCR8. RMSD traces show the overall stability of the dsRBDs during the MD simulations and highlight the rearrangement of the domains relative to each other in the Core simulation. (A) RMSD traces of dsRBD1 (red), dsRBD2 (green), and Core (black) simulations relative to their respective starting structure. (B) RMSD trace of the isolated dsRBD1 (red) compared to the RMSD trace of dsRBD1 from the Core simulation (grey). (C) RMSD trace of the isolated dsRBD2 (green) compared to the RMSD trace of dsRBD2 from the Core simulation (grey). In both B and C the total RMSD from the Core (black) simulation is given as a reference. Ribbon bundle from the Core simulation where dsRBD1 (red) is superimposed (D) and where dsRBD2 (green) is superimposed (E). Both bundles are created by taking structures from the simulation every 50 ns.
and Methods). Analysis proceeded following the calculation of 250 ns isothermal-isobaric (NPT) trajectories of each construct.

Protein stability was checked by analyzing the backbone RMSD from the starting crystal structure over the course of the trajectory. Each construct is stable over the simulation time scale as indicated by the plateau in the time trace of the RMSD (Figure 2.2). The lowest RMSD is seen in dsRBD1, which indicates higher stability for this domain and smaller deviation from the starting crystal structure. A slightly higher RMSD is seen in dsRBD2 indicating that this domain is more dynamic than dsRBD1. Even excluding the loop that was added to the crystal structure in order to run the simulation, dsRBD2 still shows a higher RMSD than dsRBD1. dsRBD1 is more closely related to the canonical dsRBDs (XlrBpa, Drosophila Staufen and RNase III) than dsRBD2, which could also explain the difference in the dynamics of the two domains. This difference in the dynamics of the two isolated domains could lead to slightly different binding affinities. Early in the simulations, the Core reaches a RMSD of about 3.0 Å, which then rises further to 4.5 Å after 60 ns and is maintained at that level for the rest of the simulation. As with dsRBD2, the high RMSD in the Core can be partially attributed to enhanced flexibility in the added loops that are not present in the crystal structure.

The high RMSD seen in the Core simulation is sufficient to cause concern in a single globular domain that the structure is unstable. However, DGCR8 is not a single globular domain, and the somewhat large RMSD is reasonable for a multiple-domain protein if it can be attributed to the two domains reorienting themselves relative to each other while still retaining their overall structure. Mathematically, this would tend to inflate the RMSD because no single reference structure would exist that serves well for
the RMSD calculation over the entire time course. The RMSD trace shows that the Core simulation reaches a steady state of 4.5 Å deviation from the starting crystal conformation, indicating that the Core does reach a stable state, albeit one that deviates significantly from the starting structure. Rigid-body-type reorientations of the two domains with respect to one another are distinct from, for example, complete local unfolding of secondary elements in the linker. Therefore, to draw accurate conclusions from the RMSD traces, one must further explore the limiting example the Core most closely resembles.

If the two domains in the Core simulation are stable individually, then superposition of the residues corresponding to each dsRBD in the Core should yield low RMSD values comparable to those observed in the isolated-domain simulations. In dsRBD1 (Figure 2.2B), the RMSD trace from the isolated domain simulation and the RMSD trace from the individual domain in the Core simulation superimpose extensively onto each other. In dsRBD2 (Figure 2.2C), the isolated domain simulation has slightly higher RMSD value throughout the simulation compared to the individual domain in the Core simulation, implying that features in the Core stabilize this domain.

The piecewise RMSD traces only provide information on the deviation of the individual domains being superimposed from their starting conformations in the crystal structure, therefore, to obtain information on the rest of the protein, ribbon bundles of the structure were analyzed (Figure 2.2D, E). As expected from the RMSD results, the domain being superimposed to generate each bundle does not deviate much; but the rest of the protein varies in its orientation with respect to the superimposed domain, consistent with the idea that the two domains may be reorienting relative to one another.
on the simulation timescale. Even though the linker and the non-superimposed domain fluctuate quite a bit in their relative orientation, the secondary elements of these regions are retained throughout the trajectories. These bundles rule out local unfolding of the secondary elements in the linker, as they are clearly retained. Additionally, from the ribbon bundles we find that the reason that dsRBD2 has a higher RMSD than dsRBD1 is because loop 1′ of dsRBD2 fluctuates more than loop 1 of dsRBD1 when superimposed. Thus, from the ribbon bundle structures, we have shown that the high RMSD from the Core is due not to instability of the structure, but the rigid-body movement of the domains relative to each other.

2.4.2 ANM Analysis:

We were concerned that the global reorientation of the dsRBDs seen in the simulations might have been an artifact, so we sought an independent measurement to verify the global reorientation before continuing with a more in depth analysis of the correlations of the domain from the simulation. Elastic network models provide an inexpensive alternative to traditional atomistic normal mode analysis and have been shown to accurately reproduce anisotropic displacement factors in high resolution crystals, as well as the structural transitions in the functional cycle of the chaperonin GroEL. The ANM analysis, in particular, utilizes the same starting crystal structure used in our simulation with a fundamentally different description of the forces underlying atomic displacements, making it an effective verification tool for our study. We therefore performed ANM analysis of the Core (residues 505 to 701). The largest amplitude eigenmode of the ANM calculation features correlated hinging motion in the dsRBDs, reminiscent of a butterfly flapping its wings (snapshots of structures spanned by the
**Figure 2.3:** Ribbon diagram representation of ANM analysis of DGCR8-Core. Ribbon diagrams spanning the range of motion in the largest-amplitude ANM illustrate the correlated movement of the dsRBD wings towards each other. Colorized regions highlight the sections of maximal displacement, with dsRBD1 and dsRBD2 colored identically in the given structure to guide the eye.
motion along this eigenmode are shown in Figure 2.3). The motions of the domains are centered on the α-helix formed by the C-terminal residues that packs between the dsRBDs. Keeping with the butterfly analogy, the two wings (formed by the two dsRBDs) flap in an anti-correlated movement that changes the distance between the two proposed RNA binding surfaces, rather than in a correlated twisting motion that would leave the distance between the binding surfaces unchanged. The largest amplitude motion is seen in loop 1 of both domains (residues 536 to 541 in dsRBD1 and 641 to 650 in dsRBD2), which is proximal to the proposed dsRNA-binding interface. The similarity between the ANM results and the features of the structural bundles from the MD simulation is striking. These results demonstrate that motions at the domain interface dynamically impact the conformation of the RNA-binding surface and may provide an adaptive separation distance that is necessary to allow optimal interactions with a variety of different pri-miRNAs with heterogeneous structures.15, 19, 20, 44, 45

2.4.3 Correlated Dynamics in dsRBDs:

Our initial motivation for calculating the MD trajectories was due to the unique prearrangement of the dsRBDs in the DGCR8-Core in the absence of RNA, which suggested an RNA-binding mechanism dependent on the collective arrangement of the two dsRNA-binding surfaces. As shown with the preliminary RMSD and ANM analysis, the function of these dynamics may be to adapt the separation distance and relative orientation of the two dsRNA-binding surfaces with respect to one another, so that all of the heterogeneous pri-miRNA-binding targets can all be recognized and bound with reasonably similar affinity.
One strength of MD simulations is their ability to reveal correlated dynamics, especially over long distances, through principal component analysis $^{46, 47}$ or a variety of related techniques that have been developed to monitor essential dynamics of selected degrees of freedom that are believed to be the most functionally relevant.$^{48-50}$ The application of traditional Cartesian essential-dynamics methods requires that the snapshots of the trajectory first be superimposed to remove the effects of translational and rotational diffusion and therefore also requires the existence of a unique reference frame for superimposing the entire macromolecule. As was seen in the RMSD traces (Figure 2.2), this is not possible for the overall Core, due to reorientational dynamics of the dsRBDs with respect to each other. For the individual domains, however, removal of global motions is practical, and thus allows collective analysis of the individual site dynamics discussed above.

Analysis of the Cα atomic fluctuations produces the covariance matrices shown in Figure 2.4. From Figure 2.4A and B, it is evident that extensive correlations are observed within the individual domains, as assayed in the dsRBD1 and dsRBD2 simulations, respectively. Although the collective dynamics of both domains in the Core simulation are qualitatively similar to those of the isolated domain (upper triangles, Figure 2.4, C and D), there are key differences. dsRBD1 shows few differences in collective dynamics between the Core and isolated simulations (see the difference matrix in the lower triangle of Figure 2.4C). However, there are two notable exceptions: dsRBD1 makes contact with dsRBD2 and H5 in the C-terminus through the S4-S5 loop and the C-terminal end of α-helix H2 (Figure 2.1B). These two regions of the primary structure pack against one another in the domain, but their dynamics become
Figures 2.4: Cα correlations of the dsRBDS from DGCR8. Cα correlation matrices reveal the collective backbone motions of isolated dsRBD1 (A), isolated dsRBD2 (B), dsRBD1 from the Core simulation (C), and dsRBD2 from the Core simulation (D). To highlight differences between the collective dynamics of the domains in isolation and in the Core, a difference matrix between panels A and C or panels B and D is shown below the diagonal in panels C and D, respectively. The color bar on the right shows the color scale indicating strong positive correlation (red), strong negative correlation (blue) and non-correlated motion (green) used in the figure.
significantly more correlated in the context of the Core (Figure 2.4C, lower triangle). Finally, constraint of the C-terminal residues of dsRBD1 by packing in the linker region of the Core strengthens correlation between the residues at the C-terminus of H2 and those in the N-terminal region of α-helix H1 (Figure 2.4C, lower triangle).

In contrast to dsRBD1, where the changes in correlated motion between the isolated and Core trajectories are limited to a small number of well-defined sites, the changes in dynamics of dsRBD2 are diffuse and encompass the entire domain (Figure 2.4D, lower triangle). Overall, the collective nature of the dynamics increases in the Core, with positive correlations becoming more positive and anticorrelations more negative, suggesting a global stiffening of the domain. Although these results are informative, they do not address the question of whether the collective dynamics in the Core as a whole produce functionally relevant changes in the character of the dsRNA-binding interface, because motions of the two domains relative to one another cannot be assessed by this method. To access the collective dynamics of the Core computationally, it is necessary to apply a method that analytically removes the effects of global rotational and translational motion without reference to a particular snapshot or average position.

2.4.4 Isotropically Distributed Ensemble Analysis:

Several methods exist for assessing collective dynamics computationally without first removing global rotational motion, including isotropically distributed ensemble analysis (IDE), and reorientational eigenmode dynamics. These methods have been applied successfully to describe the amplitude and timescale of internal dynamics
in the single domain proteins ubiquitin$^{51-53}$ calbindin$^{54}$ and RNA binding domain I (RBD1) from U1A$^{31}$ and have also provided insight into the collective dynamics of model RNA hairpins$^{55-57}$ and the complex between UI snRNA and U1A RBD1$^{58}$.

Here, we have chosen to apply IDE analysis to the C$\alpha$ atoms of the DGCR8-Core because it allows us to quantify the collective dynamics in the MD simulation and test whether they are consistent with the ANM predictions. The covariance matrix constructed in IDE can be diagonalized, producing three eigenmodes corresponding to overall rotational motion and N-3 ($N = 197$, the number of C$\alpha$ atoms) modes corresponding to internal motion. Plotting a parameter ($\kappa$) that describes the number of atoms significantly affected by a given mode as a function of the eigenvalues ($\lambda$) for the modes (Figure 2.5A) reveals the qualitative difference between the global and internal modes. The collectivity profile in Figure 2.5A clearly indicates that global and internal motion are separable into discrete subsets of modes (notice the significant gap between the three largest eigenvalues, corresponding to the global modes, and the fourth), allowing reconstruction of a covariance matrix of internal motions from the N-3 internal modes and their eigenvalues. The overall IDE matrix from the Core simulation is shown in the upper triangle of Figure 2.5B, and the internal motion matrix is shown below the diagonal in the lower triangle. The features of the total IDE matrix reveal the presence of strongly anticorrelated motion are of each RBD with respect to the other under the influence of the three global eigenmodes. This is consistent with the more-qualitative conclusions drawn from the bundles generated by superimposing each domain individually (Figure 2.2, D and E), and confirms that the distance separating the two
Figure 2.5: IDE analysis of DGCR8-Core. IDE analysis reveals motions correlating the conformations of dsRBD1 and dsRBD2 in the Core. (A) Mode collectivity plotted as a function of the eigenvalue for each mode reveals the separation of motion into three global reorientational modes (upper right) and the N-3 internal modes. (B) The IDE matrix from the Core simulation is represented through its cross-correlation coefficients (upper triangle), demonstrating the anticorrelated nature of the global reorientation of the two dsRBDs, with respect to each other. Subtracting the contributions from the three global modes before constructing the cross-correlation coefficients preserves only the effects of internal dynamics (lower triangle). The color bar on the right shows the color scale indicating strong positive correlation (red), strong negative correlation (blue) and non-correlated motion (green) used in the figure.
RNA-binding surfaces of DGCR8 is adapted dynamically on the timescale of these simulations.

Analysis of the internal dynamics of the Core reveals further adjustments made to the relative orientation of the two RNA-binding surfaces with respect to one another. The lower triangle of Figure 2.5B reveals a complex pattern of correlated and anticorrelated motions spanning the whole protein. Visualization of their effects can be achieved by projecting displacements along the largest-amplitude internal modes onto the starting structure from the simulation. In Figure 2.6, the projected structures with the most extreme excursions along the first internal eigenmode of the IDE analysis are represented as ribbons, along with that of the starting conformation. Displacement of the RNA-binding loops is seen, along with a slight twist of the two domains relative to each other. The central α-helix, on the other hand, remains largely unmoved under the influence of this motion, confirming its role as a pivot for the dynamics of the ensemble. Lastly, the portions of dsRBD1 and dsRBD2 in direct contact with one another are displaced in a correlated way, preserving contact, and further supporting the suggestion that direct domain-domain contacts must be maintained to facilitate the repositioning of the RNA-binding surfaces.

2.4.5 dsRBD Domain Interface:

The results we have discussed so far present an intriguing picture of the impact global motions have on the relative positions of the two RNA-binding surfaces in DGCR8, but up to now we have avoided a detailed discussion of the atomic level interactions that underlie them. As a starting point for investigating the atomistic
Figure 2.6: Ribbon diagram representation of IDE analysis of DGCR8-Core. Ribbon diagrams spanning the range of motion sampled by projecting the limiting excursions along the largest-amplitude internal eigenmode of the IDE matrix onto the starting crystal structure. Colorized regions highlight the sections of maximal displacement, with dsRBD1 and dsRBD2 colored identically in the given structure to guide the eye.
mechanism of RNA binding by DGCR8, we revisited the four hydrogen bond interactions between the C-terminal α-helix (H5) and the dsRBDs presented by Sohn et al. (Figure 2.7). Distance calculations indicate that throughout the simulations the carbonyl oxygen of Ile-575 is an average of 4.5 Å away from the hydroxyl oxygen of Ser-693. A similar average is seen for the amine nitrogen of Arg-630 to the carbonyl oxygen of Met-697 (4.2 Å). A higher average is seen for the amine nitrogen of Arg-630 with the carboxylate oxygen of Glu-701 and the carbonyl oxygen of Ile-575 with the amine nitrogen of Arg-696 (8.2 Å and 7.6 Å, respectively). Therefore, out of the four hydrogen bonds predicted from the crystal structure, only two seem likely to contribute significantly to the stability of the domain-linker interface formed by α-helix H5. Based on our MD simulations, we conclude that the dynamically preserved hydrogen bonds found between Arg-630 and Met-697; and Ile-575 and Ser-693, are the ones responsible for preserving contact between α-helix H5 and the dsRBDs.

These results led us to investigate the dynamic stability of other crystallographically observed interactions spanning the domain interface to clarify which are most involved in holding the interface together, and which simply appear to be preferred in the temporally and spatially averaged crystallographic model. In particular, we are drawn to the interface between the S4-S5 loop of dsRBD1 and H1´ of dsRBD2, both of which undergo correlated displacements in the largest-amplitude internal mode of the IDE matrix. Recall that the S4-S5 loop and the end of helix H2 were also the portions of dsRBD1 found to experience the greatest difference in their Ca correlations by Cartesian principal component analysis of the isolated domain and Core simulations. In addition to an extensive van der Waals interface between the dsRBDs, including
**Figure 2.7**: Interfacial Interactions of DGCR8-Core. Expanded views of the Core crystal structure showing the interfacial interactions between dsRBD1 (red) and dsRBD2 (green). Stabilization of the interface comes from (A) multiple hydrogen bonds formed between the C-terminal helix and either dsRBD1 or dsRBD2; (B) a variety of packing interactions, including a hydrogen bond between Asn-631 and Gly-550; and (C) a salt bridge between Asp-549 and Lys-659. All distances shown are between the indicated pair of atoms in the crystal structure. Their variation in the MD simulations is discussed in the text.
these regions of dsRBD1, further investigation of the crystal structure reveals two particularly important interactions spanning the dsRBD1-dsRBD2 interface: a hydrogen bond formed by Asn-631 and the backbone of Gly-550 (Figure 2.7B) and a salt bridged formed by Lys-659 and Asp-549 (Figure 2.7C). The hydrogen bond formed by the amide nitrogen of Asn-631 and the carbonyl in the backbone of Gly-550 is predicted from the crystal structure, with a distance of 3.6 Å. Throughout the simulation, an average distance of 5.4 Å is seen between the carbonyl oxygen of Gly-550 and the side-chain nitrogen atom of Asn-631, with a distance of less than 4.0 Å observed in 35% of the snapshots. This behavior indicates that over the timescale of the simulation, the hydrogen bond is retained, although dynamic excursions allow it to intermittently break. Additional van der Waals contacts between the side chain of Val-551 and various residues in dsRBD1 further strengthen this interface.

The interface between dsRBD1 and dsRBD2 is capped by a salt bridge between Lys-659 and Asp-549, as evidenced by the 4 Å approach of the charged moieties towards each other in the crystal structure (Figure 2.7C). The face of the imidazole ring of His-660 is orientated to provide packing interactions for the aliphatic chain of the Lys-659 side chain, further stabilizing its approach toward Asp-549. Additionally, the imidazole ring’s NH moiety is 5.9 Å away from the carboxylate of Asp-549 in the crystal, reinforcing the stabilizing transdomain interactions involving this residue. An early event in the simulation reorients the S4-S5 hairpin, moving Lys-659 and Asp-549 more than 15 Å apart from each other, but they return to an average separation of 7.9 Å by the 60 ns mark, with an approach of less than 5 Å observed in 22% of the snapshots, and remain oriented to maintain an electrostatic interaction for the remainder of the
trajectory. The salt bridge between Asp-549 and Lys-659 is likely to be vital to the protein’s function and structure, given that the nearby His-660 provides stability and has a pKa that can be titrated near physiological conditions to enhance stability.

2.5 Conclusion:

We have presented the first MD simulations of the DGCR8-Core; a dsRNA-binding protein that contains two dsRBDs in tandem. Our results show that the dsRBDs are connected through an extensive network of interactions in the dsRBD1-dsRBD2 interface and in the dynamically integrated interface of each domain with residues from the well-folded linker region and C-terminal α-helix. These interactions, including key hydrogen bonds and salt bridges, are consistent with the initial configuration in the crystal structure, and are largely preserved in a 250 ns all-atom MD trajectory. Moreover, our results show that these interfacial connections are vital for maintaining the dynamics of the protein as a whole, which serve to fine-tune the distance separating the two dsRNA-binding surfaces and their orientations with respect to one another.

More importantly, our results demonstrate which residues might be effectively targeted in mutagenesis studies aimed at disrupting the collective dynamics of the Core under the hypothesis that this will reduce the functional benefits of pre-organizing the dsRNA-binding site. It is predicted that loosening of transdomain interface will reduce the affinity of the Core for pri-miRNA because it will impede cooperative function of the two domains. Additionally, such a disruption may render the protein less competent to bind pri-miRNA transcripts preferentially over other dsRNA molecules with different
structures, providing insight into how DGCR8 selects pri-miRNA from the complex nuclear pool of partially dsRNAs.

2.6 Acknowledgments:

We thank Prof. Phil Bevilacqua and Mr. Wayne Mullinax for helpful discussions. This work was supported by startup funds from Pennsylvania State University.

2.7 References


29. Castrignano, T.; Chillemi, G.; Varani, G.; Desideri, A., Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals


Chapter 3

Dynamic Origins of Differential RNA Binding Function in Two dsRBDS from miRNA “Microprocessor” Complex


All EMSAs in this chapter were performed by Kaycee Quarles. Figure 3.3 was generated by both Kaycee Quarles and Christopher Wostenberg. The manuscript that this chapter is based on was reviewed and approved by all the authors.

3.1 Abstract:

MicroRNAs (miRNAs) affect gene regulation by base-pairing with messenger RNA and contribute to the control of cellular homeostasis. The first step in miRNA maturation is conducted in the nucleus by the “microprocessor” complex: made up of an RNase III enzyme, Drosha, that contains one dsRNA binding domain (dsRBD); and DGCR8, that contains two dsRBDS in tandem. The crystal structure of DGCR8-Core (493-720), containing both dsRBDS, and the NMR solution structure of Drosha-dsRBD (1259-1337) have been reported, but the solution dynamics have not been explored for any of these dsRBDS. To better define the mechanism of dsRNA binding and thus the nuclear maturation step of miRNA processing, we report NMR spin relaxation and MD
simulations of Drosha-dsRBD (1259-1337) and DGCR8-dsRBD1 (505-583). The study was motivated by electrophoretic mobility shift assays (EMSAs) of the two dsRBDs, which showed that Drosha-dsRBD does not bind a representative miRNA but isolated DGCR8-dsRBD1 does ($K_d = 9.4 \pm 0.4 \mu M$). Our results show that loop 2 in both dsRBDs is highly dynamic but the pattern of the correlations observed in MD is different for the two proteins. Additionally, the extended loop 1 of Drosha-dsRBD is more flexible than the corresponding loop in DGCR8-dsRBD1 and shows no correlation with loop 2, which potentially explains the lack of dsRNA binding by Drosha-dsRBD in the absence of the RNase III domains. The results presented in this study provide key structural and dynamic features of dsRBDs that contribute to the binding mechanism of these domains to dsRNA.

3.2 Introduction:

MicroRNAs (miRNAs) are a large class of gene-regulating RNAs that have been shown to contribute to the control of cellular homeostasis in multicellular organisms by regulating such biologically important processes as apoptosis, cell cycle progression, and cell-type differentiation. Mature miRNAs, roughly 22 nucleotides, regulate gene translation by base pairing with messenger RNA and it has been reported that a majority of human genes (>60%) are regulated post-transcriptionally by at least one miRNA. Disease states leading to cancer, Alzheimer’s disease, heart disease, and autoimmune diseases have been directly linked to aberrant cellular levels of specific miRNAs. Maturation of miRNA is a two step process, with each step being independent and spatially separate. The first step involves cleavage of the single-
stranded RNA (ssRNA) tail of primary miRNA (pri-miRNA) in the nucleus by the “microprocessor” complex, which contains the RNase III enzyme Drosha and its cofactor, a double-stranded RNA (dsRNA) binding protein DGCR8 (known as Pasha in *Drosophila*). After pre-miRNA is transported to the cytosol by exportin-5, the RNase III enzyme Dicer cleaves the terminal loop, thus forming mature miRNA.

Throughout the maturation of miRNA, dsRNA binding domains (dsRBDs) play vital roles in recruitment and positioning for cleavage. dsRBDs are one of the most common RNA-binding motifs, having been reported in viral-encoded proteins and in both prokaryotic and eukaryotic organisms. Also, dsRBDs are found in proteins in both the cytoplasm and the nucleus of eukaryotes, as evidenced by the compartmentalization of the first and second stages of miRNA processing in the nucleus and cytosol, respectively. The dsRBD is a ~65-68 amino acid motif that forms an αβββα topology, where the two α-helices lie on one face of 3 anti-parallel β-sheets with the other surface of the sheet being solvent-exposed.

The preferred binding partner of dsRBDs is A-form double-helix RNA, while dsRBDs show little or no affinity for ssRNA, dsDNA, and DNA-RNA hybrids. As few as 11 base pairs of dsRNA (coinciding with one turn of canonical A-form helix) have been shown to interact with a single dsRBD. Non-sequence specific recognition of dsRNA by dsRBDs is mediated by conserved positively charged residues and hydrogen-bond acceptors and donors in loop 2 (L2) and loop 4 (L4), the N-terminal region of helix 2 (H2), and additional contacts from helix 1 (H1). These positions of sequence conservation are highlighted in Figure 3.1 for each dsRBD to be discussed in depth. The dsRBD binding mode differs from the mode employed by the structurally
similar RNA recognition motifs (RRMs), exemplified by the two RRsMs of U1A, that bind ssRNA by utilizing conserved residues in the central β-strands.\textsuperscript{11, 23} In dsRBDs, the distance between loop 2 and loop 4 is hypothesized to be a critical determinant of the ability to bind dsRNA.\textsuperscript{21} The hypothesis originates from analysis of the co-crystal structure of Drosophila Staufen-dsRBD3, in which the spacing of these two loops matches the spacing between the minor and major groove of A-form RNA. The observed spacing is facilitated by the stacking of each loop on opposing faces of a conserved phenylalanine (Figure 3.1).\textsuperscript{20}

In addition to the value of static structure determination, measurement of protein dynamics through NMR spin relaxation experiments has proven to be a powerful step in defining the molecular mechanism of RNA binding by both the RRM\textsuperscript{24, 25} and the dsRBD.\textsuperscript{20, 26} NMR dynamics measurements have been reported for the two dsRBDs of the anti-viral response protein, PKR,\textsuperscript{27} and the dsRBD3 of Staufen free and bound to dsRNA (heteronuclear NOE only).\textsuperscript{20} Both showed increased flexibility in loop 2, which in the case of Staufen-dsRBD3 is retained in the bound state.\textsuperscript{20, 27} In addition, the PKR dynamics indicate that decreased dynamics in loop 2 correlate with decreased dsRNA binding since PKR-dsRBD2, which has a higher order parameter ($S^2$) in loop 2, shows a significantly lower affinity for dsRNA.\textsuperscript{27}

One of the first dsRBD containing proteins to be purified and studied was the RNase III enzyme from Escherichia coli.\textsuperscript{15, 16, 28} Generally, dsRBDs coexist with the RNase III domain in RNase III proteins to the broadest evolutionary extent, except in
Figure 3.1: Alignment of a select few dsRBDs. Alignment using ClustalW2 for the three dsRBDs from the human microprocessor complex and four other dsRBDs discussed in this chapter showing conserved residues in varying shades of gray, with black corresponding to complete conservation. Approximate location of the secondary elements is shown above the alignment with H representing an α-helix, B representing a β-strand, and L representing a loop/turn.
two mycoplasma and Dicer from *Giardia intestinalis*. RNase III enzymes can still function without dsRBDs, which is illustrated by the naturally occurring RNase III enzymes lacking a dsRBD and data from Sun *et al.* where the dsRBDs of *E. coli* RNase III were removed and function was still retained. RNase III function requires that two RNase III domains dimerize to form an active catalytic site, where each domain is responsible for hydrolysis of one strand of dsRNA, leaving a two nucleotide 3′ overhang - a notable feature of products of RNase III enzymes. At present, the structural biology of dsRBDs derived from RNase III enzymes is limited to solution structures of human Drosha and *E. coli* RNase III, a crystal structure derived from mouse Dicer and two structures of yeast Rnt1p determined by NMR and crystallography, respectively. Only one crystal structure has been reported of an RNase III dsRBD bound to RNA. No reports of NMR measured dynamics for RNase III-derived dsRBDs appear in the literature.

Drosha, a class II RNase III enzyme, is a 1,374 amino acid residue protein in its primary human isoform that contains a proline-rich region and an arginine-serine-rich region at the N-terminus, with two RNase III domains and a dsRNA binding domain at the C-terminus (Figure 3.2A). Association with the dsRNA binding protein DGCR8 is required for Drosha to specifically cleave pri-miRNA roughly eleven base pairs from the ss-ds RNA junction. To date, only the solution structure of the dsRBD of Drosha (1259-1337) has been determined (Figure 3.2C). Drosha-dsRBD adopts the αβββα secondary structure seen in all dsRBDs but also features a unique extension of loop 1 that is negatively charged (Figure 3.1 and Figure 3.2C). Although unconfirmed experimentally, sequence features important for RNA recognition are conserved in
Figure 3.2: Cartoon representation and tertiary structure of Drosha-dsRBD and DGCR8-dsRBD. Schematic representation of the primary sequence of (A) Drosha and (B) DGCR8. (C) A ribbon diagram representing the solution structure of Drosha-dsRBD (PDB 2KHX, residues 1259-1337) shows the extended loops 1 (L1) and 2 (L2). (D) A ribbon diagram representing the crystal structure of DGCR8-dsRBD1 (PDB 2YT4, residues 505-583) shows a less elongated fold of the dsRBD than Drosha-dsRBD.
Drosha-dsRBD, suggesting that Drosha retains the ability to bind dsRNA and is in fact necessary for pri-miRNA processing in vitro. Zheng and Cullen demonstrated that the arginine-rich region of Drosha is responsible for ssRNA recognition and that Drosha is also capable of recognizing terminal loops.

In order for Drosha to specifically cleave pri-miRNA, it must be associated with another dsRNA binding protein, DGCR8, which is a 773 amino acid residue protein that contains two dsRBDs in tandem at its C-terminus (Figure 3.2B). The rest of the protein is intrinsically disordered and without known function, except for a WW motif that is proposed to mediate heme-activated dimerization. The crystal structure of the Core region of DGCR8 (residues 492-720 of the human sequence) revealed that two dsRBDs are arranged in a pseudo-2-fold symmetry and packed against a well-defined secondary structure formed from the linker and the C-terminal tail of the construct. This well-organized arrangement of the dsRBDs is in contrast to the NMR structure of PKR, which shows a flexible linker and structural independence of the two dsRBDs in the apo-state. The two dsRBDs of DGCR8 contribute to high-affinity binding of pri-miRNA, $K_d= 2.1 \pm 1.1 \, \mu M$ for pri-miR-16-1. The binding affinities for the individual dsRBDs of DGCR8 have not been reported. Like Drosha-dsRBD, no NMR dynamics measurements have been reported for DGCR8-Core or its individual dsRBDs; however, we have recently reported molecular dynamics simulations of the Core and its dsRBDs that suggest a functional role for conformational dynamics in pri-miRNA binding.

Here we present both NMR and MD studies of Drosha-dsRBD, as well as NMR dynamics studies of DGCR8-dsRBD1, in order to provide a dynamic profile of these dsRBDs. Combined with binding assays utilizing the same dsRBDs, this work presents
a unified picture for the role of dynamics in binding. The present study aims to incorporate qualitative and quantitative data on structural and dynamic contributions to dsRNA binding by dsRBDs, resulting in a predictive model transferrable to other dsRBDs.

3.3 Materials and Methods:

3.3.1 Protein Preparation:

Drosha-dsRBD (1259-1337) was amplified by PCR from an ATCC plasmid (I.M.A.G.E Clone ID 5114643). The PCR product was then cloned into pET47b (Novagen), which encodes a 6X His tag and a 3C protease recognition site upstream of the cloning site, followed by transformation into BL21(DE3) competent cells. A liter growth in M9 minimal media with $^{15}$N-ammonium chloride as the only nitrogen source and either $^{12}$C- or $^{13}$C-glucose as the only carbon source produced uniformly $^{15}$N or uniformly $^{15}$N, $^{13}$C labeled protein. Expression was induced using 500 $\mu$L 1.0 M IPTG at $\text{OD}_{600} \approx 0.5$ and the cells were harvested after 3.5 hr. The cells were lysed by sonication at 4 °C. The suspension was centrifuged for 30 min at 11,500 r.p.m in a Beckman Coulter Allegra 25R using a TA-14-50 rotor. The clear supernatant was then passed over a Ni-NTA (Novagen) column and the protein was eluted with imidazole (200 mM). The His-tag was cleaved using 3C protease at 4 °C overnight while also dialyzing away the imidazole. The content of the dialysis bag was then passed over the same Ni-NTA column and the flow through was collected. The protein was concentrated and buffer exchanged using an Amicon Ultra centrifugal filter device (Millipore) that contained a PES 3,000 MWCO membrane. Drosha was buffer exchanged into 100mM cacodylate pH 7.0, 100 mM potassium chloride, 1 mM
dithiothreitol, 0.02% sodium azide, and 10% deuterium oxide. Final concentration of the $^{15}$N-NMR sample used for spin relaxation measurements was determined by UV absorption using $\varepsilon = 5600 \text{ M}^{-1}\text{ cm}^{-1}$ at 278 nm to be 380 µM.

A synthetic DGCR8 (493-720) gene was purchased from Geneart and DGCR8-dsRBD1 (505-720) was amplified by PCR. The PCR product was then cloned into pET49b (Novagen), which encodes a 6x His tag, a GST fusion tag and a 3C protease recognition site upstream of the cloning site, then transformed into BL21(DE3) competent cells. The protein expression and purification was the same as for Drosha-dsRBD, except that DGCR8-dsRBD1 was buffer exchanged into 20mM HEPES pH 7.0, 100 mM sodium chloride, 1 mM dithiothreitol, 0.02% sodium azide, 5mM β-mercaptoethanol and 10% deuterium oxide. Final concentration of the $^{15}$N-NMR sample used for spin relaxation measurements was determined by UV absorption using $\varepsilon = 4200 \text{ M}^{-1}\text{ cm}^{-1}$ at 278 nm to be 350 µM.

3.3.2 RNA Preparation:

pri-miR-16-1 DNA containing a T7 promoter sequence at the 5’ end and an inverted BsaI cut site at the 3’ end from Geneart was cloned into pUC19 (New England Biolabs) then transformed into DH5α competent cells. These cells were grown overnight in LB media at 37 °C until an OD$_{600}$ of approximately 3.75 was reached. The cells were then lysed and the DNA was purified using a Plasmid Maxi Kit (Omega). The recovered DNA was digested with BsaI overnight at 50 °C. After digestion, calf intestinal alkaline phosphatase was added and incubated another 30 min. at 37 °C to prevent self-ligation. Post digestion, the linearized DNA was extracted with phenol-chloroform and precipitated with ethanol. The pri-miR-16-1 was thereafter transcribed.
by T7 polymerase in a 10 mL reaction mixture of 25 μg/mL linearized DNA, 40 mM Tris pH 8.0, 25 mM MgCl₂, 2mM dithiothreitol, 1 mM spermidine, and 4 mM each of free NTPs (pH 8.0). The transcription incubated at 37 °C for 3 hr. The mixture was then purified by polyacrylamide gel electrophoresis in which the desired RNA band was cut out of the gel and soaked overnight at 4 °C in a TEN₂₅₀ solution. The RNA was then purified from the supernatant by ethanol precipitation and quantified by UV-Vis absorption; using ε = 1,125,400 M⁻¹ cm⁻¹ at 260 nm.

3.3.3 Electrophoretic Mobility Shift Assay (EMSAs):

The pri-miR-16-1 RNA prepared by in vitro transcription was dialyzed in 10 mM cacodylate pH 6.0 and 30 mM KCl. The proteins were dialyzed in 100 mM cacodylate pH 6.0, 100 mM KCl, and 0.35 μg/mL β-mercaptoethanol. The RNA was radiolabeled using [γ³²P]-ATP. Prior to mixing with protein, the RNA was denatured at 95 °C for 1 min. and renatured at 1 °C for 1 min. The binding reactions incubated at room temperature for 30 min. to ensure full equilibration in the presence of 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 5% glycerol, 100 μg/mL Bovine Serum Albumin, 1 mM dithiothreitol, and 0.1 mg/mL herring sperm DNA to prevent the complex from sticking in the wells. Subsequently, the binding reactions were run on a 0.25X TBE, 10% acrylamide gel at 12 V cm⁻¹ at 4 °C for 3 hr., with each lane containing 20 μCi. Signal from free and bound pri-miR-16-1 was quantified by imaging the gel on a Typhoon-9410 imager and uncertainties estimated by averaging the intensities in each lane from duplicate gels. The resulting fraction bound curves were fit to both a general model assuming N-independent and identical binding sites and a fully cooperative model,
analogous to that used for Hill analysis. Data fitting was performed using the Levenberg-Marquardt model as implemented in Matlab (MathWorks).

3.3.4 NMR Methods:

Standard triple resonance NMR techniques were used to assign the backbone resonances of Drosha-dsRBD on a Bruker Avance III 850 MHz spectrometer (chemical shifts are reported in Appendix A Table A.1). The backbone and side chain resonances of DGCR8-dsRBD1 were assigned using data from Bruker Avance III 500 and 600 MHz spectrometers (chemical shifts are reported in Appendix A Table A.2). Representative $^{15}$N-HSQC spectra for both proteins are shown in Appendix A (Figure A.4). The spin relaxation experiments were performed on a Bruker Avance III 500 MHz and 600 MHz spectrometers using standard $^{15}$N relaxation methods. All spectrometers were equipped with TCI cryoprobes for maximum sensitivity and the experiments were performed at 25 °C. Spectra were processed by NMRpipe and analyzed with SPARKY (SPARKY3.113; T.D. Goddard and D. G. Kneller, University of California, San Francisco, CA). Data were analyzed in Matlab.

3.3.5 Model-free Analysis:

Lipari-Szabo Model-free fitting was performed using the program ModelFree 4.20 with diffusion tensor fitting performed using the quadric method. The coordinates from the NMR structure of Drosha-dsRBD (2KHX) and the crystal structure of DGCR8 (2YT4) were used as structural references for diffusion tensor determination. $T_1$, $T_2$, and NOE data were fit to a model including the axially symmetric global diffusion parameters with $S^2$ and $\tau_{int}$ (model 2), except residue 516-518 of Drosha-dsRBD, which were fit with $S^2$ and $R_{ex}$ (model 3), to model internal motions.
3.3.6 Simulations:

MD trajectories were run in the AMBER 10.0 software package using the ff99SB force field. Simulations were carried out in explicit solvent represented by the SPC water model under particle mesh Ewald periodic boundary conditions. DGCR8-dsRBD1 and Drosha-dsRBD MD simulations were run as previously reported. All 10 configurations from the Drosha-dsRBD pdb file (2KHX) were used for initial analysis. Chloride counterions sufficient to neutralize the net positive charge on the proteins were added and the resulting systems were solvated such that no solute atom was within 10 Å of a box edge. The 10 configurations were energy minimized as previously reported. Since structure 6 in the pdb file had the lowest energy after minimization, the rest of the simulation was continued with this structure. Following the initial equilibration period, 250 ns of dynamics were run in an isothermal – isobaric (NPT) simulation for each construct. Snapshots from each trajectory were stored to disk every 1.0 ps. The analysis of the trajectories was done in AMBER using the ptraj program. Molecular graphics images were created using the UCSF Chimera package. Additional analysis and visualization was accomplished in Matlab. MD derived order parameters were obtained by using iRED analysis of MD trajectories averaged over 5ns windows, as previously reported.

3.4 Results:

3.4.1 Drosha-dsRBD and DGCR8-dsRBD1 Binding to pri-miR-16-1:

The present study was motivated by electrophoretic mobility shift assays (EMSAs) of pri-miR-16-1 (sequence shown in Figure 3.3A) binding by Drosha-dsRBD (1259-1337) and DGCR8-dsRBD1 (505-583). pri-miR-16-1 was chosen as the model
Figure 3.3: EMSA of pri-miR-16-1 binding to DGCR8-dsRBD1. (A) Predicted secondary structure of pri-miR-16-1, with the sequence of the mature miRNA shown in red, and the region removed by Drosha cleavage indicated through lower-case letters. (B) Representative gel showing addition of DGCR8-dsRBD1 (2-200 μM) to 0.25 nM pri-miR-16-1. (C) Fitted EMSA fraction bound as a function of DGCR8-dsRBD1 concentration with data points and uncertainties from the fitting represented by filled circles and error bars, and the best fit to the data (see text) represented as a gray line. The resulting binding affinity from the fit is $K_d = 9.4 \pm 0.4 \mu M$ with a Hill coefficient of $n = 2.34 \pm 0.01$. 
pri-miRNA since it has previously been utilized in other studies with DGCR8 and Drosha. We observe no binding of pri-miR-16-1 by Drosha-dsRBD, even at 250 μM (see Appendix A Figure A.1). Therefore, we conclude that Drosha-dsRBD does not possess intrinsic binding affinity for pri-miR-16-1 in the absence of the Drosha-RNase III domains and is unlikely to bind other pri-miRNAs as well. In comparison, the EMSAs of DGCR8-dsRBD1 demonstrates that DGCR8-dsRBD1 binds to pri-miR-16-1 with a $K_d = 9.4 \pm 0.4$ μM (Figure 3.3B,C), which is less than 3-fold weaker binding compared with DGCR8-Core under the same conditions ($K_d = 3.7 \pm 1.0$; Appendix A Figure A.2). The steep slope of the transition contributes to a measured Hill coefficient of $n = 2.34 \pm 0.01$ that is consistent with cooperative binding of more than one DGCR8-dsRBD1 per pri-miR-16-1. The steep slope could also be explained by loss of free protein during equilibration or dissociation during gel electrophoresis, but gels run with lower incubation times show multiple intermediates reinforcing the assumption that the steep slope is indicative of cooperative binding (Quarles and Showalter, unpublished results).

3.4.2 NMR Spin Relaxation:

Information on picosecond-nanosecond time scale dynamics can be obtained from backbone $^{15}$N NMR spin relaxation data. We have measured $^{15}$N $T_1$, $T_2$, and $({^1H}) - ^{15}$N NOE NMR spin relaxation at 500 MHz and 600 MHz field strength for both Drosha-dsRBD and DGCR8-dsRBD1 in the RNA-free state (Figure 3.4). Analysis of the spin relaxation data using the quadric method indicates that Drosha-dsRBD is more anisotropic in solution having $D_{||}/D_\perp = 1.40$ and $\tau_{iso} = 6.29$ ns compared to $D_{||}/D_\perp = 1.29$ and $\tau_{iso} = 7.20$ ns for DGCR8-dsRBD1. In both cases, the anisotropy of these domains is significantly higher than previously reported for the isolated dsRBDs of PKR,
**Figure 3.4:** $^{15}$N spin relaxation data of Drosha-dsRBD and DGCR8-dsRBD1. $^{15}$N spin relaxation data for (A) Drosha-dsRBD and (B) DGCR8-dsRBD1 collected at 500 MHz (blue and red, respectively) and 600 MHz (gray). The data shows that the extended loop 1 of Drosha-dsRBD is more flexible than that of DGCR8-dsRBD1 on the ps-ns timescale. Loop 2 is also seen to be highly dynamic in both proteins. The secondary structure elements for the respective dsRBDs, as well as the positions of loop 1 and loop 2, are represented as colored bars above the plots.
$D_{\parallel}/D_{\perp} \approx 1.06$ ($\tau_{\text{iso}}$ was not reported).\textsuperscript{27} This trend is qualitatively observed in the ribbon diagrams, where DGCR8-dsRBD1 (Figure 3.2D) is less elongated than Drosha-dsRBD (Figure 3.2C).

Complete model-free analysis of the spin relaxation data yields an estimation of the amplitude of ps-ns timescale backbone conformational dynamics through the generalized order parameter $S^2$ (Figure 3.5).\textsuperscript{59} Loop 2 of both Drosha-dsRBD and DGCR8-dsRBD1 show increased dynamics compared to the rest of the domain, as expected based on Staufen-dsRBD\textsuperscript{320} and PKR\textsuperscript{42} apo-dynamics. Even though loop 2 is highly flexible in both domains under investigation, the C-terminal residue of the region, which corresponds to Tyr-1298 and Phe-542 in Drosha-dsRBD and DGCR8-dsRBD1, respectively, shows a high order parameter compared with its neighbors. These residues correspond with the highly conserved aromatic residue that is hypothesized to be a critical determinant of dsRNA binding by dsRBDs, since it maintains the spacing between loop 2 and loop 4.\textsuperscript{21} In addition, the high order parameter indicates that the conserved aromatic residue is very rigid on the fast timescale.

Unlike Staufen-dsRBD\textsuperscript{3},\textsuperscript{20} there is no indication of increased dynamics in loop 4 for the apo-state of either dsRBD, but this is consistent with data from PKR.\textsuperscript{27} In all structures used for alignment, except Staufen-dsRBD3 and DGCR8-dsRBD2 (Figure 3.1), there is a conserved serine or threonine in loop 4 forming an N-cap for H2,\textsuperscript{61} and likely restricting the dynamics of the loop as a result. Instead of a serine or threonine, Staufen-dsRBD3 has a glycine in this position, which explains the increased flexibility for this region as glycine is a poor N-cap residue\textsuperscript{61} and is intrinsically more flexible.
Figure 3.5: Order parameters of Drosha-dsRBD and DGCR8-dsRBD1. Order parameter ($S^2$) plots for (A) Drosha-dsRBD and (B) DGCR8-dsRBD1 show that loop 2 in both proteins and loop 1 of Drosha-dsRBD are the most dynamic regions of the domains. Experimental data (blue and red lines for Drosha-dsRBD and DGCR8-dsRBD1, respectively) is plotted against MD predicted order parameters (gray). The secondary structure elements for the respective dsRBDs, as well as the positions of loops 1 and 2, are represented as colored bars above the plots. $S^2$ is represented colorimetrically in the MD derived ribbon bundles for (A) Drosha-dsRBD and (B) DGCR8-dsRBD1, with passage from dark blue and dark red towards yellow indicating increased flexibility, based on the experimental order parameters. The conserved aromatic residues (Tyr-1298 of Drosha-dsRBD and Phe-542 of DGCR8-dsRBD1) are shown in space-filling mode to demonstrate how these residues are orientated to preserve the spacing between loop 2 and loop 4. Both bundles are created by taking the structures from the simulation every 50 ns and superimposing them to remove translation and rotation of the center of mass.
The asparagine residue found in DGCR8-dsRBD2 should be capable of forming a strong N-cap and, in this sense, is a conservative mutation from the consensus. Our data indicates that Staufen-dsRBD3 is a unique case among dsRBDs that does not contain a good N-cap residue for helix 2, which causes the loop preceding (loop 4) to have increased flexibility.

The key difference in the order parameter profiles of the two dsRBDs investigated in this study is that the Drosha-dsRBD order parameter profile (Figure 3.5A) indicates that loop 1 is the most flexible region of the domain, while this same loop in DGCR8-dsRBD1 shows no increased flexibility in comparison to the rest of the domain. This trend is likely due to the addition of five amino acid residues in the loop of Drosha, which is a unique feature of Drosha-dsRBD compared with other dsRBDs (Figure 3.1) and is likely the cause for the increase in anisotropy of the domain compared to DGCR8-dsRBD1 and the two dsRBDs of PKR. Additionally the greater flexibility of loop 1 in Drosha-dsRBD is likely disruptive to Drosha-dsRBD’s ability to bind dsRNA unassisted.

**3.4.3 Molecular Dynamic Simulations:**

Two MD simulations were run: DGCR8-dsRBD1 (505-583) derived from the crystal structure of DGCR8-Core (PDB 2YT4) and one derived from the lowest energy structure, with respect to the AMBER ff99SB force field, in the solution ensemble of Drosha-dsRBD (PDB 2KHX, 1259-1337). The DGCR8-dsRBD1 trajectory has previously been reported, but the analysis reported in this study is novel (except where noted). The Drosha-dsRBD trajectory however has not previously been reported. Protein stability in the 250 ns simulations was verified by analyzing root mean-square
Figure 3.6: RMSD of Drosha-dsRBD and DGCR8-dsRBD1. The global dynamics of Drosha-dsRBD and DGCR8-dsRBD1 observed in MD simulations indicates qualitative differences between the two domains. The RMSD traces of Drosha-dsRBD (blue line) and DGCR8-dsRBD1 (red line) demonstrate that both proteins are stable over the MD simulations, although Drosha-dsRBD is more dynamic overall, as seen in the MD structure bundles found in Figure 3.5.
deviation (RMSD) from the starting structure over the course of the trajectories (Figure 3.6). The RMSD of DGCR8-dsRBD1 (as previously reported) was very low (1.5 Å) for a large majority of the simulation (Figure 3.6, red line), indicating that this construct is highly stable. The RMSD of Drosha-dsRBD is higher than DGCR8-dsRBD1 (Figure 3.6, blue line) but eventually plateaus at 3.0 Å, demonstrating that this construct is also stable over the simulation timescale. The higher RMSD seen in the Drosha-dsRBD simulation is typical of a simulation initiated from an NMR solution structure rather than a crystal structure, as was used for DGCR8-dsRBD1. Additionally, Drosha-dsRBD loop 1 and loop 2 are larger and more dynamic than the same loops in DGCR8-dsRBD1, which inflates the calculated RMSD by roughly 1.0 Å (data not shown). As final evidence that the high RMSD is not due to local unfolding of secondary elements or tertiary contacts, ribbon bundles for both simulations were generated from the trajectories (Figure 3.5). The ribbon bundles clearly show that the overall folds are retained throughout both simulations. Additionally, the ribbon diagrams show the increased flexibility of loop 2 and loop 4 in Drosha-dsRBD compared with those in DGCR8-dsRBD1 throughout the simulation.

**3.4.4 MD-Derived Order Parameters:**

The generalized Model-free order parameter $S^2$ was computationally predicted using iRED analysis of molecular dynamics (MD) trajectories averaged over 5ns windows (Figure 3.5, gray lines). Qualitatively, the same global trends observed in the experimental data are reproduced computationally in both Drosha-dsRBD and DGCR8-dsRBD1 (Figure 3.5). The notable exception is that loop 3 of DGCR8-dsRBD1 is computationally predicted to be more dynamic than the adjacent secondary structural
elements, which is not observed experimentally. Loop 3 of DGCR8-dsRBD1 has been shown computationally to mediate interfacial interactions with dsRBD2 in the Core structure of DGCR8. The interfacial interactions cause loop 3 of DGCR8-dsRBD1 to be bent towards dsRBD2 in the crystal structure, a conformation enabled by the positive Φ angle observed for Asp-549 in the crystal structure. This bend of loop 3 is not observed in the solution structure of isolated DGCR8-dsRBD1 (PDB 1X47). Therefore the disagreement between experimental and computational data for loop 3 of DGCR8-dsRBD1 is likely due to the crystal structure used for the MD simulation and the solution structure of DGCR8-dsRBD1 being different for this region.

Another observation to note is that for Drosha-dsRBD, there is an offset in the baseline between the experimental and computational order parameter profiles which is not observed in DGCR8-dsRBD1. Intriguingly, it is the experimental order parameters that adopt a low average value compared to the MD, and compared with both sets of order parameters for DGCR8-dsRBD1, and not the other way around. The generally high order parameters seen for the secondary elements in Drosha-dsRBD support the conclusion that the conformation of the protein in the MD simulation was dynamically stable, but substantially drifted from the initial configuration during equilibration, producing the high RMSD observed (Figure 3.6).

### 3.4.5 Correlated Dynamics in the dsRBDS:

A major advantage of running MD simulations is the retention of dynamic correlations in the dataset. Such correlations can be extracted through principal component analysis or a variety of related techniques that have been developed to monitor dynamics that are believed to be functionally relevant. Analysis of Cα
Figure 3.7:  $\alpha$ correlations of Drosha-dsRBD and DGCR8-dsRBD1.  $\alpha$ correlation matrices reveal the collective backbone motions of (A) Drosha-dsRBD and (B) DGCR8-dsRBD1.  The color bar on the right shows the scale indicating strong positive correlation (red), strong negative correlation (blue), and non-correlated motion (green).  Labels above each panel indicate the location of secondary structural elements within the sequence.
atomic fluctuations in our MD trajectories produces the covariance matrices shown in Figure 3.7. Note the Cα correlation for DGCR8-dsRBD1 has already been reported, but the analysis discussed below is novel. The correlation plots for Drosha-dsRBD and DGCR8-dsRBD1 show similar features, with the features in Drosha-dsRBD indicating a greater absolute magnitude of correlation for a higher percentage of the residues than is the case for DGCR8-dsRBD1. One exception is the positive correlation between loop 1 and loop 2 in DGCR8-dsRBD1, which is absent in Drosha-dsRBD. Previous studies have shown that both loop 2 and helix 1 are involved in binding dsRNA, so it makes sense that loop 2 and loop 1 (which is connected to helix 1) show correlated motions in the apo-state that would facilitate dsRNA binding. It has already been mentioned that loop 1 of Drosha is structurally and dynamically unique compared to other dsRBDs and the correlated dynamics further expand on this. The lack of correlations between the two loops in Drosha-dsRBD could further contribute to Drosha-dsRBD not being able to bind dsRNA in the absence of the RNase III domains.

In both domains under investigation, anti-correlated motions exist between loop 2 and helix 2, with only minor anti-correlations seen between loop 2 and loop 4. Hence, the pivot for the anti-correlated motions observed between loop 2 and helix 2 may be loop 4 itself and the contacts it makes with the C-terminal region of loop 2 and the N-terminal region of sheet 2. Moreover, this yields an alternative explanation for the importance of the conserved aromatic residue (Tyr-1298 and Phe-542 for Drosha-dsRBD and DGCR8-dsRBD1, respectively). The role of a static spacer could easily be filled by a bulky aliphatic side chain, but such a residue would provide a highly flexible stacking surface compared to the rigid ring of a phenylalanine or tyrosine residue. The
rigidity of the aromatic moiety from the phenylalanine or tyrosine residue may be preferred evolutionarily because of its enhanced ability to transduce dynamic correlations between its two stacking partners.

3.5 Discussion:

Both dsRBDs studied are part of the microprocessor complex, which is involved in the nuclear maturation of miRNA, but the binding data presented here points to different roles for the two domains in the complex. The results presented in this study identify key structural and dynamic features of dsRBDs that contribute to the binding mechanism of these domains to dsRNA. Previous studies have shown extensive cooperativity between the dsRBDs of the protein RDE-4, which is involved in the processing of long continuously double-stranded small-interfering RNAs. Especially when considered in light of the correlated conformational dynamics our previous MD studies of DGCR8-Core have revealed, this work points to a central role for cooperative complex assembly in the microprocessor as a means of establishing specificity in the pri-miRNA cleavage reaction.

3.5.1 Intrinsic Binding Affinity of the dsRBDs:

The slightly weaker binding of DGCR8-dsRBD1 reported here ($K_d = 9.4 \pm 0.4 \mu M$) compared to DGCR8-Core ($K_d = 3.7 \pm 1.0 \mu M$; Appendix A Figure A.2) is noteworthy, as it suggests a dominant role for dsRBD1 in RNA binding, with dsRBD2 present to fine-tune the interaction. This is similar to the respective roles of PKR-dsRBD1 and PKR-dsRBD2; albeit PKR-dsRBD1 binds ~30-fold weaker than the two domains in tandem.

Analysis of the EMSA results with a binding model that assumes N-independent and identical sites on pri-miR-16-1 for DGCR8-dsRBD1 resulted in poor fit quality (see
Appendix A Figure A.3) that was completely relieved through the assumption of fully cooperative binding by more than one copy of the protein. The result that more than one dsRBD can be loaded onto a single pri-miRNA is unsurprising. The construct possesses approximately three full turns of (interrupted) A-form helix (Figure 3.3A), and there are at least three dsRBDs in the microprocessor (assuming a 1:1:1 DGCR8:Drosha:miRNA stoichiometry). Hill analysis of binding data in the Supporting Information of Sohn et al. also points to the possibility of cooperative loading of more than one DGCR8-Core molecule onto the same pri-miRNA. DGCR8-dsRBD1 packs against dsRBD2 in the crystal structure of DGCR8-Core, for which we have previously reported dynamic correlations between the two dsRBDs by MD. The apparent cooperative loading of multiple DGCR8-dsRBD1 molecules onto pri-miR-16-1 could be a manifestation of the intrinsic capability of the domain to assemble and further supports our hypothesis that collective positioning of the dsRBDs in the microprocessor imparts some or all of the specificity inherent in the endonuclease activity of the complex. Further studies will be needed to quantitatively define the stoichiometry in the complex.

While our results show DGCR8-dsRBD1 possesses intrinsic RNA binding affinity, we also observe a lack of pri-miRNA binding in vitro for the isolated Drosha-dsRBD that is likely due to both structural and dynamic features that are inconsistent with those of other studied dsRBDs known to bind dsRNA. First, dsRBDs known to bind dsRNA, like PKR-dsRBD and Staufen-dsRBD, generally contain three lysine residues that are on the N-terminal part of helix 2 and positioned between helix 1 and the β-sheet on the protein surface (Figure 3.1). Even though Drosha-dsRBD lacks the conserved lysine residues, Meuller et al. show a positively charged region on the Drosha-dsRBD
surface which could facilitate the binding of the negatively charged phosphate backbone in this region, albeit in an atypical manner compared with models from dsRNA bound to dsRBDs in the literature.\textsuperscript{31} Second, loop 1 of Drosha-dsRBD is five residues longer than other dsRBDs, and this increase in size causes increased flexibility in the loop along with increased anisotropy of the domain (Figure 3.4 and Figure 3.5). The increased dynamics of loop 1 in Drosha-dsRBD, which is negatively charged, allows for it to sample conformational space that potentially would be occupied by dsRNA, thus hindering binding due to charge repulsion. Also, in contrast to DGCR8-dsRBD1 (Figure 3.7), loop 1 of Drosha-dsRBD shows no correlated motions with the rest of the protein construct, which would be an effective means of maintaining its distance from the presumptive binding surface if they were present. However, our data do not rule out the possibility that Drosha-dsRBD contributes to dsRNA binding in more inclusive constructs containing the RNase III domains, and ongoing studies in our laboratory aim to address this possibility.

3.5.2 Dynamic Profile of Drosha-dsRBD and DGCR8-dsRBD1:

Dynamically, Drosha-dsRBD is not a total anomaly among dsRBDs: it shows increased flexibility in loop 2 like Staufen-dsRBD3,\textsuperscript{20} PKR-dsRBD1 and PKR-dsRBD2,\textsuperscript{27} and DGCR8-dsRBD1. The increased flexibility of loop 2 has been hypothesized to be critical for binding, since the loop needs to rotate toward the RNA in order to interact with the 2'-hydroxyl groups from the minor groove of dsRNA.\textsuperscript{20} In the case of PKR-dsRBD2, the higher order parameter in loop 2 compared to PKR-dsRBD1 correlated with a much lower affinity of dsRNA.\textsuperscript{27} While being important, increased flexibility of
loop 2 is not by itself a sufficient determinant of dsRNA binding, since Drosha-dsRBD which possesses this feature does not bind dsRNA.

Even though loop 2 is dynamic, the C-terminal region of the loop in the protein still remains in close proximity to loop 4. Our data indicates that loop 4 is not dynamic on the fast timescale, which is consistent with data from PKR, but differs from Staufen-dsRBD3. Loop 4 is not dynamic in these cases due to a conserved serine or threonine that forms a good N-cap for helix 2, while Staufen-dsRBD3 contains a glycine in this position that causes increased flexibility. However, the added dynamics of loop 4 in Staufen-dsRBD3 is not detrimental to binding since it binds dsRNA with micromolar affinity.

3.5.3 Distance Between Loop 2 and Loop 4:

The overall good agreement between the experimental and computational order parameters for the two dsRBDS led us to further investigate the molecular dynamic simulations for structural and correlated features of the two domains. Chang and Ramos hypothesized that a critical determinant of RNA binding by dsRBDS is the distance separating loop 2 and loop 4, which corresponds with the spacing between the minor and major groove of A-form RNA. In the case of Staufen-dsRBD3, the distance between the two loops is maintained by a phenylalanine, which based on alignment is strongly conserved (as either phenylalanine or tyrosine) and is present in both DGCR8-dsRBD1 and Drosha-dsRBD (Figure 3.1A). As previously mentioned, the order parameter for Phe-542 in DGCR8-dsRBD1 and Tyr-1298 in Drosha-dsRBD is high in comparison to its neighbors, indicating that these residues are rigid and could potentially be involved in maintaining the spacing between the aforementioned loops.
As a further investigation, we calculated the distance between α-carbons in loop 2 and loop 4 found in our MD trajectories. For DGCR8-dsRBD1, Glu-540 in loop 2 maintains an average distance from Ala-558 and Ser-559 of 4.7 Å and 4.4 Å, respectively. More than 99% of the time, Glu-540 is within 6 Å of both of these residues. In Drosha, the average distance between Thr-1297 and Pro-1315 is 4.7 Å and they are kept within 6 Å of each other 99% of the time as well. The order parameters for the residues in loop 2 (Glu-540 and Thr-1297 for DGCR8-dsRBD1 and Drosha-dsRBD, respectively) indicate that these residues are dynamic while the residues in loop 4 are not, but the distance between the residues in loop 2 and loop 4 are maintained on average throughout the simulation. In both cases, no hydrogen bonds or salt bridges are formed between the residues mentioned to preserve the close proximity of the atoms. Instead, for DGCR8-dsRBD1, a salt bridge is formed between Glu-540 in loop 2 and Lys-567 in helix 2. This salt bridge is evident in the crystal structure with the charged moieties being 4.8 Å apart. Throughout the simulation, the average distance of the two charged atoms is 5.8 Å. A similar salt bridge is not seen in Drosha-dsRBD since in the Drosha-dsRBD structure a glycine residue is present instead of a lysine. In fact, the Lys-567 of DGCR8-dsRBD is not a conserved residue in the other dsRBD sequences either (Figure 3.1); thus this would only help to explain the close proximity of loop 2 and loop 4 observed in DGCR8-dsRBD, but not dsRBDs as a whole.

The conserved aromatic residue hypothesized to be responsible for maintaining the spacing between these two loops is observed to potentially be a pivot for anti-correlated motions between loop 2 and helix 2, which caused it to have a high order parameter...
compared to neighboring residues and to maintain the loop 2 to loop 4 spacing. In the case of DGCR8-dsRBD1, the additional salt bridge observed between Glu-540 and Lys-567 further controls the amount of correlation and dynamics of these two regions. The anti-correlated motions of these regions of the domain are potentially responsible for the ability of dsRBDS in the miRNA maturation pathway to bind a large variety of heterogeneous dsRNA precursors. Further NMR spin relaxation and MD simulations are needed to verify the generality of our conclusions and the predictive quality of our model for the effect of dynamics on dsRNA binding.

3.6 Acknowledgment:

We thank Prof. Phil Bevilacqua for helpful evaluation of the manuscript and members of the Bevilacqua laboratory for assistance in initiating the EMSA assays.

3.7 References:


Chapter 4

The Role of the Dicer-dsRBD on Processing Small Regulatory RNAs

[This chapter represents a manuscript in progress with the following author list
Christopher Wostenberg, Jim Cole and Scott A. Showalter]

All AUC experiments were performed by Jim Cole.

4.1 Abstract:

One of the most exciting recent developments in RNA biology has been the discovery of small non-coding RNAs that affect gene expression through the RNA interference (RNAi) mechanism. Two major classes of RNAs involved in RNAi are small interfering RNA (siRNA) and microRNA (miRNA). Dicer, an RNase III enzyme, plays a central role in the RNAi pathway by cleaving precursors of both of these classes of RNAs to form mature siRNAs and miRNAs, which are then loaded into the RNA-induced silencing complex (RISC). miRNAs and siRNAs precursors are quite different structurally; miRNA precursors are short, imperfect hairpins while siRNA precursors are long, perfect duplexes, nonetheless Dicer is able to process both. Dicer, like the majority of RNase III enzymes, contains a dsRNA binding domain (dsRBD), but the data are sparse on the exact role it plays in the mechanism of Dicer binding and cleavage. To further explore the role of human Dicer-dsRBD, we determined its binding affinity to various RNAs modeling both pre-miRNA and pre-siRNA. Our study shows that Dicer-
dsRBD binding is influenced by the presence of a terminal loop, which is observed in miRNA precursors. In addition, NMR spin relaxation and MD simulations provide an overview of the role dynamics contribute to binding. We compare this current study with our previous studies of the dsRBDs from Drosha and DGCR8, to give an overall mechanistic view of dsRBD binding of dsRNA.

4.2 Introduction:

In the past decade there has been a surge of interest in the role of small regulatory RNAs, most notably microRNAs (miRNAs) and small interfering RNAs (siRNAs), in gene regulation. Both of these RNA classes function in RNA interference (RNAi) by affecting gene translation through base pairing with messenger RNA (mRNA) via their association with Argonaute-(Ago) family proteins. The roles of RNAi include defense against viruses, regulation of development, and maintenance of cellular homeostasis. siRNAs can be derived either endogenously from repetitive sequences or exogenously from viral RNAs, whereas miRNAs are only endogenously transcribed. Even though the biosynthesis of siRNAs and miRNAs start at different points, they are both processed by the RNase III enzyme Dicer into ~21nt RNAs prior to associating with Ago-family proteins forming the RNA-induced silencing complex (RISC). From the literature it is unclear the exact mechanism Dicer utilizes to discriminate between dsRNA targets and process them into functional regulatory RNAs.

The Dicer processing step common to both RNAi pathways is intriguing biochemically because of the extensive structural differences between miRNA and siRNA precursors. Dicer, a class III RNase III enzyme, is a 1922 amino acid residue
protein in humans that contains a helicase domain, a domain of unknown function (DUF283), PAZ domain on the N-terminal side of the RNase III domains, and a dsRNA binding domain (dsRBD) on the C-terminal side (Figure 4.1A). The mechanism of action for RNase III enzymes requires that two RNase III domains dimerize forming an active catalytic site, where each domain is responsible for the hydrolysis of one strand of dsRNA, leaving a characteristic 3′ two-nucleotide overhang product. Based on crystallographic data from *Giardia intestinalis* the PAZ domain of Dicer binds the 3′ two-nucleotide overhang and positions the catalytic sites 65 Å away from the overhang, corresponding with roughly 25 base-pairs of A-form RNA, the length of mature siRNAs in *Giardia*. Dicer from *Giardia* is a minimal construct because many domains that human Dicer contains, i.e., the helicase domain, the DUF, and the dsRBD, are absent. A new model for the Dicer mechanism was proposed for higher eukaryotes in the recent Cryo-EM reconstruction of human Dicer. As in *Giardia*, human Dicer is proposed to recognize the 2 nt overhang on the 3′ end through interaction with its PAZ domain, while the helicase on the opposite side of the enzyme is utilized in an ATP-dependent manner to retain the dsRNA for processive cleavage of long siRNA precursors. The reconstruction of human Dicer positions the dsRBD adjacent to the RNase III domains and on the opposite side of the “ruler” from the PAZ domain, but no discussion is provided by the authors of a mechanistic role for the dsRBD.

In most cases, dsRBDs co-exist with the RNase III domain to the broadest evolutionary extent, although function without the a dsRBD has been demonstrated. The isolated dsRBD from human Drosha, another RNase III enzyme
Figure 4.1: Cartoon representation, tertiary structure, and primary sequence of Dicer-dsRBD. (A) Schematic representation of the primary sequence of Dicer with approximate location of domains above the cartoon. (B) A ribbon diagram representing the crystal structure of mouse Dicer-dRBD (PDB 3C4B, residues 1833 to 1900), which is 100% identical to the human sequence (residues 1849 to 1916). (C) Primary sequence of human Dicer-dsRBD (residues 1850 to 1922), with the approximate location of secondary elements shown above the sequence, with H representing an α-helix, B representing a β-sheet and L representing a loop/turn.
in the miRNA maturation pathway, is unable to bind dsRNA due to structural features and dynamics.\textsuperscript{16} The mouse Dicer-dsRBD, whose protein sequence is one hundred percent identical to human Dicer-dsRBD, shows the canonical structure for dsRBDs.\textsuperscript{17} One study shows that a human Dicer construct lacking the dsRBD is able to bind pre-miRNA and pre-siRNA, but the initial rate of cleavage was greatly reduced compared with wild-type Dicer, suggesting that the dsRBD is critical for function.\textsuperscript{18} Another study indirectly suggests that the Dicer-dsRBD with part of the RNase IIIb domain is capable of binding dsRNA, as it inhibits full length Dicer binding.\textsuperscript{19} While this data suggests Dicer-dsRBD is competent to bind, no direct demonstration of such activity exists and the role dsRBD-mediated binding might play in the Dicer mechanism remains poorly defined.

To further test the role of the Dicer-dsRBD, we expressed it in isolation and have determined its binding affinity to various RNAs modeling pre-miRNA and pre-siRNA. Our previous work allows us to compare and contrast the binding affinity and dynamics of the isolated dsRBD of Drosha and the first dsRBD of its cofactor, DGCR8, showing that non-canonical loop 1 of Drosha-dsRBD has increased dynamics that hindered dsRNA binding (chapter 3).\textsuperscript{16} Similar to our previous study, Dicer-dsRBD binding ability is determined by EMSAs, while dynamics are studied with both NMR spin relaxation and MD simulations. We conclusively show that Dicer-dsRBD in isolation is able to bind dsRNA and has similar dynamics to DGCR8-dsRBD1. In addition, Dicer-dsRBD binding is only minimally influenced by the length of the RNA and that a ss-ds junction caused by a large terminal loop or ssRNA tail aids in positioning Dicer-dsRBD on the RNA. The dynamic profile of Dicer-dsRBD presented in this study, together with results in the
literature, start to provide a clear picture of how dynamics contribute to the dsRBD binding mechanism.

4.3 Materials and Methods:

4.3.1 Protein Preparation:

A synthetic Dicer gene was purchased from Geneart and a Dicer-dsRBD construct (1850-1922) was PCR amplified. The PCR product was then cloned into pET47b (Novagen), which encodes a 6X His tag and a 3C protease recognition site upstream of the cloning site. Next, the plasmid was transformed into BL21 (DE3) competent cells. For NMR experiments, the cells were grown at 37 ºC with shaking in a liter of M9 minimal media with [15N] ammonium chloride as the only nitrogen source and either [12C]- or [13C] glucose as the only carbon source to produce uniformly 15N or uniformly 15N, 13C- labeled protein. The cells were induced with 500 µL of 1.0 M IPTG when OD600 was between 0.5 and 0.6, then harvested after 3.5 hrs. For EMSAs, the cells were grown in a 500 mL of LB media and induced with 250 µL of 1.0 M IPTG when the OD600 was between 0.8 and 1.0, then harvested after 3.5 hrs. The cells were lysed by sonication at 4 ºC and the resulting suspension was centrifuged for 30 min. at 11,500 rpm in a Beckman Coulter Allegra 25R using a TA-14-50 rotor. The clear supernatant was then passed over a Ni-NTA (Novagen) column, and the protein was eluted with imidazole (200 mM). The His tag was cleaved using 3C protease at 4 ºC overnight while also dialyzing away the imidazole. The content of the dialysis bag was then passed over the same Ni-NTA column, and the flow-through was collected. The protein was concentrated and buffer exchanged using an Amicon Ultra centrifugal filter device (Millipore) that contained a PES 3000 MWCO membrane. The final buffer was 50 mM
cacodylate, pH 6.0 and 50 mM potassium chloride. The final concentration of the protein was determined by UV absorption using a $\varepsilon = 2800 \text{ M}^{-1} \text{ cm}^{-1}$ at 278 nm.

4.3.2 RNA Preparation:

To test the ability of the Dicer-dsRBD to discriminate between miRNA and siRNA precursors, it was necessary that our model pre-miRNA contained a natural nucleotide sequence with a 2 nt 3' overhang. This necessitates that a different method than the standard T7 transcription be used, because the standard procedure requires a tandem G sequence on the 5' of the RNA. Therefore, pre-miR-16-1 DNA containing the sequence for a self cleaving hammerhead ribozyme and a T7 promoter sequence on the 5' end and an inverted BsaI cut site at the 3' end was purchased from Integrated DNA Technologies (IDT) as a sense strand (sequence 5' to 3': GTC AGA ATT CTA ATA CGA CTC ACT ATA GGG AGC GTG CTG CTA CTG CTG ATG AGC GCG AAA GCG CGA AAG GAT TCC GAA AGG GAT CCT ATA GCA GCACG) and an antisense strand (sequence 5' to 3': CTG CGC ATG CGG TCT CCT TCA GCA GCA CAG TTA ATA CTG GAG ATA ATT TTA GAA TCT TAA CGC CAA TAT TTA CGT GCT GCT ATA GGA TCC C). The strands were PCR extended to be self complimentary, then the dsDNA was inserted into pUC19 (New England Biolabs). Figure B.1 in Appendix B shows a schematic of this construct. Next, the plasmid containing the construct was transformed into DH5α competent cells. The cells were grown overnight in LB media at 37 °C to an OD$_{600}$ of approximately 3.0. The cells were lysed, and the DNA was purified using a Plasmid Maxi Kit (Omega). The recovered DNA was digested with BsaI overnight at 50 °C. After digestion, calf intestinal alkaline phosphatase was added and incubated another 30 minutes at 37 °C to prevent self-ligation. Post-digestion, the linearized DNA
was extracted with phenol-chloroform and precipitated with ethanol. The RNA was then transcribed by T7 polymerase in a 10 mL reaction mixture of 25 µg/mL linearized DNA, 40 mM Tris, pH 8.0, 25 mM MgCl₂, 2mM dithiothreitol, 1 mM spermidine, and 4 mM of each free NTP (pH 8.0) at 37 °C for 3 hrs. Transcription yielded three major products on a denaturing polyacrylamide gel: pre-miR-16-1(65nt), hammerhead ribozyme (58nt) and uncleaved RNA (124nt). The pre-miR-16-1 band was cut out of the gel and soaked overnight at 4 °C in a TEN₂⁵₀ solution. The RNA was then purified from the supernatant by ethanol precipitation and quantified by UV-Vis absorption; using $\varepsilon = 666,700 \text{ M}^{-1} \text{ cm}^{-1}$ at 260 nm.

DNA for the top and bottom strands for the 33 bp (top strand sequence 5’ to 3’: GGA TAT TTA CGT GCT GCT AAG GCA CTG CTG ACC TAT AGT GAG TCG TAT TAA TTT C, bottom strand sequence 5’ to 3’: GGT CAG CAG TGC CTT AGC AGC ACG TAA ATA TCC TAT AGT GAG TCG TAT TAA TTT C) and 44 bp (top strand sequence 5’ to 3’: GGT CTT AAC GCC AAT ATT TAC GTG CTG CTA AGG CAC TGC TGA CCT ATA GTG AGT CGT ATT AAT TTC bottom strand sequence 5’ to 3’: GGT CAG TGC CTT AGC AGC ACG TAA ATA TTG GCG TTA AGA CCT ATA GTG AGT CGT ATT AAT TTC) duplex RNAs were purchased from IDT containing a T7 promoter site on the 3’ end (5’ to 3’: TAT AGT GAG TCG TAT TAA TTT C). Also, DNA complementary to the T7 promoter site (5’ to 3’: GAA ATT AAT ACG ACT CAC TAT A) was purchased from IDT and annealed to the above DNAs to promote T7 transcription in a hemi-duplex method using the same conditions as pre-miR-16-1. The RNAs were purified similar to the pre-miR-16-1. The RNAs were quantified by UV-Vis absorption using $\varepsilon = 326,800 \text{ M}^{-1} \text{ cm}^{-1}$, $\varepsilon = 319,300 \text{ M}^{-1} \text{ cm}^{-1}$, $\varepsilon = 440,900 \text{ M}^{-1} \text{ cm}^{-1}$, and $\varepsilon = 422,900$
M$^{-1}$ cm$^{-1}$ at 260 nm for the 33 bp top, 33 bp bottom, 44 bp top, and 44 bp bottom strand, respectively.

The 12 bp (top strand sequence 5′ to 3′: GUC AGC AGU GCC, bottom strand is self complementary), 16 bp (top strand sequence 5′ to 3′: GUC AGC AGU GCC UUA G, bottom strand is self complementary) and 22 bp (top strand sequence 5′ to 3′: GUC AGC AGU GCC UUA GCA GCA C, bottom strand is self complementary) top and bottom strands as well as ds16-flanking top (sequence 5′ to 3′: CUC UUA UGA UAG CAA UGU CAG CAG UGC CUU AG), ds16-flanking bottom (sequence 5′ to 3′: CUA AGG CAC AGC UGA CCA CAA CCG ACA CUU CU), ds16-terta-stable (sequence 5′ to 3′: GUC AGC AGU GCC UUA GUU CGC UAA GGC ACU GCU GAC), ds16-tetra-U (sequence 5′ to 3′: GUC AGC AGU GCC UUA GUU UUC UAA GGC ACU GCU GAC), the ds16-hexa-U (sequence 5′ to 3′: GUC AGC AGU GCC UUA GUU UUU UCU AAG GCA CUG CUG AC) and ds16-octa-U (sequence 5′ to 3′: GUC AGC AGU GCC UUA GUU UUU UUU CUA AGG CAC UG C UGA C) RNAs were purchased from Dharmacon without any post synthesis purification or deprotection. The RNAs were deprotected according to the Dharmacon protocol, and the RNAs were spun down to dryness. The RNAs were resuspended in water to get a concentration of roughly 100 µM. The concentration was verified by UV-Vis using the following molar extinction coefficients at 260 nm for the 12 bp top, 12 bp bottom, 16 bp top, 16 bp bottom, 22 bp top, 22 bp bottom, ds16-flank top, ds16-flank bottom, ds16-tetra-stable, ds16-tetra-U, ds16-hexa-U and ds16-octa-U respectively: 113,900 M$^{-1}$ cm$^{-1}$, 110,300 M$^{-1}$ cm$^{-1}$, 156,900 M$^{-1}$ cm$^{-1}$, 151,300 M$^{-1}$ cm$^{-1}$, 211,500 M$^{-1}$ cm$^{-1}$, 206,300 M$^{-1}$ cm$^{-1}$, 318,000 M$^{-1}$ cm$^{-1}$, 304,700 M$^{-1}$ cm$^{-1}$, 344,200 M$^{-1}$ cm$^{-1}$, 351,400 M$^{-1}$ cm$^{-1}$, 370,800 M$^{-1}$ cm$^{-1}$, and 385,900 M$^{-1}$ cm$^{-1}$.
4.3.3 Electrophoretic Mobility Shift Assay:

The RNAs (top strand only for duplexes) were radiolabeled using $[^\gamma-^{32}\text{P}]\text{ATP}$, then the renatured RNAs were gel purified on an 8% native gel. For the duplex RNAs, the radiolabeled top-strand RNA was mixed with a 20 fold molar excess of cold bottom strand prior to purification. The binding reactions incubated at room temperature for 30 min. to ensure full equilibration in the presence of 50 mM Tris-HCl, pH 7.5, 200 mM sodium chloride, 5% glycerol, 100 μg/mL Bovine Serum Albumin, 1 mM dithiothreitol, and 0.1 mg/mL herring sperm DNA to prevent the complex from sticking in the wells. Subsequently, the binding reactions were run on a 0.25X TBE, 10% acrylamide gel at 12 V cm$^{-1}$ at 4 °C for 3 hrs, with each lane containing 20 μCi. Signal from free and bound RNA was quantified by imaging the gel on a Typhoon-9410 imager, and uncertainties estimated by averaging the intensities in each lane from duplicate gels. The resulting fraction bound curves were fit to a fully cooperative model, analogous to that used for Hill analysis. Data fitting was performed using the Levenberg-Marquardt model as implemented in Matlab (MathWorks).

4.3.4 NMR Methods:

Standard triple resonance NMR techniques$^{20, 21}$ were used to assign the backbone resonances of Dicer-dsRBD on a Bruker Avance III 500 MHz spectrometer (chemical shifts are reported in Appendix B Table B.1). The spin relaxation experiments were performed on a Bruker Avance III 500 MHz and 600 MHz spectrometers using standard $^{15}$N relaxation methods.$^{22, 23}$ All spectrometers were equipped with TCI cryoprobes for maximum sensitivity and the experiments were performed at 25 °C. Spectra were processed by NMRpipe and analyzed with SPARKY (SPARKY3.113; T.D.
Goddard and D. G. Kneller, University of California, San Francisco, CA). Data were analyzed in Matlab.

4.3.5 Model-free Analysis:

Lipari-Szabo Model-free fitting was performed using the program ModelFree 4.20,24 with diffusion tensor fitting performed using the quadric method.25, 26 The coordinates from the crystal structure of Dicer (2C4B) were used as structural references for diffusion tensor determination. $T_1$, $T_2$, NOE data were fit to a model including the axially symmetric global diffusion parameters with $S^2$ and $\tau_{int}$ (model 2), except Asp-1875 which required an $R_{ex}$ term (model 3).

4.3.6 Simulations:

MD trajectories were run in the AMBER 10.0 software package27 using the ff99SB28, 29 force field. Simulations were carried out in explicit solvent represented by the SPC water model30 under particle mesh Ewald periodic boundary conditions.31 Dicer-dsRBD MD simulations were run as previously reported16 using the crystal structure (2C4B, residues 1833-1900 of the mouse sequence corresponding with 1849-1916 of the human sequence). Nine chloride counterions were added to neutralize the net positive charge on the protein, then the resulting system was solvated such that no solute atom was within 10 Å of a box edge, this required 7031 water molecules. The starting configuration was energy minimized as previously reported.32 Following the initial equilibration period, 250 ns of dynamics were run in an isothermal – isobaric (NPT) simulation for the construct. Snapshots from each trajectory were stored to disk every 1.0 ps. The analysis of the trajectories was done in AMBER using the ptraj program.27 Molecular graphics images were created using the UCSF Chimera
Additional analysis and visualization was accomplished in Matlab. MD derived order parameters were obtained by using iRED analysis of MD trajectories averaged over 5ns windows, as previously reported.32,34

4.4 Results:

4.4.1 Binding of Dicer-dsRBD with Pre-miRNA and Perfect Duplex RNAs:

Electrophoretic mobility shift assays (EMSAs) were used to monitor the binding of Dicer-dsRBD (1850-1922) in isolation. The initial study was done with pre-miR-16-1, because it represents a dsRNA that Dicer would encounter in the cell and it correlates with previous work done with pri-miR-16-1 by our group and others.35-39 Dicer-dsRBD is able to bind pre-miR-16-1 with a $K_d = 2.15 \pm 0.01 \mu M$ (Figure 4.2 and Table 4.1) when fit to a Hill equation binding model, as used in other binding studies.35,40 Dicer-dsRBD binds tighter to pre-miR-16-1 than the first dsRBD of DGCR8 in isolation binds to pri-miR-16-1 ($K_d = 9.4 \pm 0.4 \mu M$); and Dicer-dsRBD also binds slightly tighter than DGCR8-Core, which contains two dsRBDs in tandem, does to pri-miR-16-1 ($K_d 3.7 \pm 0.1 \mu M$).35 Given the dual-role of Dicer in the cell, we also wanted to determine if Dicer-dsRBD is able to discriminate between pre-miRNA and pre-siRNA. To test the ability of Dicer-dsRBD to bind perfect Watson-Crick base paired dsRNA, which is a typical structure found in siRNA precursors, we designed a 44 bp RNA, based on pri-miR-16-1, where the top strand was from the native sequence of pri-miR-16-1, starting from the ss-ds junction and the bottom strand was the exact complement. Dicer-dsRBD has a similar binding affinity for this construct as for the pre-miRNA (Figure 4.2, Table 4.1). In order to check length dependence of dsRNA on binding, we designed three more perfect duplex RNAs based on pri-miR-16-1 of various lengths by starting with the ds44 and
**Figure 4.2:** EMSA of pre-miR-16-1 and ds44 binding by Dicer-dsRBD. EMSA of Dicer-dsRBD binding (A) pre-miR-16-1 with a $K_d = 2.15 \pm 0.01 \mu M$ and (B) ds44 with a $K_d = 2.39 \pm 0.02 \mu M$. The predicted secondary structure of the RNAs are shown above the representative gels, which were run with various Dicer-dsRBD concentration (0.25 – 50.12 µM) to 0.125 nM RNA. Fraction bound, from the EMSA data, versus Dicer-dsRBD concentration was fitted using a fully cooperative model (gray line).
consecutively removing one turn of A-form helix (11 bp) from the 3’ end with respect to the top strand. Additionally, we designed a perfect duplex 16mer in the same manner because it has been reported that dsRBDs span 16 bp of A-form helix dsRNA in complex.\textsuperscript{14} In the majority of cases, dsRBDs have shown no preference for RNA sequence\textsuperscript{14, 41} (exception being ADAR2\textsuperscript{42}), so the affinity for these constructs should be based primarily on length and not sequence differences. As the length of the dsRNA is decreased, the binding affinity decreases monotonically (Table 4.1). Note that no large decrease in binding affinity is observed as the RNA is changed from ds33 to ds22, which is significant because ds22 represents the approximate length of a Dicer cleavage product for both siRNA and miRNA precursors. Therefore, Dicer-dsRBD affinity alone is not sufficient to discriminate between reactant and product of the enzymatic reaction.

### 4.4.2 Effect of Hairpin Structure on Binding:

A major difference between pre-miRNA and pre-siRNA is that pre-miRNA has a hairpin structure that pre-siRNA lacks. As stated earlier, dsRBDs generally do not recognize RNA sequence, but are able to recognize structural features, specifically the structure of loops, as suggested by Rnt1p-dsRBD.\textsuperscript{43, 44} To test if Dicer-dsRBD is able to discriminate between pre-miRNA and pre-siRNA based on the terminal loop structure we designed four different loop constructs attached to ds16: a ds16-tetra-stable containing the thermostable UUCG tetraloop, ds16-tetra-U containing a poly-U tetraloop, ds16-hexa-U containing a poly-U hexaloop, and ds16-octa-U containing a poly-U octaloop. Poly-U loops were chosen because uracils do not stack upon each other, thus ensuring the formation of the desired loop sizes. Dicer-dsRBD has the same
Figure 4.3: EMSA of ds16-tetra-stable and ds16-octa-U binding by Dicer-dsRBD.

EMSA of Dicer-dsRBD binding (A) ds16-tetra-stable with a $K_d = 9.1 \pm 0.1 \mu M$ and (B) ds16-octa $K_d = 4.70 \pm 0.02 \mu M$. The predicted secondary structure of the RNAs are shown above the representative gels, which were run with various Dicer-dsRBD concentration (0.25 – 50.12 µM) to 0.125 nM RNA. Fraction bound, from the EMSA data, versus Dicer-dsRBD concentration was fitted using a fully cooperative model (gray line).
affinity for the ds16 RNA with and without the thermostable UUCG tetraloop (Figure 4.3A and Table 4.1). It is only upon addition of the poly-U octalooop that Dicer-dsRBD binding affinity increases for the RNA, albeit it is a modest two-fold increase (Figure 4.3B and Table 4.1). Although a factor of two change in affinity is modest, it is reproducible; and the addition of a poly-U octalooop has roughly the same effect on $K_d$ as doubling the length of the RNA, as exemplified by it having a similar binding affinity to the ds33 construct (Table 4.1). Many miRNAs feature loops comparable in size to the highly flexible poly-U loop used here and have been confirmed to adopt highly disordered conformations through SHAPE reactivity (Quarles and Showalter, unpublished results).

Besides recognizing the structure of the loop, it is possible that Dicer-dsRBD is recognizing the ss-ds junction caused by the large poly-U octalooop and not the structure of the loop itself. Therefore, ds16-flank was designed, where 16 nucleotides were attached to the top and the bottom strand of the ds16 construct creating a ssRNA tail, which showed a similar binding affinity as ds16-octa-U by Dicer-dsRBD. Together, this suggests that Dicer-dsRBD binding is influenced by the presence of a ss-ds junction created by either a large terminal loop or a ssRNA tail.

**Table 4.1:** Binding Affinity of Dicer-dsRBD for Various RNA Constructs by EMSA

<table>
<thead>
<tr>
<th>RNA Construct</th>
<th>Dissociation Constant ($K_d$, µM)</th>
<th>Hill Coefficient (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-miR-16-1</td>
<td>$2.15 \pm 0.01$</td>
<td>$2.23 \pm 0.01$</td>
</tr>
<tr>
<td>ds44</td>
<td>$2.39 \pm 0.02$</td>
<td>$3.18 \pm 0.01$</td>
</tr>
<tr>
<td>ds33</td>
<td>$4.92 \pm 0.03$</td>
<td>$2.81 \pm 0.03$</td>
</tr>
<tr>
<td>ds22</td>
<td>$6.54 \pm 0.02$</td>
<td>$3.4 \pm 0.2$</td>
</tr>
<tr>
<td>ds16</td>
<td>$8.94 \pm 0.09$</td>
<td>$3.4 \pm 0.2$</td>
</tr>
<tr>
<td>ds12</td>
<td>$15.9 \pm 0.1$</td>
<td>$3.8 \pm 0.1$</td>
</tr>
<tr>
<td>ds16-tetra-stable</td>
<td>$9.1 \pm 0.1$</td>
<td>$2.92 \pm 0.03$</td>
</tr>
<tr>
<td>ds16-tetra-U</td>
<td>$10.67 \pm 0.03$</td>
<td>$4.6 \pm 0.3$</td>
</tr>
<tr>
<td>ds16-hexa-U</td>
<td>$8.94 \pm 0.08$</td>
<td>$3.77 \pm 0.05$</td>
</tr>
<tr>
<td>ds16-octa-U</td>
<td>$4.70 \pm 0.02$</td>
<td>$2.84 \pm 0.02$</td>
</tr>
<tr>
<td>ds16-flank</td>
<td>$4.62 \pm 0.02$</td>
<td>$2.52 \pm 0.02$</td>
</tr>
</tbody>
</table>
4.4.3 NMR Spin Relaxation:

Backbone resonances of Dicer-dsRBD in the apo-state were assigned using standard triple resonance NMR techniques on a Bruker Avance III 500 MHz spectrometer (chemical shifts are reported in Appendix B Table B.1). Additionally, we tried to obtain backbone assignments for the holo-state of Dicer-dsRBD with ds16, but peaks were severely line broadened, indicating intermediate exchange on the NMR timescale (data not shown). Apo-Dicer-dsRBD spin relaxation (\(^{15}\text{N} T_1, T_2,\) and \(^{1}\text{H}\)\(^{15}\text{N}\) NOE NMR) was measured at 500 MHz and 600 MHz field strength (Figure 4.4). The quadric method used to analyze the spin relaxation data reveals that apo-Dicer-dsRBD tumbles anisotropically in solution with a \(D_{||}/D_{\perp} = 0.51\) and \(\tau_{\text{iso}} = 6.35\) ns, which agrees well with previous dsRBDS from DGCR8 and Drosha (\(\tau_{\text{iso}} = 7.20\) ns and 6.29, respectively). These rotational tumbling times are representative of a monomeric assembly state for a globular protein domain of this size (\(~8\) kDa); this was further supported through dynamic light scattering (DLS) and NMR diffusion measurements for Dicer-dsRBD (Appendix B Figure B.3 and Figure 6.4).

Our previous data suggest a correlation between backbone flexibility, as modeled by the generalized order parameter \((S^2)\), and dsRBD binding competence. Picosecond to nanosecond timescale backbone conformational dynamics analyzed through the generalized order parameters \((S^2)\) have been obtained by complete model-free analysis of the apo-Dicer-dsRBD spin relaxation data (Figure 4.5A). Dicer-dsRBD shows lower order parameters in loop 3 and loop 4, indicating higher flexibility, with a minimal decrease in the order parameters for loop 2. Among dsRBDS previously
Figure 4.4: $^{15}$N spin relaxation of Dicer-dsRBD. $^{15}$N spin relaxation data for Dicer-dsRBD collected at 500 MHz (purple) and 600 MHz (gray). The data shows that the most dynamic regions of Dicer-dsRBD are the loops, most notably loop 4, on the picosecond to nanosecond time scale. The secondary elements are represented as purple bars above the plot.
studied, only Staufen-dsRBD3 has shown increased dynamics in loop 4 (by \( ^1\text{H} \cdot ^{15}\text{N} \) NOE NMR spin relaxation, not \( S^2 \)). The other region of increased flexibility in Staufen-dsRBD3 is loop 2, which shows increased dynamics in all of the other dsRBDs studied to date. Dicer-dsRBD displayed elevated dynamics in this loop too, albeit only slightly. An alignment of Dicer-dsRBD sequence with other dsRBDs indicates that the decrease in the dynamic characteristics of loop 2 could be due to Dicer-dsRBD loop 2 being three residues shorter than the canonical length (Figure 1.2). The decreased dynamics of loop 2 is likely a facilitating component in the high binding affinity of Dicer-dsRBD.

4.4.4 Molecular Dynamics Simulations of Dicer-dsRBD:

In connection with experimental data, MD simulations can provide useful dynamic information on the mechanism of proteins. Residues 1833 to 1900 from the crystal structure of the C-terminal region of mouse Dicer (2C4B) were used as the starting point for molecular dynamics (MD) simulations. The sequence used for the MD simulations is one hundred percent identical to residues 1849 to 1916 in human Dicer. Root-mean-square deviation (RMSD) from the starting structure over the course of the trajectories verified that the Dicer-dsRBD was stable over the 250 ns simulation (Figure 4.5B). The Dicer-dsRBD construct is highly stable; its RMSD is very low (1.0 Å) for a large majority of the simulation (figure 4.5B). This is the lowest RMSD we have reported for any dsRBD in isolation, highlighting the high stability of the backbone of Dicer-dsRBD as compared with other dsRBDs. Further evidence for the stability of the Dicer-dsRBD comes from ribbon bundles (Figure 4.5C) of the simulations overlapping well and showing no loss in secondary structure elements.
Figure 4.5: Order parameters and RMSD of Dicer-dsRBD. (A) Order parameter ($S^2$) plot for Dicer-dsRBD show that the most flexible regions in the protein are loops 3 and 4. Experimental data (purple) is plotted against MD predicted (gray) order parameters. The secondary elements are represented as purple bars above the plot. (B) The overall stability of Dicer-dsRBD during the 250 ns MD simulation is demonstrated by the low average RMSD (< 1.0 Å). (C) MD-derived ribbon bundle for Dicer-dsRBD also shows the overall stability of the construct. Increased flexibility, derived from the experimental order parameters, is depicted colorimetrically on the ribbon bundle as passage from purple (high order parameter) to yellow (low order parameter).
iRED analysis of the MD trajectories averaged over 5 ns windows gave computational Model-free order parameters, S^2, (Figure 4.5A, gray line). Qualitative comparison of the computationally with the experimentally derived order parameters show the same global trends. Note, there exists an offset between the experimental and computational order parameter profiles; a similar observation was seen with Drosha-dsRBD where the experimental order parameters have a lower average than those derived from MD.

In our previous studies of isolated dsRBDs, we utilized principal component analysis (PCA) to investigate correlated dynamics within the dsRBD of DGCR8 and Drosha. The same analysis was done with Dicer-dsRBD (Figure 4.6) demonstrating that Dicer-dsRBD has similar correlated motions as DGCR8-dsRBD1 (see figure 2.4A and figure 3.7B). The major positive correlation in Dicer-dsRBD is between loop 1 and loop 2, which is also observed in DGCR8-dsRBD1. While, the major negative correlations in Dicer-dsRBD is between loop 2 and strand 3 as well as with helix 2. These anti-correlated motions are not as strong in DGCR8-dsRBD1, which is a potential contributing factor to Dicer-dsRBD binding slightly tighter.

4.5 Discussion:

In this chapter, we have focused on the binding affinities between Dicer-dsRBD, in isolation from the rest of the protein, with various RNAs. Our results show that Dicer-dsRBD is able to discriminate pre-miRNA and pre-siRNA based on the length and presence of a terminal loop and thus it may be involved in the mechanism of discrimination between these two distinct biological targets. Previously we have studied
Figure 4.6: Cα correlations of Dicer-dsRBD. Cα correlation matrix reveals the collective backbone motions of Dicer-dsRBD. The color bar on the right shows the scale indicating strong positive correlation (red), strong negative correlation (blue), and noncorrelated motion (green). Labels above the panel indicate the location of secondary structural elements within the sequence.
pri-miRNA binding and the protein backbone dynamics of the dsRBDs involved in miRNA maturation by the Drosha-DGCR8 complex.\textsuperscript{35} The present Dicer studies, our previously reported data, and information from additional studies in the literature, yield broadly consistent trends correlating specific patterns of backbone dynamics and RNA binding activity in dsRBDs.

4.5.1 Dicer-dsRBD Role in Discriminating Between siRNA and miRNA precursors:

Dicer is involved in cleavage of siRNA precursors and miRNA precursors. Both of these types of small regulatory RNAs have different structural features that could be exploited to promote matching with appropriate pathway-specific Ago proteins by Dicer. Notably, canonical siRNA precursors are long dsRNAs composed almost exclusively of Watson-Crick base pairs, while pre-miRNA are roughly 26 bp of dsRNA with a terminal loop and internal imperfections (loops and bulges). From our length study, Dicer-dsRBD shows a modest monotonic decrease in binding affinity as the length of dsRNA ligand decreases from 44 to 12 bp. Thus, Dicer-dsRBD is able to discriminate dsRNAs based on length. Even with only 12 bp of RNA, Dicer-dsRBD is able to bind, albeit with a minor drop-off in binding affinity going from ds16 to ds12. This result is in agreement with literature that states a single dsRBD can bind a minimum of 11 bp.\textsuperscript{14, 49} As noted earlier, no large drop-off in binding affinity is observed going from ds33 to ds22, suggesting Dicer-dsRBD is unable to distinguish between the reactant and the product of Dicer cleavage and is therefore unlikely to serve a critical role in product release.

Other than duplex length, the most striking factor for dsRNA discrimination is differences in secondary structure. Both pre-siRNA and pre-miRNA have 3’ two
nucleotide overhangs, which the PAZ domain recognizes.\textsuperscript{2, 11} The major structural difference between pre-siRNA and pre-miRNA is the presence of the terminal loop on pre-miRNAs. Dicer-dsRBD binding is influenced by the ss-ds junction caused by large terminal loops – as indicated by the higher binding affinity of the ds16-octa-U (K\textsubscript{d} = 4.70 ± 0.02 µM) versus the binding affinity for ds16 without a loop (K\textsubscript{d} = 8.94 ± 0.09 µM). Although this effect is small in isolation, cooperative contributions in the context of full Dicer involving the dsRBD could yield substantial discriminatory value and perform a major role in the mechanism.

As the PAZ domain binds on the 3′ end and the dsRBD is able to discriminate RNA based on the presence of ss-ds junction caused by large terminal loops on pre-miRNAs, we propose a model where these two domains in Dicer bind opposite ends of the pre-miRNA to correctly place the active sites of the RNase III domains. Further studies need to be done with full length Dicer and a Dicer mutant devoid of the dsRBD to support this novel role for the Dicer-dsRBD in discriminating between Dicer substrates. That said, our model agrees with the recent cryo-EM constructed low resolution structure, where the PAZ domain and the dsRBD are separated by a ruler domain and cleavage is between these two domains;\textsuperscript{12} and provides a plausible model for previously unaddressed features of pre-miRNA recognition.

\textbf{4.5.2 Dicer-dsRBD Binding Compared to Other dsRBDs:}

Among the isolated or tandem dsRBDs from the miRNA processing pathways that we have studied to date, Dicer-dsRBD has the highest binding affinity.\textsuperscript{35} Based on length alone, our length dependence studies suggest DGCR8-dsRBD1 should bind pri-miRNA tighter than Dicer-dsRBD binds pre-miRNA, because pri-miRNA is one turn of
A-form RNA longer than pre-miRNA. This is not the case; our data illustrate that Dicer-dsRBD is better at binding dsRNA than DGCR8-dsRBD1 and in fact binds more tightly than the core construct of DGCR8, which contains two dsRBDs.

Protein backbone dynamics have been shown to play a vital role in binding,\textsuperscript{22, 50, 51} therefore we performed NMR spin relaxation on Dicer-dsRBD to establish a dynamic profile for this domain in the unbound state. Dicer is unusual in that loop 2 shows only a minimal increase in flexibility when compared to loop 3 in both the experimental and computational data (Figure 4.5A). In our previous work, loop 2 of DGCR8-dsRBD1 has dramatically lower order parameters when compared with the rest of the domain, but Dicer-dsRBD loop 2 is only slightly more flexible. Loop 2 in Dicer-dsRBD is three residues shorter, thus explaining the reduced dynamics of this region, which appear not to be detrimental to binding. The decreased dynamics in loop 2 of Dicer-dsRBD correlated with it binding dsRNA tighter than DGCR8-dsRBD1 merits future studies, so at this time we do not imply a causal relationship between the two.

Dicer-dsRBD shows additional dynamics in loop 3 and loop 4, which were not observed in DGCR8-dsRBD1. Dicer-dsRBD is the first reported case of a dsRBD having flexibility in loop 3 of the domain. Loop 3 is on the opposite side of the canonical RNA binding interface of loop 2 and loop 4, so its dynamics appear unrelated to dsRNA binding unless they are correlated with the dynamics of another region of the domain that is involved in binding. Cα correlations plots from the MD simulations of Dicer-dsRBD show similar correlations as DGCR8-dsRBD1 (Figure 4.6).\textsuperscript{16, 35} Thus, the Cα correlations of loop 3 are not unique for Dicer-dsRBD and would not explain dynamic contribution loop 3 has for binding. Without further studies, we are not sure the role
dynamics of loop 3 has on binding dsRNA, if it even does. The dynamics of loop 3 could be an artifact of having the Dicer-dsRBD in isolation, where in the context of the whole protein it could be less dynamic due to packing interactions with another domain. Dicer-dsRBD is not the first case of increased dynamics in loop 4; based on \( \{^{1}H\}-^{15}N \) NOE NMR spin relaxation, Staufen-dsRBD3 has increased dynamics in loop 4 in both the apo and holo state.\(^46\) Both Dicer-dsRBD and Staufen-dsRBD3 demonstrate that the flexibility observed in this loop is not detrimental to binding of dsRNA by a dsRBD.

In closing, this work along with previous studies starts to provide a dynamic profile for the binding mechanism of dsRBDs. While the dynamics of the loops vary from one dsRBD in the apo-state to the next, the overall C\(\alpha\) correlations remain relatively consistent. Taken together, this clearly demonstrates that dynamics are a key attribute in determining the overall mechanism of dsRBD binding.

4.6 Acknowledgment:

I would like to thank Kaycee Quarles and Rico Acevedo for their help in preparing some of the RNA constructs and helpful discussion. Also, I thank Prof. Phil Bevilacqua and members of the Bevilacqua laboratory for assistance with RNA design and helpful discussion. This work was supported by a grant from the NIH to S.A.S (RO1GMO98451).
4.7 References:


Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular

28. Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C., HIV-1 protease flaps

29. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.,
Comparison of multiple amber force fields and development of improved protein

30. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.,


32. Showalter, S. A.; Bruschweiler, R., Validation of molecular dynamics simulations
of biomolecules using NMR spin relaxation as benchmarks: Application to the

33. Pettersen, E. F. G., T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng,
E. C.; Ferrin, T. E., UCSF Chimera a visualization system for exploratory research and

34. Prompers, J. J.; Bruschweiler, R., General framework for studying the dynamics
of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J.


43. Wu, H.; Henras, A.; Chanfreau, G.; Feigon, J., Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Apr 26 2004, 101* (22), 8307-12.


Chapter 5

Future Directions for Studying dsRBDs
in the miRNA Maturation Pathway

[The following chapter is designed as proposed research stemming from the results presented in chapter 2, chapter 3 and chapter 4]

5.1 Cooperativity Between the Two dsRBDs in DGCR8-Core:

The RNase III enzyme Drosha is known to cleave dsRNA non-specifically,\(^1\) illustrating the importance of the cofactor DGCR8 in the maturation of miRNAs. In addition, it has been shown that both dsRBDs of DGCR8 are necessary for high affinity binding of dsRNA,\(^2\) thus leading to the hypothesis that the two dsRBDs bind dsRNA with positive cooperativity. In this case, positive cooperativity means that the binding of dsRNA to one of the dsRBD increases the affinity of binding for the other dsRBD, thus the cooperative free energy (\(\Delta G_c\)) should be negative (favorable free energy). This type of cooperativity is observed for PKR, an antiviral response protein containing two dsRBDs in tandem, where PKR-dsRBD1 first anchors dsRNA inducing PKR-dsRBD2 to then bind the RNA.\(^3\) A major difference between DGCR8 and PKR, is that DGCR8 has a well-ordered linker between the two dsRBDs,\(^2\) whereas PKR has a flexible linker.\(^3\)\(^,\)\(^4\)
Figure 5.1: Thermodynamic cycles. Examples of thermodynamic cycles used for sections 5.1 and 5.2 are shown, where $\Delta G_C$ is the free energy associated with cooperativity. (A) Scheme for determining cooperativity of binding between the two dsRBDs of DGCR8-Core (section 5.1). (B) Scheme for determining the effect intermolecular interactions have on binding.
Figure 5.1A shows the thermodynamic cycle that can be utilized to understand the cooperativity between the two dsRBDs in DGCR8-Core in the proposed study. Three binding dissociation constants (K_D) can be measured and the corresponding free energy (ΔG) can be calculated for the three constructs, dsRBD1 (ΔG_1), dsRBD2 (ΔG_2), and Core (ΔG_3), using the following relationship:

\[ ΔG = +RT \ln(K_D) \] (5.1)

where R is the gas constant (8.3144 J K⁻¹ mol⁻¹) and T is the absolute temperature. Due to free energy being a state function, it is path independent; therefore the free energy of Core binding pri-miRNA should be equal to the free energy of dsRBD1 binding pri-miRNA plus the free energy of dsRBD2 binding pri-miRNA plus the free energy of cooperativity. If the measured Core binding free energy is non-additive with respect to the free energy of the individual domains (and the deviation from the additive binding is statistically significant), the excess or deficit free energy is attributed to cooperativity (ΔG_c in Figure 5.1A). Therefore the free energy of cooperativity will be experimentally determined as the difference between the free energy of the Core binding pri-miRNA with the summation of the free energy of the individual dsRBDs binding pri-miRNA. An alternative to using the three constructs described above is to create two mutants of the Core, each of which eliminate the thermodynamic significance of protein-RNA contracts mediated by, for example, charged arginine and lysine residues in the desired dsRBD. This can be readily accomplished, because Sohn et al. have already reported mutant constructs of the DGCR8-Core that eliminate the binding ability of each of the individual dsRBDs,² although they opted to not carry out their analysis through to evaluating cooperativity in the system.
In chapter 2, dissociation constants for DGCR8-dsRBD1 and DGCR8-Core were measured using EMSAs. Initial EMSAs of DGCR8-dsRBD2 gave a $K_D = 4.3 \pm 0.1 \mu M$, but this dissociation constant was not able to be reproduced. The major hindrance is the stability of DGCR8-dsRBD2, which tends to precipitate out of solution within hours, making it difficult to reproduce EMSAs. Additionally, the current DGCR8-dsRBD2 construct contains the C-terminal helix and most of the linker, which affects the dissociation constant of the construct by providing a surface for dimerization. A more minimal DGCR8-dsRBD2 construct that is void of both the C-terminal helix and the linker is even more unstable. The alternative method of mutating out the binding surface in the individual dsRBDs from the Core might need to be utilized if the stability of DGCR8-dsRBD2 cannot be improved to allow for reproducible EMSAs.

5.2 Interfacial Interactions of DGCR8-Core:

The crystal structure of DGCR8-Core revealed that the two dsRBDs are arranged in a pseudo two-fold symmetry and packed against a well defined secondary structure formed from the linker and the $\alpha$-helix formed at the C-terminal tail of the construct.\(^2\) The $\alpha$-helix forms hydrophobic interactions with $\alpha$-helix H2 and $\beta$-strand S5 of both dsRBDs, with additional interactions from helix H1’ of dsRBD2; also it forms several hydrogen bonds with the dsRBDs.\(^2\) MD simulations confirm that the most likely hydrogen bonds between the dsRBDs and the $\alpha$-helix are the hydrogen bond between Arg-630 and Met-697; and the hydrogen bond between Ile-575 and Ser-693 (Chapter 2).\(^5\) The stability of further interfacial interactions was confirmed in the MD simulations, notably a hydrogen bond formed by the amide nitrogen of Asn-631 and the carbonyl in the backbone Gly-550 and a salt bridge formed by amine nitrogen of Lys-659 and the
carbonyl of Asp-549 (Chapter 2). These interfacial interactions in the absence of RNA are in contrast with the only other protein of known structure with tandem dsRBDs, PKR. Therefore, I hypothesize that the interfacial interactions of DGCR8 are critical for its mechanism of binding a diverse family of pri-miRNA; with sufficient specificity to exclude binding of other pools of dsRNA.

A representative thermodynamic cycle is shown in Figure 5.1B to aid in the explanation of the mutant studies. As a specific example, the salt bridge between Asp-549 and Lys-659 will be discussed, but the same logic applies to the other interfacial interactions. In this case, the hypothesis is that the salt bridge aids in pri-miRNA recognition by DGCR8, thus with it eliminated the binding free energy should be higher (less favorable) and the change in free energy of the double mutant is equal to sum of the change of free energy of single point mutants (i.e., once one charge is removed, there is no further penalty for removing the other). The absolute change caused by each of the isolated point mutants need not be precisely the same because of additional factors (such as changes to van der Waals contacts) will also have an impact on the measured free energy. First, three mutants have been purchased; an Asp-549 to alanine (single point mutation), a Lys-659 to alanine (single point mutation) and a double mutant where both residues are changed to an alanine. The residues were mutated to alanines, since alanine is small, which prevents introducing steric strain, and it is a non-charged, non-polar residue, which eliminates the salt bridge and prevents further interactions from occurring. Four binding constants can be measured and from that the corresponding free energy can be calculated; wild type (\( \Delta G_{WT} \)), the two single mutants (\( \Delta G_{M1} \) and \( \Delta G_{M2} \)) and the double mutant (\( \Delta G_D \)). By comparing \( \Delta G_{WT} \) to \( \Delta G_{M1} \)
and \( \Delta G_{M2} \), I would expect \( \Delta G_{WT} \) to be more negative than \( \Delta G_{M1} \) and \( \Delta G_{M2} \), this alone does not prove that the salt bridge is the cause for the higher free energy. Instead of analyzing the free energies directly, the change in free energy (\( \Delta \Delta G \)) will be analyzed to determine the extent of cooperativity of the two mutants. If the salt bridge is the cause of the less favorable free energy of the single point mutant, then the change in free energy of the double mutant (\( \Delta \Delta G_3 \)) should equal the sum of the change of free energy of single point mutants (\( \Delta \Delta G_1 \) and \( \Delta \Delta G_2 \)), therefore \( \Delta \Delta G_c \) should be zero.

In order to eliminate the interfacial interactions, single alanine mutants of Asn-631, Lys-659 and Asp-549 were purchased along with the double alanine mutant, Asp-549/Lys-659. Remember that the hydrogen bond is between Asn-631 and the backbone carbonyl of Gly-550, which cannot be mutated to an alanine to break the hydrogen bond. Additional double mutants that eliminate both the hydrogen bond and the salt bridge were purchased as well; Asn-631/Lys-659 and Asp-549/Asn-631. Initial EMSAs gave similar dissociation constants for the single and double mutants as wild-type DGCR8-Core with pri-miR-16-1 (data not shown). The reason for the lack of differences in binding affinities could be due to DGCR8-Core being able to bind to multiple regions of dsRNA, as indicated by the Hill coefficient greater than one. Therefore, a one-to-one binding interaction between DGCR8-Core and dsRNA needs to be established, and a good starting point is the model dsRNAs that were used in the Dicer-dsRBD study (chapter 4). AUC (in collaboration with Jim Cole’s group at UConn) can be utilized to verify that wild-type DGCR8-Core binds the model dsRNA in a one-to-one fashion prior to continuing EMSAs with the mutants. Another method for determining the stoichiometry is monitoring the CD change at 260 nm. In the case of
PKR, the dsRBD relaxes the winding of the RNA as it is titrated in causing an increase in the CD intensity until the stoichiometric ratio of RNA to protein is reached, then the CD intensity is unchanged with increasing protein.\textsuperscript{5}

After positive results from these mutants are collected, then a His-660 alanine mutant can be made to further explore the role this amino acid residue plays in stabilizing the salt bridge between Asp-549 and Lys-659. Its close proximity to the salt bridge throughout the MD simulations (chapter 2) suggests that the imidazole moiety is important in maintaining the salt bridge with fluctuations in pH. Again, EMSAs of the His-660 mutant with a model dsRNA can be done, with the expectation that it should show similar binding affinity as wild-type at cellular pH. Next, the pH of the binding reaction can be varied for both the wild-type and His-660 mutant to elucidate the effect the His-660 has maintaining the salt bridge. Additionally, it might be a good idea to perform the same pH study with the salt bridge eliminated, i.e., using the Asp-549/Lys-659 double mutant, to verify similar results as the His-660 mutant. Further, NMR pKa studies can be utilized to determine the pKa of the His-660 and the effect pH has on the local environment around the salt bridge.

Finally, the conclusion from the MD simulations was that the domain-linker interface formed by the C-terminal helix is stabilized by only two out the four crystallographically suggested hydrogen bonds (chapter 2).\textsuperscript{5} Therefore to test this single and double alanine mutants can be made and studied as previously described in this section. This study should include all four hydrogen bonds suggested by the crystal structure to validate the results from the simulations. Like the Asn-631 to Gly-550 hydrogen bond between the two dsRBDs, the hydrogen bond between Arg-630 and
Met-697 involves the backbone carbonyl, so only a single mutant can be studied. All the rest of the hydrogen bonds are between side chains, therefore both single and double mutants can be made. Also, similar to the interfacial interactions, mutants can be made to eliminate multiple hydrogen bonds in this region of the protein.

5.3 Effect of Dynamics in Loop 1 of dsRBD on Binding of dsRNA:

From Chapter 3, it was demonstrated that Drosh-dsRBD in isolation was incapable of binding dsRNA, which was hypothesized to be due to the non-canonical loop 1. Generally, dsRBDs have a small loop 1 that is not very dynamic as illustrated in DGCR8-dsRBD1 (chapter 3) and Dicer-dsRBD (Chapter 4). In the case of the Drosha-dsRBD, loop 1 is five amino acids longer and is negatively charged (Chapter 3). The extension of loop 1 in Drosha causes increased flexibility in this region, which is hypothesized to be detrimental to dsRNA binding due to charge-charge repulsion between the negatively charged loop and the negatively charged phosphate backbone of the dsRNA. Alternatively, the apo-state dynamic nature of loop 1 in Drosha-dsRBD could be a hindrance to dsRNA binding due to a loss in entropy in the bound state causing a larger dissociation constant value.

In order to test the hypothesis that increased dynamics in loop 1 leads to a hindrance in dsRNA binding by a dsRBD, two chimera constructs were purchased: Drosha-dsRBD with the smaller loop 1 from Dicer-dsRBD and Dicer-dsRBD with the larger loop 1 from Drosha-dsRBD. I would expect the Drosha chimera to be able to bind dsRNA albeit weaker than other dsRBDs studied, while the Dicer chimera should have weaker binding affinity than wild-type Dicer-dsRBD. In contrast, preliminary
EMSAs demonstrated that the smaller loop 1 from Dicer-dsRBD was not enough to recover dsRNA binding ability of Drosha-dsRBD. Therefore, other factors like the lack of conserved lysines and arginines in the dsRNA interface of helix 2 of Drosha-dsRBD must contribute to the incapability of Drosha-dsRBD to bind dsRNA in isolation. This could be studied through mutagenesis of Drosha-dsRBD to incorporate lysines and/or arginines along the dsRNA binding interface of helix 2. Likewise, the addition of the larger loop 1 to Dicer-dsRBD had no effect on its binding affinity. It is possible that in the context of Dicer-dsRBD, the loop 1 from Drosha-dsRBD is not as dynamic as it was observed to be in Drosha-dsRBD. Therefore, it is critical to do NMR spin relaxation experiments on both the chimeras to verify that swapping of the loops caused the desired dynamic effects. Additionally, it would be useful to purchase DGCR8-dsRBD1 with loop 1 from Drosha-dsRBD as a further verification that the large loop 1 from Drosha-dsRBD does or does not have an effect on dsRNA binding. It has been demonstrated that Dicer-dsRBD binds dsRNA with a higher binding affinity (Chapter 4) than DGCR8-dsRBD1, so it is conceivable that Dicer-dsRBD binding would be unaffected by the addition of loop 1, but that DGCR8-dsRBD1 would be. Taken together, the preliminary EMSA data suggests that the initial hypothesis stating that loop 1 being dynamic hinders dsRNA from binding along the α-helical binding interface is incorrect.

5.4 Importance of Spacing Between Loop 2 and Loop 4:

Alignment of 15 dsRBDs (Figure 1.2) revealed that a conserved aromatic residue (a phenylalanine or tyrosine) is positioned at the beginning of β-strand 2. Chang and Ramos hypothesized that the conserved aromatic residue is vital for maintaining the
distance between loop 2 and loop 4 to match the spacing between the major and minor groove of A-form dsRNA. In Chapter 3, it was shown that the aromatic residue might act as a pivot for anticorrelated motions between loop 2 and helix 2, regions involved binding dsRNA. Dicer-dsRBD is abnormal in the fact that it has a valine instead of the conserved aromatic (Figure 1.2), but Tyr-1897 in helix 2 of Dicer-dsRBD instead might be capable of maintaining the spacing between loop 2 and loop 4 to match the spacing between the major and minor groove of the dsRNA. Even with the lack of the conserved aromatic in the exact location suggested by alignment, Dicer-dsRBD is able to bind dsRNA quite well and shows similar apo-state dynamics as DGCR8-dsRBD1 (chapter 3 and chapter 4).

To modify the spacing between loop 2 and loop 4, mutagenesis of the Phe-542 from DGCR8-dsRBD1 can be performed. First, an alanine mutant can be made to decrease the size of the amino acid residue, thus potentially decreasing the spacing if no other amino acid residues are involved in the packing. Additionally, this mutant might have increased dynamics affecting the spacing between the two loops as well as the binding ability. In order to cause the spacing to increase between the two loops, a tryptophan mutant can be made. The last mutant to make is a lysine mutant because it can mimic the packing ability of an aromatic residue but is more dynamic. Similar mutants of Dicer-dsRBD can be made as well, mutating both Val-1878 and Tyr-1897, to verify the affects in another dsRBD. Like the chimeras from the previous section, both NMR spin relaxation and EMSAs will be performed on the mutants to determine the dynamics and binding affinity for each mutant. Based on preliminary experimental results, MD simulations of the mutants can be done to access the ability of the mutants
to maintain the anticorrelated motions between loop 2 and helix 2 observed in the wild-type DGCR8-dsRBD1.

5.5 Effect of Internal Loops and Bulges in dsRNA has on Dicer-dsRBD Binding:

Dicer is responsible for the maturation of two different classes of RNAs; siRNA and miRNA. siRNAs are derived either endogenously from repetitive sequences or exogenously from viral RNAs, in both cases the siRNA precursor is long and perfectly duplexed RNA. On the other hand miRNAs are endogenously transcribed as a hairpin structure with imperfections (internal loops and bulges) in the A-form helix. In chapter 4, it was demonstrated that Dicer-dsRBD is able to modestly discriminate between pre-miRNA and pre-siRNA based on the ss-ds junction caused by large terminal loops, but the study failed to extensively look at the role imperfections in the A-form helix, i.e., internal loops and bulges have the binding affinity.

All the experiments discussed below can use the ds16 as the basis RNA because Dicer-dsRBD binds it in a one-on-one stoichiometry based on AUC (Figure B.2). Also, the secondary structure of all constructs proposed can be computationally predicted using mFold to verify no alternative folds are probable. Likewise, the secondary structures can be experimentally verified by using structure mapping techniques, most notably SHAPE and endonucleases. The experimental data can be used as further constraints in mFold and other online servers to predict the structure of RNA constructs. As in previous sections, binding affinity for Dicer-dsRBD with the various model RNAs will be determined using EMSAs.
As mentioned earlier, a major feature of all pre-miRNAs is the existence of a terminal loop, but pre-miRNAs are also predicted to have internal loops and bulges that deform the A-form helix. First, variable sized bulges can be introduced in the middle of ds16 ranging from one to five nucleotides. Addition of a bulge in the middle of ds16 has the possibility of disrupting Dicer-dsRBD binding altogether because it would deform the A-form helix structure. If the structure of the A-form helix is vital for Dicer-dsRBD binding, then even the addition of a one nucleotide bulge will halt binding. Due to the imperfections in pre-miRNAs, Dicer-dsRBD might be able to tolerate slight deformations in the A-form helix, but not larger deformations. Therefore, RNAs with small bulges (one to two nucleotides) might be able to be bound by Dicer-dsRBD, but not RNAs with larger bulges. This basic setup will also be used to study the effect of symmetric and asymmetric internal loops on binding.

Finally, the role of the imperfections in the miRNAs precursors could be to correctly position and prevent sliding of the dsRBD on the RNA, which can only be tested with longer duplex constructs that allow for more than one dsRBD to be bound initially. To study this effect of imperfections (bulges and internal loops) on binding, imperfections will be introduced into ds22 and ds33 where the imperfection disrupts the A-form helix, but enough space exists for Dicer-dsRBD to bind to either side of the bulge.
5.6 References:


Chapter 6

Atomistic Simulations Reveal Structural Disorder in the RAP74-FCP1 Complex


Figure 6.5 and 6.7 were made by Sushant Kumar and W. G. Noid. All authors participated in the writing of the manuscript this chapter is derived from.

6.1 Abstract:

We report atomically detailed molecular dynamics simulations characterizing the interaction of the RAP74 winged-helix domain with the intrinsically disordered C-terminal of FCP1. The RAP74-FCP1 complex promotes the essential dephosphorylation of RNA polymerase II prior to initiation of transcription. Although disordered in solution, the C-terminal of FCP1 forms an amphipathic helix when bound to RAP74. Our simulations demonstrate that this interaction also reorganizes and stabilizes RAP74. These simulations illuminate the significance of hydrophobic contacts for stabilizing disordered protein complexes, provide new insight into the mechanism of protein binding by winged-helix domains, and also reveal “dynamic fuzziness” in the complex as FCP1 retains significant flexibility after binding. In conjunction with our recent NMR experiments identifying residual structure in unbound FCP1, these
simulations suggest that FCP1 loses relatively little conformational entropy upon binding and that the associated coupled folding-binding transition may be less sharp than expected.

6.2 Introduction:

Intrinsically disordered proteins (IDPs), or intrinsically disordered regions (IDRs) within longer polypeptide chains, are an intriguing class of proteins that perform many vital cellular functions, despite lacking a well defined equilibrium structure. Because their sequences are relatively enriched in polar residues and depleted in hydrophobic residues, IDPs cannot form a stable hydrophobic core and instead sample an ensemble of disordered conformations in vivo. The unique physical and sequence properties of IDPs have been suggested to confer a wide array of biological advantages for protein-protein interactions, including many-to-one signaling, high-specificity low-affinity binding, and rapid binding kinetics.

Many IDPs fold to a stable structure when interacting with appropriate binding partners, but the resulting disordered protein complexes significantly differ from those involving well-folded proteins. In contrast to ordered protein complexes, which involve a relatively small fraction of the protein surface and are stabilized by polar interactions of residues that are distant in the protein sequence, disordered protein complexes form extensive interfaces that are largely stabilized by hydrophobic interactions of residues that are close in sequence. Nussinov and coworkers have suggested that the coupled folding-binding of IDPs results from interactions that reshape the free energy landscape for disordered proteins. Moreover, disordered proteins often retain
considerable disorder after binding, resulting in somewhat “fuzzy” complexes. The landscape reshaping paradigm has recently been adapted to the description of fuzzy complex formation, which may reflect traces of landscape reshaping from the process of their formation. The Sic1 ubiquitin ligase system, which has been extensively characterized by NMR spectroscopy, offers unparalleled insight into the nature of fuzzy complexes and their free energy landscapes. Recent atomistic simulations of IDP complexes have also contributed insight into the mechanisms of folding-upon-binding events and the possible significance of “fly-casting” for this process. Nevertheless, many questions remain regarding the general principles governing IDP binding, and about FCP1 binding in particular.

Bioinformatic studies have revealed that IDRs are especially prevalent in eukaryotic organisms, where they function in transcription factors, acidic transactivation domains, interactions bridging the pre-initiation complex and mediator, and in other components of the eukaryotic transcription machinery. Disordered regions can effectively link globular domains, and the relatively large surface area of disordered protein complexes optimizes molecular recognition while requiring a minimal number of amino acid residues to achieve specificity. The RAP74-FCP1 interaction investigated in the present paper is one particularly interesting example of a disordered protein complex involved in regulating transcription.

Eukaryotic transcription of mRNA by RNA polymerase II (PolII) is tightly regulated by a collection of transcription factors, including transcription factor IIF (TFIIF), and also by the phosphorylation state of a repeated heptapeptide sequence in the disordered C-terminal domain (CTD) of PolII. The CTD repeated heptapeptide
becomes hyperphosphorylated during the course of transcription. The TFIIF-associating CTD Phosphatase (FCP1) performs an essential function in regulating transcription by dephosphorylating the CTD so that PolII can be recycled for another round of transcription. This function is promoted by an interaction between the disordered C-terminal domain of FCP1 and the C-terminal winged-helix (WH) domain of the TFIIF RAP74 subunit.

The C-terminal domain of FCP1 bears striking similarity to acidic transactivation domains and can function as a transactivator in vitro when fused to the GAL4 DNA binding domain. Similarly to other transactivation domains, the FCP1 C-terminal domain is enriched in acidic residues and contains relatively few hydrophobic residues, which are organized in several clusters. Consequently, the FCP1 tail cannot form a stable hydrophobic core and remains largely disordered in solution, although it should be noted that the Showalter laboratory has obtained complete chemical shift assignments that have recently identified nascent helical character in the disordered ensemble for this tail.

In the presence of the C-terminal WH domain of RAP74, the C-terminal 17 residues of FCP1 fold to form a well-defined amphipathic helix. This helix binds to a hydrophobic groove lined by two helices of RAP74. Crystal and NMR structures have provided insight into molecular interactions between the proteins, but do not provide a quantitative understanding of the spatio-temporal dynamics involved in binding and in maintaining the complex. The current literature contains no reports of experimental (e.g. NMR spin relaxation) or computational investigations into the dynamics of the RAP74-FCP1 interaction. Consequently, detailed simulation studies of the RAP74-
FCP1 system should contribute additional insight into the spatio-temporal dynamics of binding IDPs and the interactions stabilizing these complexes.

The FCP1-RAP74 complex is also intriguing because this interaction is mediated by the basic WH domain of RAP74. The WH motif is a variation on the all α-helical helix-turn-helix motif that generally functions as a DNA binding domain. The WH domain has a compact α/β structure consisting of three α helices (H1, H2, and H3), three β strands (S1, S2, S3), and two wings (W1 and W2) arranged in order H1-S1-H2-H3-S2-W1-S3-W2; RAP74 contains an additional short helix (termed H2.5) between H2 and H3 (annotated in Figure 6.1). The canonical mode of DNA binding by WH domains involves extensive polar interactions mediated by the recognition helix, H3. The WH domain in the RAP30 chain of TFIIF also displays chemical shift changes in H1 upon DNA binding, suggesting an alternate binding mode involving this helix, long range conformational change induced by ligand binding, or both.

The structure of FCP1 in complex with RAP74 (Figure 6.1A) reveals a binding site created by the shallow groove between H2 and H3, adjacent to the canonical DNA binding interface. Previous studies have shown that the hydrophobic patch bordered by H2 and H3 mediates protein-protein interactions, particularly when the domains are bound to DNA. This hypothesis is supported by the DNA mediated heterodimerization of the winged-helix domains from DP2 and E2F4 in which H3 from each protein interacts simultaneously with the major groove of DNA and with the H2/H3 interface of the adjacent protein. The quaternary Cul1-Rbx1-Skp1-Fbox°Skp2 complex also features a protein-protein interaction mediated by the H2/H3 interface. To our knowledge, no
Figure 6.1: Sequence and tertiary complex of RAP74 and FCP1. Sequence schematic and a ribbon diagram representing the crystal structure of FCP1 (green) bound to RAP74 (red) (PDB: 1J2X). (A) Secondary structures are annotated over the primary sequence of each protein. (B) The ribbon diagram shows that FCP1 lays across helix 2 (H2) and helix 3 (H3) of RAP74.
other high resolution structures of protein-protein complexes involving a winged-helix domain have been experimentally determined in the absence of DNA. The role of the H2/H3 interface in mediating protein-protein interactions is further supported by mutagenesis studies of the T4 transcription factor MotA\textsuperscript{41} and by the structure of p53 bound to MDM2 (which bears topological similarity to a WH).\textsuperscript{42} Consequently, simulation studies of the RAP74-FCP1 complex should identify interactions stabilizing this complex and reveal fundamental insight into the potentially novel protein-protein interactions involving WH domains.

Motivated by these considerations, we have performed atomically detailed explicit solvent molecular dynamics studies of the RAP74 winged-helix in the apo-state and in complex with FCP1. Our \textit{in silico} studies of apo-RAP74 and the RAP74-FCP1 complex elucidate the atomistic interactions stabilizing the complex, their effect upon protein dynamics, and the mode of binding FCP1 by RAP74. These simulations demonstrate that this binding not only induces FCP1 folding, but also reorganizes and stabilizes the RAP74 WH domain. Most significantly, though, these simulations reveal that FCP1 retains substantial conformational flexibility in the RAP74-FCP1 complex, that is, the complex is fuzzy. The significant disorder of bound FCP1, coupled with our previous NMR studies revealing partial order in unbound FCP1, suggests a relatively limited entropic cost associated with the folding-upon-binding of IDPs.

6.3 Materials and Methods:

Two MD simulations were performed with the AMBER 11.0 software package\textsuperscript{43} using the ff99SB\textsuperscript{44,45} force field and the SPC water model\textsuperscript{46} under particle-mesh Ewald
periodic boundary conditions. Both simulations were derived from the crystal structure of RAP74 in complex with FCP1 (pdb 1J2X): RAP74 apo and RAP74-FCP1 holo. The RAP74 apo simulation modeled residues 451-517 from human RAP74; the holo simulation also modeled residues 944-961 from human FCP1. Two and seven chloride ions were added to neutralize the net positive charge of the proteins in the holo and apo simulations, respectively. The resulting systems were solvated such that no solute atom was within 10 Å of a box edge, requiring 6019 waters for the holo simulation and 5340 for the apo simulation. The starting configurations were energy-minimized and equilibrated as previously reported. After an initial equilibration period, each system was simulated for 250 ns with a 2.0 fs timestep in the isothermal-isobaric (NPT) ensemble. Snapshots from each trajectory were stored to disk every 1.0 ps. The trajectories were analyzed with the AMBER software package, using the ptraj program, with VMD, with MATLAB (The MathWorks, Natick, MA), with GROMACS utilities, and also with in-house software. Molecular graphics images were created using the UCSF Chimera package.

6.4 Results and Discussion:

6.4.1 Atomic Displacements in the Simulations:

The backbone root mean-square deviation (RMSD) from the crystal structure was calculated to determine protein stability during each simulation. The unstructured tails of RAP74 (N-terminal residues 451-454 and C-terminal residue 517) and FCP1 (residues 944 and 961) were excluded from the calculation. The apo simulation (Figure 6.2A, blue) reaches a maximum backbone RMSD near 3.0 Å, which is slightly higher
Figure 6.2: RMSD, RMSF, and PDBs of apo-RAP74 and RAP74-FCP1. (A) RMSD traces show the overall stability of the two simulations, RAP74 apo simulation (blue) and the holo simulation (black). The isolated proteins from the holo simulation are also plotted, FCP1 (green) and RAP74 (red), which shows that the high RMSD in the holo simulation comes from deviations in FCP1. (B) Experimental B-factors (black) and computational RMSF by residue (colored as per panel A). Ribbon bundles of the apo state (C) and the holo state (D) show there is no loss of secondary structure during the simulations. All bundles were created by taking structures from the simulation every 50 ns.
than expected for a single globular domain. Nevertheless, the Ca root mean-square fluctuation (RMSF) (Figure 6.2B) and ribbon bundle (Figure 6.2C) demonstrate that RAP74 retains its secondary structure and that the relatively high RMSD reflects the mobility of the wing subdomain and helix 2.5. The holo simulation (Figure 3.2A, black) has a slightly higher RMSD with a maximum of approximately 3.9 Å. The corresponding RMSF and ribbon bundle demonstrate (Figure 6.2B,D) greater flexibility in the RAP74 wing subdomain (upper left), as observed for the apo-state. In both simulations the average backbone RMSD of RAP74 is well below 2.0 Å (1.7 Å and 1.4 Å in the apo and the holo trajectories, respectively), indicating the overall stability of the trajectories.

Previous studies have demonstrated that the ff99SB parameters used in these simulations accurately reproduce the structure, dynamics, and in particular, the magnitude of loop dynamics for small globular systems,\textsuperscript{44, 48, 53} however, it is still important to cross-validate the current simulations with experimental data. Given that no reports of NMR spin relaxation measurement exist for this system, we compare the crystallographic B-factors reported for the complex and the RMSFs observed in the trajectories. Qualitatively, it is clear that the RMSF of RAP74 tracks the experimental B-factors well in all regions except the wing subdomain, which appears to be overly flexible in both trajectories (Figure 6.2B). The largest B-factors found in the complex are almost all found in FCP1. While our trajectory appears to reproduce the extent of motion in the N-terminal portion of the peptide well, the RMSF of the C-terminal region is much lower than the B-factors, suggesting either the presence of dynamics occurring on timescales longer than the simulation or heterogeneity in the crystal.
6.4.2 Burial of the Solvent-Accessible Surface Area (SASA) in the RAP74-FCP1 Complex:

Upon binding RAP74, the disordered C-terminus of FCP1 folds into an amphipathic helix that presents it hydrophobic face to cover a long, but relatively shallow hydrophobic groove on the RAP74 surface.\textsuperscript{33, 36} Nussinov and coworkers have suggested that intrinsic disorder confers a functional advantage by allowing a larger fraction of the binding protein to contribute to interface formation.\textsuperscript{9} Analysis of the co-crystal structure supports this hypothesis. To further characterize this interface, we calculated the average solvent accessible surface area (SASA) of each residue in the RAP74-FCP1 complex using configurations sampled from the holo simulation. We then compared these averages with the corresponding averages for residues in each isolated protein after removing its binding partner from the sampled configurations. A per residue comparison of SASA between residues in the isolated proteins and in the complex (Figure 6.3) demonstrates that less than 15% of residues experience a significant decrease of SASA in the complex. During the simulation, the complex buries an average of 1256 Å\textsuperscript{2} of SASA, including 614 Å\textsuperscript{2} from RAP74 and 642 Å\textsuperscript{2} from FCP1, neglecting changes in FCP1 SASA due to forming an α-helix. FCP1 illustrates the ability of IDPs to achieve high specificity through burial of a large surface area while utilizing a minimal number of amino acid residues to do so.\textsuperscript{9} The majority of the buried surface area arises from the hydrophobic face of FCP1 covering the hydrophobic groove of RAP74 and, in particular, burial of Met-949, Leu-953, and Leu-957.

These changes in SASA also reflect reorganization of the RAP74 binding pocket. A total of eight residues from RAP74 become solvent inaccessible in the bound state:
**Figure 6.3:** Difference in SASA of RAP74 and FCP1 when bound. The difference in SASA shows the specific residues from RAP74 and FCP1 that are buried upon binding. In FCP1 (green), the major residues that are buried are Met-949, Leu-953, and Leu-957, which are all along the hydrophobic face of the amphipathic α-helix. In RAP74 (red), the major residues that are buried are Thr-470, Lys-471 and Leu-474 from helix 2, Ser-486, Val-490, Ala-494, and Lys-498 from helix 3, and Arg-504 from strand 2. Secondary structure elements are depicted above the graph.
Thr-470, Lys-471, and Leu-474 from H2; Ser-486, Val-490, Ala-494, and Lys-498 from H3; and Arg-504 from S2. Comparison of SASAs for the holo and apo simulations suggests that the RAP74 hydrophobic groove expands to accommodate FCP1. This expansion exposes several residues for interacting with FCP1, including Thr-470 (in H2), which has been implicated in an important polar contact with Glu-954, as well as Leu-474 (in H2), Val-490, Val-492, and Leu-493 (all in H3), which form contacts with key hydrophobic residues in FCP1: Met-949, Leu-953, and Leu-957.

6.4.3 Interactions Stabilizing the RAP74-FCP1 Complex:

Complexes involving two ordered proteins tend to be stabilized by polar interactions, whereas complexes involving IDPs are often depleted in polar contacts. As described above, the hydrophobic face of the amphipathic FCP1 helix lies across the RAP74 hydrophobic binding groove, but acidic residues near the FCP1 helix termini are positioned in close proximity with basic residues flanking the RAP74 binding pocket. The present simulations supplement previous experimental structural studies by microscopically characterizing the key interactions that stabilize the RAP74-FCP1 complex and by providing a dynamic description of the complex.

6.4.3.1 Hydrophobic Interactions:

Both crystallographic and NMR structures indicate that FCP1 buries three key hydrophobic residues in the shallow hydrophobic groove of RAP74: Met-949, Leu-953, and Leu-957. Mutation of these residues to alanine significantly decreases the complex affinity, further demonstrating that these residues are critical for proper binding. These three residues form an extensive network of interaction with RAP74, as evidenced by
the multiple methyl-methyl NOEs recorded;\textsuperscript{33} this result stands in contrast to the unusually large crystallographic B-factors reported for Leu-953 and Leu-957, which suggests that the C-terminal portion of RAP74-bound FCP1 may be poorly ordered in the crystal environment. Appendix C Table C.1 documents the average heavy atom distances separating RAP74 and FCP1 residues identified in the initial structural reports as forming significant contacts. When the heavy atom distances in Table C.1 (Appendix C) are shortened to account for the fact that NOEs measure H-H distances, our RAP74-FCP1 simulation is seen to preserve all contacts mediated by the methyl groups of Met-949, Leu-953, and Leu-957. As described below, our atomically detailed simulation of the RAP74-FCP1 complex clarifies the significance and stability of these contacts.

Closer inspection of the complex reveals that the RAP74 binding pocket is shallow in the vicinity of Met-949 and that this residue appears to cap the interaction interface on the N-terminal end of FCP1. The backbone of the neighboring FCP1 residues retains a stable $\alpha$-helical conformation and both RMSF and B-factors report relatively small fluctuations in the Met-949 backbone. However, the simulation suggests that the Met-949 sidechain fluctuates dramatically, repeatedly escaping from and re-entering the RAP74 hydrophobic groove during the trajectory (indicated in Figure 6.4A by snapshots at 50 ns, 100 ns, 150 ns and 200 ns). The escape of its sidechain significantly increases the Met-949 SASA and correlates with reversible transitions in the rotomeric state of the torsion angle $\chi_2$ (Figure 6.4B and F, respectively) that occur on the nanosecond timescale, but do not appear to correlate strongly with any other fluctuations in Met-949 torsions.
Figure 6.4: Dynamic motion of Met-949. The reversible escape of Met-949 from the RAP74 binding groove is largely due to jumps in $\chi_2$. (A) Representative snapshots taken at 50 ns, 100 ns, 150 ns, and 200 ns zoomed in to show Met-949 (colored by atom) from FCP1 (green ribbon) and its excursions out from and into the RAP74 binding groove (red van der Waals spheres). (B) The SASA of Met-949 as a function of time shows increased solvent exposure when $\chi_2$ is approximately ± 90°. The remaining panels display Met-949 torsion angles $\phi$ (C), $\psi$ (D), $\chi_1$ (E), and $\chi_2$ (F) as a function of time. The angle $\chi_3$ is unconstrained and not shown for clarity.
In contrast, RAP74 residues Thr-470, Leu-474, Leu-493, and Ala-494 form a pocket that locks the Leu-953 sidechain in place and reduces the fluctuations in Leu-953 $\chi_1$ and $\chi_2$ torsions (see Appendix C, Figure C.1 and C.3). The sidechain carbons of Leu-953 remain within 5Å of each RAP74 methyl forming the binding pocket over 40% of the time. Because Leu-957 forms half as many van der Waals contacts as Leu-953, Leu-957 demonstrates greater conformational freedom and its $\chi_1$ and $\chi_2$ torsions sample significant fluctuations, which reposition its methyl groups in the RAP74 groove (see Appendix C, Figure C.2; a less constrained leucine residue is shown in Figure C.3 for contrast). Despite its backbone flexibility, the sidechain of Leu-957 does not escape from the binding groove as Met-949 does. The very low RMSF near Leu-953 and Leu-957, compared with the anomalously high B-factors in this region, suggest that these residues may undergo motions in vitro that are too slow to characterize with MD. Future NMR relaxation-dispersion studies are needed to support this hypothesis. Nevertheless, our MD simulations clearly support the mutagenesis studies by demonstrating the importance of hydrophobic contacts and, in particular, Leu-953 and Leu-957 for stabilizing the RAP74-FCP1 complex.

6.4.3.2 Polar Interactions:

In addition to contacts involving the three key hydrophobic residues (i.e., Met-949, Leu-953 and Leu-957), alanine scanning mutations performed by Kamada et al. suggested that two polar contacts in FCP1 are essential for effective binding to RAP74. Both the crystal and solution NMR structure indicate that Asp-947 at the N-terminal of the FCP1 helix forms a salt bridge with Lys-471 in H2 of RAP74. Mutation of either Asp-947 to alanine or of Lys-471 to glutamate disrupts the RAP74-
Our MD simulations further substantiate the importance of the Lys-471 – Asp-947 salt bridge (Appendix C, Table C.1). Intriguingly, despite the many basic residues lining the RAP74 binding groove and the acidic nature of FCP1, this is the only salt bridge directly confirmed by mutagenesis to significantly influence binding affinity. Upon the basis of NOEs observed between Lys-498 and Leu-960, Nguyen et al. also suggested that the neighboring Asp-959 at the C-terminus of FCP1 may form a salt bridge with Lys-498 in RAP74 H3.33 However, this interaction appears to be poorly preserved in the holo simulation, suggesting that this salt bridge may be relatively weak.

### 6.4.4 FCP1 Alters RAP74 Structure and Dynamics:

The RAP74-FCP1 interaction not only induces FCP1 to fold, but also appears to significantly alter the structure and fluctuations in the folded RAP74 WH domain. Except for fluctuating contacts that involve the terminal residues, the RAP74-FCP1 complex neither breaks nor forms stable intramolecular hydrogen bonds or salt bridges that are not present in apo-RAP74. However, detailed analysis of the simulations reveals that interactions with FCP1 expose key interacting residues in RAP74, as described above, and also lead to a reorganization and stabilization of RAP74 secondary structures.

#### 6.4.4.1 Secondary Structure Analysis:

The FCP1 binding cleft on RAP74 is composed primarily of three α-helix secondary structures, suggesting that helix dipole interactions, which significantly stabilize folded helix bundles, may play a significant electrostatic role in stabilizing the complex.55 When bound to RAP74, the FCP1 helix dipole is equally anti-parallel with
H2 (average angle of -52 +/- 10 degrees) and parallel with the larger dipole of H3 (average angle of +56 +/- 10 degrees), contradicting this premise. However, FCP1 has little effect upon the structure and stability of H2. The distributions for the H2 end-to-end distance and backbone dipole moment are essentially unchanged between the apo and holo simulations. Moreover, the angle between H2 and H3 remains approximately 60° in both simulations, although this angle fluctuates slightly less in the holo simulation (standard deviation reduces from 7 to 4°).

In contrast, interactions with FCP1 significantly stabilize the canonical DNA recognition helix H3. In the apo state, the first turn of H3 (involving Ser-486, Glu-487, Gln-488, and Thr-489) fluctuates in and out of alignment to form a kink at an angle of roughly 50° with respect to the rest of the helix. This kink correlates with an expansion of 2 Å between the alpha carbons of the first and second turns of H3 (Figure 6.5). However, in the presence of bound FCP1, H3 retains a rigid helical structure with stable hydrogen bonding between successive turns and a uniform orientation along the helix.

FCP1 also induces several changes in the structure and fluctuations of regions that are spatially separated from the RAP74 binding pocket. In the holo simulation, H1 tends to sample more compact conformations and also tilts to form an angle that is more nearly (anti-) parallel with respect to H2, giving rise to a periodicity in the difference of SASA between the apo-RAP74 and RAP74-FCP1 complex simulations (Figure C.5, Appendix C). FCP1 also dramatically stabilizes the interaction between S1 and S3. In the apo state, the twist vectors of S1 and S3 form an angle of 50° with very large fluctuations. In contrast, in the holo state, the twist vectors are almost antiparallel with much smaller fluctuations. The RAP74 reorganization also significantly reduces
Figure 6.5: The N-terminal turn of helix 3 in apo-RAP74 instability. The N-terminal turn of helix 3 is unstable in apo-RAP74. (A) Representative snapshots from 100 ns, 125 ns, and 240 ns reveal the reversible unfolding of H3. Panels (B) and (C) present time traces from simulations of apo-RAP74 (black) and the RAP74-FCP1 complex (red). Panel (B) corresponds to the first (486-489) and second (490-493) turns; panel (C) corresponds to the second and third (494-497) turns. In each case, the top and bottom graphs describe the angle formed by successive turns and the corresponding distance, respectively. The direction of each turn was calculated from the average of the helix directors for the four alpha carbons. The distance between turns was calculated from centers of geometry for the alpha carbons. Disorder in H3 of apo-RAP74 is localized to the first turn.
the solvent exposure of strands S2 and S3, including residues Met-511 and Phe-513, both of which make several contacts with FCP1 in the crystal structure. Appendix C contains an expanded analysis of helical interactions in the apo-RAP74 WH and in the RAP74-FCP1 complex.

6.4.4.2 Covariance Analysis:

Covariance analysis provides further evidence that FCP1 significantly stabilizes and rigidifies RAP74. Figure 6.6 presents covariance matrices calculated from the fluctuations of alpha carbons in apo and holo simulations. RAP74 clearly samples many more and much larger correlated motions in the apo state (Figure 6.6A) than in the holo state (Figure 6.6B). The covariance matrix for the holo state reveals that the fluctuations of the central FCP1 residues (950 to 955) are characterized by positive correlations with RAP74 H3 (residues 485 to 499) and strong negative correlation with the RAP74 wing domain (residues 505 to 513). These positive correlations with H3 likely result from the many intermolecular interactions described above. However, despite the presence of many intermolecular interactions between H2 (residues 470 to 476) and FCP1, the covariance matrix indicates relatively little correlated motion involving these regions in the holo simulation (Figure 6.6B).

6.4.5 Structural Disorder of FCP1 in Complex with RAP74:

Published crystal and NMR structures provide a static description of the RAP74-FCP1 complex and demonstrate that the disordered FCP1 C-terminus forms a helix when bound to RAP74. Our simulations indicate that the RAP74 binding groove stabilizes the orientation of FCP1 at an angle of roughly 55° with respect to both H2 and
Figure 6.6: $\alpha$ correlation of apo-RAP74 and RAP74-FCP1. $\alpha$ correlation matrices reveal decreased dynamics of RAP74 in the FCP1 bound state. (A) Correlation matrix for the unbound RAP74 simulation, shown on a consistent scale with the RAP74 portion of the RAP74-FCP1 complex below. The color bar on the right indicates strong positive correlation (red), strong negative correlation (blue), and noncorrelated motion (green). (B) The holo state shows positive correlation between helix 3 (485 to 499) and the middle of FCP1 (950 to 955) and negative correlation between the winged-domain of RAP74 (505 to 513) and the middle of FCP1 (950 to 955). Secondary structures are annotated with bars above each panel.
The groove also maintains a stable helix orientation for FCP1 as the average angle between the helix directors in successive turns is less than 14° with a standard deviation of a few degrees. However, the simulations also demonstrate that FCP1 retains significant disorder in this complex; in addition, to the conformational flexibility of Met-949 described above, the tails of FCP1 remain quite disordered, and the hydrogen bonding between FCP1 turns is relatively unstable.

The radius of gyration for the first 14 FCP1 helical residues fluctuates about 7.2 +/-0.1 Å with occasional fluctuations to a larger radius of 7.8 Å. However, the final four residues of FCP1 (Gln-958, Asp-959, Leu-960, and Met-961) sample much larger fluctuations (Figure 6.7). As discussed earlier, Met-949, Leu-953, and Leu-957 anchor the center and C-terminal regions of the FCP1 helix. The alpha carbons for these residues remain separated by a distance of 6.2 Å with a standard deviation of less than 0.3 Å, reflecting stable hydrogen bonding in this region, and thus a stable α-helix. In contrast, the N-terminal side of the FCP1 helix is significantly less well anchored to the binding groove, although the previously discussed salt bridge mediated by the sidechain of Asp-947 does help hold this end of the protein in place. The distance between the first and second turns of the helix samples significant fluctuations that correlate with an expanded radius of gyration for the helix. For much of the simulation, Ala-946 and Ala-950 anchor the N-terminal turns of the FCP1 helix by making very stable contacts that only rarely expand. However, in the last 10 ns, these two residues separate and the first turn of the FCP1 helix escapes so that the corresponding alpha carbons are separated by 10 Å. At the same time, the backbone dipole of FCP1, which had already
Figure 6.7: Radius of gyration of FCP1. Radius of gyration ($R_g$) distributions indicate disorder in bound FCP1. (A) $R_g$ for full FCP1 (944-961) (black) and for the helical residues (944-957) only (green) as a function of time. (B) The histograms of the same $R_g$ data demonstrate that the FCP1 helix undergoes minor fluctuations about its mean, while the full peptide experiences significant excursions, dominated by the disordered C-terminus.
decreased by roughly 10% from its initial value during the first 150 ns, precipitously drops during the final 10 ns of simulation.

6.5 Conclusions:

The present work reports the first simulations investigating the WH domain of the larger (RAP74) subunit of TFIIF\textsuperscript{56} and its interaction with FCP1.\textsuperscript{29} FCP1 performs an essential function by dephosphorylating the C-terminal domain of RNA Polymerase II prior to re-initiation of transcription.\textsuperscript{30} The catalytic activity of FCP1 is promoted when its C-terminal tail binds the RAP74 WH domain.\textsuperscript{57} Upon binding the hydrophobic groove of RAP74, the intrinsically disordered FCP1 tail folds to form an amphipathic helix.\textsuperscript{33, 36} Our atomically detailed simulations of RAP74 in its apo and holo state employed a force field and simulation procedures previously verified to accurately reproduce the dynamics of domains the size of RAP74.\textsuperscript{44, 48, 53} Moreover, these simulations are consistent with a considerable body of experimental results. Consequently, our study supplements previous experimental studies by providing exquisite insight into the interactions and fluctuations characterizing this complex.

As would be expected for an IDP, the RAP74-FCP1 complex significantly differs from interfaces formed by ordered proteins, which typically cover a relatively small fraction of the proteins and which are largely stabilized by interactions between polar and charged residues.\textsuperscript{7, 9} Instead, the RAP74-FCP1 complex forms an extensive interface with the RAP74 hydrophobic binding pocket covering approximately 1200 Å\textsuperscript{2} of SASA, including the hydrophobic surface of FCP1. Our simulations confirm the experimental observation that, despite the presence of many charged and polar residues at the FCP1-RAP74 interface, polar interactions play a relatively small role in
stabilizing the complex. The only significant exception is the very stable Lys-471 – Asp947 salt bridge, which binds the N-terminal side of the FCP1 helix to the N-terminal side of RAP74 H2. Although experimental studies have suggested that Glu-954 – Thr-470 and Asp-959 – Lys-498 form important contacts, these contacts are not very stable in our simulations.\textsuperscript{33, 36} It is possible that these interactions were not well modeled in the simulations; alternatively, these weak polar interactions may contribute to binding affinity through means that are not immediately obvious from observations of the final bound state, such as by distinguishing between parallel and antiparallel binding of FCP1. Intriguingly, helix dipole interactions, which significantly stabilize folded helix bundles,\textsuperscript{55} would appear to destabilize the interaction of RAP74 and FCP1, as the FCP1 dipole is oriented parallel with the relatively long H3.

Our atomistic simulations demonstrate that the RAP74-FCP1 interaction not only induces FCP1 folding, but also impacts the structure and fluctuations of RAP74. The RAP74 hydrophobic groove expands and exposes key residues to bind FCP1. The presence of FCP1 stabilizes H3, which samples kinked conformations in apo RAP74, tightens the wing domain, and also dampens the collective dynamics of the winged helix domain. Intriguingly, experiments have demonstrated that binding of the central IDR of FCP1 (centFCP1-P04) to RAP74 induces an expansion of the H2/H3 cleft and reorganizes interactions between H2.5 and H1.\textsuperscript{58} Our \textit{in silico} results reinforce the conclusions drawn from these experimental studies and suggest that reorganization of the WH domain fine-tunes the binding interaction.

The observations from our atomistic simulations are significant for providing the first quantitative description of the impact protein binding may have on the structure and
dynamics of WH domains. An alternative DNA binding mode for WH domains that directly involves H1 has been suggested by NMR chemical shifts recorded for the RAP30 WH domain in complex with DNA. We note that binding FCP1 also appears to realign H1 on the surface of RAP74 opposite the FCP1 binding groove, which may alter potential interactions between H1 and other proteins or nucleic acids in the pre-initiation complex. While low resolution footprinting experiments have suggested a possible location for H1, its precise orientation and interactions in the pre-initiation complex remain ambiguous.

Most significantly, the simulations also reveal that the FCP1-RAP74 is quite “fuzzy.” Although the RAP74 hydrophobic groove stabilizes the FCP1 helix and maintains a stable orientation between the two proteins, FCP1 retains significant conformational flexibility in the complex. The N-terminal turns of the FCP1 helix fluctuate significantly and completely separate during the last ten nanoseconds of the simulation, which may be consistent with their large B-factors in the crystal structure. The Met-949 sidechain reversibly escapes from and re-enters the RAP74 binding pocket, suggesting that interactions near this N-terminal turn of helix are not maximally stabilizing. Intriguingly, phosphorylation of nearby Ser-942 and Ser-944 by CK2 stabilizes an extra turn of helix in FCP1 and strengthens the RAP74-FCP1 interaction, indicating that fine-tuning the interaction strength in this region may play a critical regulatory role.

Clearly, the current definition of fuzzy complexes is rather vague. Consequently, it is important to compare the fluctuations in the RAP74-FCP1 complex and, in particular, the dynamics of sidechains, such as Met-949, with those that would be
expected for a folded protein complex. Several recent studies have employed methyl groups to probe sidechain dynamics in the binding interfaces of folded protein complexes. In these experimental studies, binding-induced changes range from being very minimal in the barnase-barstar complex\textsuperscript{61} to being much more dramatic, as in the globally reduced sidechain flexibility of a ternary FKBP12-mTOR complex.\textsuperscript{62} In the light of these studies, our simulations appear to suggest that FCP1 samples significantly larger conformational fluctuations than might be expected for folded protein complexes. Nevertheless, direct experimental validation is required to clarify this point and to establish quantitative bounds for distinguishing the “fuzziness” of disordered and folded protein complexes.

In closing, we note that the fuzziness observed in our simulations, in conjunction with previous and ongoing studies in our laboratories are beginning to reveal a mechanism for the binding of FCP1 and, in particular, to suggest that the disorder-order transition associated with FCP1-RAP74 interactions is more gradual than initially expected. Preformed structural elements have been widely recognized for their significant roles in defining the mechanism of IDP interactions with globular proteins.\textsuperscript{63} For example, NMR studies have reported pre-formed motifs embedded within disordered interaction regions of many proteins, including the viral surface antigens\textsuperscript{64} and the N-terminal transactivation domain of p53.\textsuperscript{65, 66} Recently, the idea of conformational fuzziness in IDP complexes has emerged as a further mechanistic signature of IDP binding that reduces the entropic barrier to binding.

In the specific case of FCP1, recent NMR experiments performed by the Showalter laboratory have demonstrated that unbound FCP1 samples helical
conformations in vitro that are sufficiently stable to yield α-helical assignment through the chemical shift index in residues 945-950.\textsuperscript{35} This region corresponds well with the region of FCP1 having the lowest B-factors in the co-crystal structure.\textsuperscript{56} In addition, preliminary studies with native-based Gō potentials suggest that FCP1-binding proceeds through an extended conformational-selection mechanism.\textsuperscript{10} The distribution of FCP1 conformations gradually narrows during the early stages of binding, but the binding-transition is crossed when RAP74 selects a partially folded FCP1 conformation (S.K and W.G.N, unpublished). Clearly, measurements of backbone NMR spin relaxation for both RAP74 and FCP1 in solution should provide much needed quantitative insights into the extent of FCP1 helix stabilization in the unbound state and confirmation of the results presented here for the bound state.

Taken together with ongoing experimental studies in our laboratory, and consistent with these themes from the literature, the present simulations suggest that FCP1 may lose relatively little conformational entropy upon binding RAP74. Consequently, the entropic cost for the coupled folding-binding transition of some IDPs may be smaller than previously anticipated and the order-disorder transition associated with IDP folding may be less pronounced than implied by static crystal structures.

\textbf{6.6 Acknowledgment:}

We thank Chad Lawrence for helpful discussion and for assistance revising the manuscript. This work was supported by a grant from the NSF to SAS (MCB-0953918), by a fellowship awarded to WGN by the Penn State Institute for Cyberscience, and by startup funds from the Pennsylvania State University.
6.7 References


Chapter 7

Future Directions for Studying the RAP74-FCP1 Complex

[The following chapter is designed as proposed research stemming from the results presented in chapter 6]

7.1 Investigate the Adaptability of the Winged-Helix Domain of RAP74 by Running MD simulations of the RAP74 with the Central Domain of FCP1:

The Omichinski group showed that both the central and the C-terminal domains of FCP1 (centFCP1 and cterFCP1, respectively) interact with the winged-helix domain of RAP74 through an acidic/hydrophobic LXXLL-like motif.1-3 Both the centFCP1 and the cterFCP1 contain conserved casein kinase 2 (CK2) phosphorylation sites and the phosphorylation of these sites increases the binding affinity of the domains to the winged-helix domain of RAP74.4 While this data would suggest that the two domains of FCP1 have a similar binding mode, the NMR solution structures of the two complexes reveal significant differences in binding.2, 3 Therefore, the binding groove of RAP74 adapts by varying the shape and charge distribution of the binding groove in order to bind either the centFCP1 or cterFCP1. To study the adaptability of RAP74 and differences in the binding mechanisms, I propose running MD simulations of the RAP74-centFCP1 and comparing it with the MD simulations of RAP74-cterFCP1 (chapter 6).
Complexes involving IDPs are mediated through hydrophobic interactions, which are observed in both the RAP74-centFCP1 and the RAP74-cterFCP1 complexes. In the case of the RAP74-cterFCP1 complex, the lack of polar interactions allows for flexibility of FCP1 when bound (chapter 6). Initially, the burial of amino acid residues in the RAP74-centFCP1 complex can be explored utilizing the difference in the solvent-accessible surface area (SASA). The NMR solution structure established that less total SASA is buried in the RAP74-centFCP1 complex (1160 Å) than the RAP74-cterFCP1 complex (1510 Å). This difference in the SASA is due in part to centFCP1 forming a smaller helix, 12 amino acid residues compared to 17 amino acid residues for cterFCP1; and centFCP1 not binding as deeply into the RAP74 binding groove. Therefore, the RAP74-centFCP1 complex should have an altered RAP74 profile in the SASA difference plot caused by the change in the shape of the RAP74 binding groove and less amino acids in the centFCP1 helix interacting with RAP74.

Based on the results from the difference in SASA profile, the torsion angles for the centFCP1 residues that show large differences in SASA can be calculated and a time trace plotted to assess the flexibility of the side chain. In the RAP74-cterFCP1 complex, two out of three amino acid residues from FCP1 (Met-949 and Leu-957) that are buried in the hydrophobic groove of RAP74 based on crystal and NMR data are quite dynamic and escape from the binding pocket (chapter 6). Due to centFCP1 not binding as deeply into the groove as cterFCP1, it is expected that the side chains will be more flexible and escape the binding pocket more readily than what was observed for cterFCP1. If it is true that the side chains of centFCP1 are more dynamic, than it reasons that the whole of centFCP1 bound to RAP74 will be more flexible as well,
leading to a more dynamic (fuzzier) complex than previously observed for cterFCP1. The flexibility of centFCP1 can further be investigated by plotting the time trace and population of the radius of gyration. A dynamic protein should have a radius of gyration that does fluctuate quite a bit over time and undergoes a large variation about its mean value. Additionally, dynamics of the two complexes can also be investigated through order parameters similar to work done on the apo state of dsRBDs in the miRNA maturation pathway (chapter 3 and chapter 4). The advantage of calculating the order parameters from the MD simulations is that they can be compared with experimental data to validate the results from the simulations.

The major reason for running the simulations is to investigate the adaptability of RAP74 for binding different regions of FCP1, in order to study this, secondary structure and covariance analysis can be performed similar to the investigation for the RAP74-cterFCP1 complex (chapter 6). In the case of the RAP74-cterFCP1 complex, cterFCP1 stabilizes and rigidifies the RAP74 structure even though cterFCP1 is dynamic. Therefore, centFCP1 potentially will stabilize and rigidify RAP74 similar to cterFCP1, but in a slightly different conformation. The difference in conformation can be verified by determining the RMSD between the two RAP74 structures.

7.2 Mimicking the Cellular Environment Through MD Simulations of the RAP74-CterFCP1 Complex with Crowding Agents and Osmolytes:

The cellular environment is very crowded due to it containing many large macromolecules: like complex sugars, proteins, and nucleic acids. This macromolecular crowding in the cell has kinetic and thermodynamic affects on many
aspects of cellular function: notably folding, stability, diffusion and association of macromolecules.5-7 Another component of the cell that contributes to the folding and stability of macromolecules is osmolytes, molecules that combat environmental stresses such as temperature and osmotic pressure.8, 9 Both in vitro and in silico experiments are generally carried out in aqueous buffers containing some ions, but tending to neglect the effect of crowding agents and osmolytes. Only recently, the effect of crowding agents and osmolytes on the folding and dynamics of proteins has been investigated using MD simulations.10-13 While experiments with crowding agents, like Ficoll and Dextran, and osmolytes, like 2,2,2-trifluoroethanol (TFE) and trimethylamine N-oxide (TMAO), of IDPs have demonstrated the propensity of disordered regions to be more structured than originally thought, suggesting that the cellular nature of IDPs might be quite different than original in vitro experiments imply.14-16 Even with the advances in computational methods for modeling crowding agents and osmolytes as well as the experimental data of IDPs with these molecules, the MD literature is still sparse of reports of MD simulations of IDPs with crowding agents and osmolytes.

Current experimental work in the Showalter lab is designed around the effect that both crowding agents and osmolytes have on the structure of apo-FCP1 in order to understand the mechanism of IDP binding in cells, in general, and RAP74 binding of FCP1, specifically. A propensity for apo-FCP1 to form a nascent \( \alpha \)-helix under increasing TFE concentrations,15 suggests a conformational selection mechanism for binding over an induced fit mechanism. To further test this hypothesis of conformational selection in the cellular environment, I propose running simulations that model both macromolecular crowding and osmolyte effects on the RAP74-FCP1 system. In order
to accomplish this three additional all atom simulations can be carried out on the RAP74-FCP1 complex: one incorporating the effect of crowding agents, another incorporating the effect of an osmolyte and finally, one incorporating both. The major reason for breaking up the simulations into three cases is to investigate the effect of crowding agents and osmolytes individually first, which is similar to the experimental setup. Another reason is that to my knowledge no simulation has been done on any system with both crowding agents and osmolytes modeled. Therefore, a divide-and-conquer method of building up the complex nature of the simulation to more accurately model the cellular environment is proposed. These simulations can be evaluated similar to the RAP74-FCP1 complex simulation presented in chapter 6. One of the key features that can be assessed are the effects both crowding agents and osmolytes have on the dynamics of both RAP74 and FCP1 in the complex. In chapter 6, it was shown that FCP1 in the complex is quite dynamic including residues Met-949 and Leu-957, which are in the hydrophobic groove of RAP74 in the crystal and NMR structures. It will be interesting to see by modeling cellular conditions if the complex is still heavily dynamic or has more rigidity. Additional simulations of apo-RAP74 could be run with the crowding agents and osmolyte to compare with the apo-RAP74 simulation in chapter 6 as well as investigate the role crowding agents and osmolytes have on the well folded, winged-helix domain of RAP74 prior to binding.

While all atom simulations are valuable for studying the structure and fluctuations of the complex, it does not effectively model the conformational space of unfolded proteins, like FCP1. Due to this limitation, the Showalter lab works in collaboration with the Noid lab to run coarse-grained simulations of apo-FCP1 and the RAP74-FCP1
complex. Therefore, it could be beneficial to run coarse-grained simulations of apo-FCP1 and the RAP74-FCP1 complex with both crowding agents and osomlytes in a similar manner as the above proposed all atom simulations. These coarse-grained simulations in connection with the experimental and computational data should provide further insight into the binding mechanism of the IDP, FCP1, binding to the winged-helix domain of RAP74.

7.3 References:


13. Yu, I.; Nakada, K.; Nagaoka, M., Spatio-temporal characteristics of the transfer free energy of apomyoglobin into the molecular crowding condition with trimethylamine


Appendix A

Supporting Information: Chapter 3


EMSAs were run by Kaycee A. Quarles. Figure A.1 was made by Kaycee Quarles and Figure A.2 was jointly made by Kayce A. Quarles and Christopher Wostenberg.

Figure A.1: EMSA of pri-miR-16-1 with Drosha-dsRBD. EMSA gel of Drosha-dsRBD that demonstrates Drosha-dsRBD does not bind pri-miR-16-1. Protein concentration (μM) is shown below the lanes.
**Figure A.2:** Different fits of EMSA of pri-miR-16-1 with DGCR8-dsRBD1. EMSA assay of pri-miR-16-1 binding to DGCR8-dsRBD1. (A) Predicted secondary structure of pri-miR-16-1 with the sequence of the mature miRNA shown in red, and the region removed by Drosha cleavage indicated through lower-case letters. (B) Representative gel showing addition of 2-200 μM DGCR8-dsRBD1 to 0.25 nM pri-miR-16-1. (C) Fitted EMSA fraction bound as a function of DGCR8-dsRBD1 concentration with data points and uncertainties represented by filled circles. The N-independent and identical sites best fit to the data is shown in red and the fully cooperative best fit to the data is shown in grey.
Figure A.3: $^{15}$N-HSQC spectra of Drosha-dsRBD and DGCR8-RBD1. The representative $^1$H,$^{15}$N-HSQC spectra of (A) Drosha-dsRBD and (B) DGCR8-dsRBD1.
Table A.1: Chemical shifts (ppm) from the backbone assignment of Drosha-dsRBD.

<table>
<thead>
<tr>
<th>Residue</th>
<th>N</th>
<th>H</th>
<th>CO</th>
<th>CA</th>
<th>CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1258</td>
<td></td>
<td></td>
<td>173.7</td>
<td>45.27</td>
<td>-</td>
</tr>
<tr>
<td>N1259</td>
<td>119.4</td>
<td>7.82</td>
<td>176.2</td>
<td>54.98</td>
<td>34.04</td>
</tr>
<tr>
<td>D1260</td>
<td>122.8</td>
<td>8.51</td>
<td>174.8</td>
<td>54.61</td>
<td>29.14</td>
</tr>
<tr>
<td>P1261</td>
<td></td>
<td></td>
<td>178.0</td>
<td>65.41</td>
<td>32.80</td>
</tr>
<tr>
<td>K1262</td>
<td>117.0</td>
<td>7.68</td>
<td>180.6</td>
<td>60.41</td>
<td>31.86</td>
</tr>
<tr>
<td>S1263</td>
<td>119.0</td>
<td>8.12</td>
<td>176.8</td>
<td>61.38</td>
<td>62.25</td>
</tr>
<tr>
<td>Q1264</td>
<td>121.9</td>
<td>8.32</td>
<td>178.5</td>
<td>58.76</td>
<td>28.87</td>
</tr>
<tr>
<td>L1265</td>
<td>120.3</td>
<td>8.44</td>
<td>176.9</td>
<td>58.27</td>
<td>41.06</td>
</tr>
<tr>
<td>Q1266</td>
<td>116.7</td>
<td>7.67</td>
<td>177.8</td>
<td>58.42</td>
<td>28.19</td>
</tr>
<tr>
<td>Q1267</td>
<td>116.8</td>
<td>8.13</td>
<td>179.0</td>
<td>58.72</td>
<td>27.97</td>
</tr>
<tr>
<td>C1268</td>
<td>199.3</td>
<td>8.16</td>
<td>177.5</td>
<td>58.37</td>
<td>39.24</td>
</tr>
<tr>
<td>C1269</td>
<td>119.9</td>
<td>8.50</td>
<td>177.6</td>
<td>62.02</td>
<td>33.332</td>
</tr>
<tr>
<td>L1270</td>
<td>120.2</td>
<td>7.92</td>
<td>179.0</td>
<td>57.56</td>
<td>41.62</td>
</tr>
<tr>
<td>T1271</td>
<td>111.4</td>
<td>7.71</td>
<td>175.4</td>
<td>64.23</td>
<td>69.66</td>
</tr>
<tr>
<td>L1272</td>
<td>121.1</td>
<td>7.40</td>
<td>176.9</td>
<td>54.82</td>
<td>41.61</td>
</tr>
<tr>
<td>R1273</td>
<td>120.8</td>
<td>7.52</td>
<td>176.0</td>
<td>56.38</td>
<td>30.65</td>
</tr>
<tr>
<td>T1274</td>
<td>119.1</td>
<td>8.61</td>
<td>174.2</td>
<td>61.44</td>
<td>70.29</td>
</tr>
<tr>
<td>E1275</td>
<td>124.3</td>
<td>8.63</td>
<td>177.1</td>
<td>57.60</td>
<td>29.52</td>
</tr>
<tr>
<td>G1276</td>
<td>111.5</td>
<td>8.62</td>
<td>173.7</td>
<td>45.17</td>
<td>-</td>
</tr>
<tr>
<td>K1277</td>
<td>118.3</td>
<td>8.26</td>
<td>174.4</td>
<td>52.77</td>
<td>39.02</td>
</tr>
<tr>
<td>E1278</td>
<td>119.9</td>
<td>8.17</td>
<td>174.1</td>
<td>52.05</td>
<td>40.88</td>
</tr>
<tr>
<td>P1279</td>
<td></td>
<td></td>
<td>176.2</td>
<td>62.90</td>
<td>32.04</td>
</tr>
<tr>
<td>D1280</td>
<td>121.7</td>
<td>8.28</td>
<td>174.6</td>
<td>54.79</td>
<td>43.25</td>
</tr>
<tr>
<td>I1281</td>
<td>116.5</td>
<td>8.17</td>
<td>174.0</td>
<td>58.06</td>
<td>40.14</td>
</tr>
<tr>
<td>P1282</td>
<td></td>
<td></td>
<td>174.8</td>
<td>63.17</td>
<td>32.05</td>
</tr>
<tr>
<td>L1283</td>
<td>124.6</td>
<td>8.49</td>
<td>174.8</td>
<td>53.70</td>
<td>45.13</td>
</tr>
<tr>
<td>Y1284</td>
<td>123.8</td>
<td>8.88</td>
<td>176.2</td>
<td>57.78</td>
<td>39.77</td>
</tr>
<tr>
<td>K1285</td>
<td>122.4</td>
<td>9.29</td>
<td>175.0</td>
<td>55.08</td>
<td>36.87</td>
</tr>
<tr>
<td>T1286</td>
<td>122.7</td>
<td>8.91</td>
<td>174.8</td>
<td>63.21</td>
<td>68.08</td>
</tr>
<tr>
<td>L1287</td>
<td>129.1</td>
<td>9.22</td>
<td>177.2</td>
<td>55.44</td>
<td>43.28</td>
</tr>
<tr>
<td>Q1288</td>
<td>115.7</td>
<td>7.66</td>
<td>173.9</td>
<td>55.49</td>
<td>32.05</td>
</tr>
<tr>
<td>T1289</td>
<td>119.5</td>
<td>8.38</td>
<td>173.3</td>
<td>61.87</td>
<td>70.65</td>
</tr>
<tr>
<td>V1290</td>
<td>121.6</td>
<td>8.22</td>
<td>174.8</td>
<td>60.29</td>
<td>35.49</td>
</tr>
<tr>
<td>G1291</td>
<td>110.6</td>
<td>8.25</td>
<td>172.1</td>
<td>43.76</td>
<td>-</td>
</tr>
<tr>
<td>P1292</td>
<td></td>
<td></td>
<td>177.7</td>
<td>62.12</td>
<td>32.37</td>
</tr>
<tr>
<td>S1293</td>
<td>114.4</td>
<td>8.78</td>
<td>177.7</td>
<td>62.06</td>
<td>62.84</td>
</tr>
<tr>
<td>H1294</td>
<td></td>
<td></td>
<td>174.8</td>
<td>56.04</td>
<td>30.36</td>
</tr>
<tr>
<td>A1295</td>
<td>128.5</td>
<td>7.59</td>
<td>174.3</td>
<td>51.61</td>
<td>18.03</td>
</tr>
<tr>
<td>R1296</td>
<td>123.0</td>
<td>8.08</td>
<td>177.0</td>
<td>55.34</td>
<td>32.15</td>
</tr>
<tr>
<td>T1297</td>
<td>116.2</td>
<td>8.57</td>
<td>173.5</td>
<td>62.39</td>
<td>70.65</td>
</tr>
<tr>
<td>Y1298</td>
<td>127.6</td>
<td>9.22</td>
<td>174.9</td>
<td>57.32</td>
<td>40.67</td>
</tr>
<tr>
<td>T1299</td>
<td>117.1</td>
<td>8.75</td>
<td>173.9</td>
<td>61.78</td>
<td>70.15</td>
</tr>
<tr>
<td>V1300</td>
<td>128.2</td>
<td>9.40</td>
<td>172.7</td>
<td>59.98</td>
<td>35.94</td>
</tr>
<tr>
<td>A1301</td>
<td>126.3</td>
<td>8.87</td>
<td>174.5</td>
<td>49.56</td>
<td>23.77</td>
</tr>
<tr>
<td>V1302</td>
<td>118.7</td>
<td>8.16</td>
<td>173.0</td>
<td>57.62</td>
<td>34.53</td>
</tr>
<tr>
<td>Y1303</td>
<td>126.7</td>
<td>9.28</td>
<td>175.2</td>
<td>56.32</td>
<td>42.90</td>
</tr>
<tr>
<td>F1304</td>
<td>119.3</td>
<td>9.04</td>
<td>175.1</td>
<td>56.66</td>
<td>42.83</td>
</tr>
<tr>
<td>K1305</td>
<td>128.3</td>
<td>9.08</td>
<td>176.3</td>
<td>57.58</td>
<td>29.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>G1306</td>
<td>104.3</td>
<td>8.94</td>
<td>173.6</td>
<td>45.16</td>
<td>-</td>
</tr>
<tr>
<td>E1307</td>
<td>121.2</td>
<td>7.73</td>
<td>175.2</td>
<td>54.28</td>
<td>31.76</td>
</tr>
<tr>
<td>R1308</td>
<td>126.7</td>
<td>8.96</td>
<td>176.7</td>
<td>56.95</td>
<td>30.06</td>
</tr>
<tr>
<td>I1309</td>
<td>121.9</td>
<td>8.95</td>
<td>174.8</td>
<td>61.37</td>
<td>39.81</td>
</tr>
<tr>
<td>G1310</td>
<td>107.4</td>
<td>7.24</td>
<td>172.2</td>
<td>45.43</td>
<td>-</td>
</tr>
<tr>
<td>C1311</td>
<td>120.8</td>
<td>9.75</td>
<td>172.3</td>
<td>57.04</td>
<td>36.41</td>
</tr>
<tr>
<td>G1312</td>
<td>111.5</td>
<td>9.35</td>
<td>170.5</td>
<td>44.58</td>
<td>-</td>
</tr>
<tr>
<td>K1313</td>
<td>117.1</td>
<td>8.52</td>
<td>176.7</td>
<td>53.96</td>
<td>37.15</td>
</tr>
<tr>
<td>G1314</td>
<td>106.0</td>
<td>8.58</td>
<td>171.0</td>
<td>45.99</td>
<td>-</td>
</tr>
<tr>
<td>P1315</td>
<td>-</td>
<td>-</td>
<td>174.7</td>
<td>63.70</td>
<td>31.81</td>
</tr>
<tr>
<td>S1316</td>
<td>111.5</td>
<td>7.38</td>
<td>173.7</td>
<td>56.11</td>
<td>66.6</td>
</tr>
<tr>
<td>I1317</td>
<td>124.4</td>
<td>8.45</td>
<td>177.4</td>
<td>65.23</td>
<td>37.08</td>
</tr>
<tr>
<td>Q1318</td>
<td>118.1</td>
<td>8.36</td>
<td>178.8</td>
<td>59.57</td>
<td>27.88</td>
</tr>
<tr>
<td>Q1319</td>
<td>116.8</td>
<td>7.81</td>
<td>178.8</td>
<td>58.92</td>
<td>29.18</td>
</tr>
<tr>
<td>A1320</td>
<td>124.7</td>
<td>8.09</td>
<td>179.0</td>
<td>55.86</td>
<td>18.44</td>
</tr>
<tr>
<td>E1321</td>
<td>117.9</td>
<td>8.72</td>
<td>178.7</td>
<td>59.66</td>
<td>29.83</td>
</tr>
<tr>
<td>M1322</td>
<td>116.8</td>
<td>7.80</td>
<td>178.6</td>
<td>58.94</td>
<td>32.01</td>
</tr>
<tr>
<td>G1323</td>
<td>106.2</td>
<td>8.15</td>
<td>176.3</td>
<td>47.27</td>
<td>-</td>
</tr>
<tr>
<td>A1324</td>
<td>127.9</td>
<td>8.60</td>
<td>178.5</td>
<td>55.08</td>
<td>17.64</td>
</tr>
<tr>
<td>A1325</td>
<td>120.3</td>
<td>8.39</td>
<td>179.5</td>
<td>55.45</td>
<td>17.45</td>
</tr>
<tr>
<td>M1326</td>
<td>116.1</td>
<td>8.30</td>
<td>178.3</td>
<td>59.09</td>
<td>32.96</td>
</tr>
<tr>
<td>D1327</td>
<td>120.4</td>
<td>7.99</td>
<td>177.7</td>
<td>57.78</td>
<td>43.24</td>
</tr>
<tr>
<td>A1328</td>
<td>119.8</td>
<td>8.10</td>
<td>180.0</td>
<td>54.81</td>
<td>19.07</td>
</tr>
<tr>
<td>L1329</td>
<td>117.1</td>
<td>8.46</td>
<td>179.3</td>
<td>58.15</td>
<td>42.25</td>
</tr>
<tr>
<td>E1330</td>
<td>117.7</td>
<td>7.83</td>
<td>178.7</td>
<td>58.72</td>
<td>30.10</td>
</tr>
<tr>
<td>K1331</td>
<td>117.5</td>
<td>7.88</td>
<td>177.6</td>
<td>57.74</td>
<td>33.88</td>
</tr>
<tr>
<td>Y1332</td>
<td>120.5</td>
<td>8.22</td>
<td>174.8</td>
<td>59.57</td>
<td>38.38</td>
</tr>
<tr>
<td>N1333</td>
<td>121.1</td>
<td>7.56</td>
<td>173.7</td>
<td>52.48</td>
<td>38.09</td>
</tr>
<tr>
<td>F1334</td>
<td>122.2</td>
<td>7.92</td>
<td>173.8</td>
<td>56.92</td>
<td>38.70</td>
</tr>
<tr>
<td>P1335</td>
<td>-</td>
<td>-</td>
<td>176.4</td>
<td>63.10</td>
<td>29.12</td>
</tr>
<tr>
<td>Q1336</td>
<td>120.6</td>
<td>8.39</td>
<td>175.2</td>
<td>55.80</td>
<td>29.39</td>
</tr>
<tr>
<td>M1337</td>
<td>127.3</td>
<td>7.99</td>
<td>-</td>
<td>56.94</td>
<td>33.76</td>
</tr>
</tbody>
</table>
Table A.2: Chemical shifts (ppm) from the backbone assignment of DGCR8-dsRBD1.

<table>
<thead>
<tr>
<th>Residue</th>
<th>N</th>
<th>H</th>
<th>CO</th>
<th>CA</th>
<th>CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>G504</td>
<td>-</td>
<td>-</td>
<td>173.7</td>
<td>45.12</td>
<td>-</td>
</tr>
<tr>
<td>I505</td>
<td>119.5</td>
<td>7.95</td>
<td>175.7</td>
<td>60.83</td>
<td>38.84</td>
</tr>
<tr>
<td>P507</td>
<td>-</td>
<td>-</td>
<td>175.7</td>
<td>63.5</td>
<td>31.83</td>
</tr>
<tr>
<td>N508</td>
<td>123.7</td>
<td>7.89</td>
<td>177.8</td>
<td>54.76</td>
<td>40.74</td>
</tr>
<tr>
<td>G509</td>
<td>114.0</td>
<td>8.29</td>
<td>174.2</td>
<td>45.55</td>
<td>-</td>
</tr>
<tr>
<td>K510</td>
<td>120.8</td>
<td>8.01</td>
<td>176.1</td>
<td>56.08</td>
<td>34.02</td>
</tr>
<tr>
<td>S511</td>
<td>118.0</td>
<td>8.47</td>
<td>174.8</td>
<td>57.54</td>
<td>65.01</td>
</tr>
<tr>
<td>E512</td>
<td>123.3</td>
<td>9.993</td>
<td>178.4</td>
<td>62.03</td>
<td>28.31</td>
</tr>
<tr>
<td>V513</td>
<td>118.0</td>
<td>7.70</td>
<td>178.7</td>
<td>66.23</td>
<td>32.58</td>
</tr>
<tr>
<td>C514</td>
<td>122.1</td>
<td>7.59</td>
<td>177.4</td>
<td>63.01</td>
<td>27.05</td>
</tr>
<tr>
<td>I515</td>
<td>121.2</td>
<td>8.39</td>
<td>179.1</td>
<td>65.15</td>
<td>38.15</td>
</tr>
<tr>
<td>L516</td>
<td>120.2</td>
<td>8.64</td>
<td>177.6</td>
<td>58.45</td>
<td>40.80</td>
</tr>
<tr>
<td>H517</td>
<td>119.6</td>
<td>8.19</td>
<td>177.8</td>
<td>60.92</td>
<td>31.10</td>
</tr>
<tr>
<td>E518</td>
<td>117.9</td>
<td>8.25</td>
<td>178.8</td>
<td>59.70</td>
<td>29.30</td>
</tr>
<tr>
<td>Y519</td>
<td>120.5</td>
<td>8.29</td>
<td>177.6</td>
<td>62.10</td>
<td>38.66</td>
</tr>
<tr>
<td>M520</td>
<td>116.8</td>
<td>8.89</td>
<td>178.7</td>
<td>59.57</td>
<td>31.21</td>
</tr>
<tr>
<td>Q521</td>
<td>117.9</td>
<td>8.09</td>
<td>177.3</td>
<td>58.76</td>
<td>28.79</td>
</tr>
<tr>
<td>R522</td>
<td>117.5</td>
<td>7.61</td>
<td>177.3</td>
<td>58.67</td>
<td>31.03</td>
</tr>
<tr>
<td>V523</td>
<td>115.9</td>
<td>8.13</td>
<td>176.7</td>
<td>64.74</td>
<td>32.69</td>
</tr>
<tr>
<td>L524</td>
<td>117.7</td>
<td>7.07</td>
<td>176.1</td>
<td>55.58</td>
<td>43.50</td>
</tr>
<tr>
<td>K525</td>
<td>114.6</td>
<td>7.99</td>
<td>175.4</td>
<td>56.62</td>
<td>28.91</td>
</tr>
<tr>
<td>V526</td>
<td>113.0</td>
<td>6.96</td>
<td>173.8</td>
<td>59.72</td>
<td>35.62</td>
</tr>
<tr>
<td>R527</td>
<td>121.3</td>
<td>8.11</td>
<td>174.3</td>
<td>52.83</td>
<td>30.88</td>
</tr>
<tr>
<td>P528</td>
<td>-</td>
<td>-</td>
<td>175.3</td>
<td>62.41</td>
<td>32.41</td>
</tr>
<tr>
<td>V529</td>
<td>122.7</td>
<td>8.17</td>
<td>174.9</td>
<td>61.87</td>
<td>33.56</td>
</tr>
<tr>
<td>Y530</td>
<td>125.4</td>
<td>8.63</td>
<td>174.9</td>
<td>57.59</td>
<td>39.97</td>
</tr>
<tr>
<td>N531</td>
<td>121.3</td>
<td>8.80</td>
<td>173.4</td>
<td>51.68</td>
<td>41.04</td>
</tr>
<tr>
<td>F532</td>
<td>120.2</td>
<td>8.76</td>
<td>175.4</td>
<td>57.48</td>
<td>41.51</td>
</tr>
<tr>
<td>F533</td>
<td>118.7</td>
<td>8.31</td>
<td>173.3</td>
<td>55.73</td>
<td>40.90</td>
</tr>
<tr>
<td>E534</td>
<td>120.7</td>
<td>8.74</td>
<td>175.7</td>
<td>55.62</td>
<td>31.21</td>
</tr>
<tr>
<td>C535</td>
<td>123.0</td>
<td>8.36</td>
<td>174.1</td>
<td>56.69</td>
<td>29.56</td>
</tr>
<tr>
<td>E536</td>
<td>124.8</td>
<td>8.86</td>
<td>175.6</td>
<td>57.01</td>
<td>29.17</td>
</tr>
<tr>
<td>N537</td>
<td>121.8</td>
<td>7.56</td>
<td>174.6</td>
<td>51.01</td>
<td>39.07</td>
</tr>
<tr>
<td>P538</td>
<td>-</td>
<td>-</td>
<td>178.3</td>
<td>64.44</td>
<td>31.86</td>
</tr>
<tr>
<td>S539</td>
<td>110.9</td>
<td>7.80</td>
<td>174.7</td>
<td>59.81</td>
<td>63.63</td>
</tr>
<tr>
<td>E540</td>
<td>118.6</td>
<td>7.17</td>
<td>171.1</td>
<td>53.48</td>
<td>30.56</td>
</tr>
<tr>
<td>P541</td>
<td>-</td>
<td>-</td>
<td>177.4</td>
<td>64.07</td>
<td>32.34</td>
</tr>
<tr>
<td>F542</td>
<td>117.8</td>
<td>7.72</td>
<td>173.5</td>
<td>57.95</td>
<td>40.19</td>
</tr>
<tr>
<td>G543</td>
<td>109.2</td>
<td>8.29</td>
<td>173.0</td>
<td>42.48</td>
<td>-</td>
</tr>
<tr>
<td>A544</td>
<td>123.1</td>
<td>8.69</td>
<td>174.7</td>
<td>50.56</td>
<td>22.73</td>
</tr>
<tr>
<td>S545</td>
<td>114.3</td>
<td>9.00</td>
<td>174.3</td>
<td>56.84</td>
<td>66.69</td>
</tr>
<tr>
<td>V546</td>
<td>122.7</td>
<td>8.31</td>
<td>175.5</td>
<td>61.12</td>
<td>33.27</td>
</tr>
<tr>
<td>T547</td>
<td>119.6</td>
<td>8.93</td>
<td>173.7</td>
<td>59.55</td>
<td>71.52</td>
</tr>
<tr>
<td>I548</td>
<td>121.3</td>
<td>8.79</td>
<td>176.7</td>
<td>61.05</td>
<td>39.37</td>
</tr>
<tr>
<td>D549</td>
<td>130.9</td>
<td>9.46</td>
<td>175.6</td>
<td>55.31</td>
<td>40.03</td>
</tr>
<tr>
<td>G550</td>
<td>101.8</td>
<td>8.53</td>
<td>173.7</td>
<td>45.23</td>
<td>-</td>
</tr>
<tr>
<td>V551</td>
<td>124.0</td>
<td>7.71</td>
<td>175.6</td>
<td>61.80</td>
<td>33.68</td>
</tr>
<tr>
<td>T552</td>
<td>124.8</td>
<td>8.74</td>
<td>174.5</td>
<td>64.39</td>
<td>69.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Y553</td>
<td>128.2</td>
<td>9.15</td>
<td>174.6</td>
<td>57.60</td>
<td>39.19</td>
</tr>
<tr>
<td>G554</td>
<td>110.1</td>
<td>7.90</td>
<td>171.2</td>
<td>45.07</td>
<td>-</td>
</tr>
<tr>
<td>S555</td>
<td>113.3</td>
<td>8.08</td>
<td>172.8</td>
<td>56.84</td>
<td>66.84</td>
</tr>
<tr>
<td>G556</td>
<td>107.5</td>
<td>8.99</td>
<td>170.6</td>
<td>45.59</td>
<td>-</td>
</tr>
<tr>
<td>T557</td>
<td>110.0</td>
<td>8.38</td>
<td>173.7</td>
<td>58.84</td>
<td>71.55</td>
</tr>
<tr>
<td>A558</td>
<td>120.3</td>
<td>9.38</td>
<td>175.9</td>
<td>51.55</td>
<td>23.98</td>
</tr>
<tr>
<td>S559</td>
<td>112.0</td>
<td>8.85</td>
<td>173.4</td>
<td>59.79</td>
<td>63.61</td>
</tr>
<tr>
<td>S560</td>
<td>113.0</td>
<td>7.45</td>
<td>-</td>
<td>56.15</td>
<td>66.72</td>
</tr>
<tr>
<td>K561</td>
<td>-</td>
<td>-</td>
<td>177.8</td>
<td>60.53</td>
<td>32.24</td>
</tr>
<tr>
<td>K562</td>
<td>118.4</td>
<td>8.13</td>
<td>179.2</td>
<td>59.70</td>
<td>32.38</td>
</tr>
<tr>
<td>L563</td>
<td>118.6</td>
<td>7.53</td>
<td>179.4</td>
<td>57.33</td>
<td>42.14</td>
</tr>
<tr>
<td>A564</td>
<td>124.4</td>
<td>8.08</td>
<td>178.9</td>
<td>56.12</td>
<td>17.15</td>
</tr>
<tr>
<td>K565</td>
<td>117.5</td>
<td>8.27</td>
<td>178.0</td>
<td>60.47</td>
<td>32.62</td>
</tr>
<tr>
<td>N566</td>
<td>116.3</td>
<td>7.44</td>
<td>177.6</td>
<td>57.04</td>
<td>38.80</td>
</tr>
<tr>
<td>K567</td>
<td>120.3</td>
<td>8.16</td>
<td>179.1</td>
<td>60.14</td>
<td>33.15</td>
</tr>
<tr>
<td>A568</td>
<td>123.3</td>
<td>8.49</td>
<td>180.9</td>
<td>55.29</td>
<td>18.12</td>
</tr>
<tr>
<td>A569</td>
<td>124.1</td>
<td>8.78</td>
<td>179.2</td>
<td>55.69</td>
<td>17.88</td>
</tr>
<tr>
<td>R570</td>
<td>120.5</td>
<td>8.65</td>
<td>177.5</td>
<td>59.89</td>
<td>29.62</td>
</tr>
<tr>
<td>A571</td>
<td>119.0</td>
<td>8.24</td>
<td>180.4</td>
<td>54.56</td>
<td>18.64</td>
</tr>
<tr>
<td>T572</td>
<td>115.7</td>
<td>7.78</td>
<td>175.7</td>
<td>68.16</td>
<td>67.98</td>
</tr>
<tr>
<td>L573</td>
<td>124.8</td>
<td>8.44</td>
<td>177.1</td>
<td>58.72</td>
<td>41.17</td>
</tr>
<tr>
<td>E574</td>
<td>114.2</td>
<td>7.65</td>
<td>177.9</td>
<td>58.74</td>
<td>29.66</td>
</tr>
<tr>
<td>I575</td>
<td>117.4</td>
<td>7.10</td>
<td>177.7</td>
<td>63.93</td>
<td>39.63</td>
</tr>
<tr>
<td>L576</td>
<td>118.0</td>
<td>7.91</td>
<td>177.5</td>
<td>57.22</td>
<td>44.06</td>
</tr>
<tr>
<td>I577</td>
<td>118.4</td>
<td>8.33</td>
<td>173.1</td>
<td>58.50</td>
<td>38.51</td>
</tr>
<tr>
<td>P578</td>
<td>-</td>
<td>-</td>
<td>177.7</td>
<td>64.46</td>
<td>31.70</td>
</tr>
<tr>
<td>D579</td>
<td>116.3</td>
<td>8.36</td>
<td>176.2</td>
<td>54.43</td>
<td>39.89</td>
</tr>
<tr>
<td>F580</td>
<td>120.6</td>
<td>8.09</td>
<td>176.0</td>
<td>60.16</td>
<td>39.54</td>
</tr>
<tr>
<td>V581</td>
<td>119.4</td>
<td>7.67</td>
<td>175.2</td>
<td>62.13</td>
<td>32.52</td>
</tr>
<tr>
<td>K582</td>
<td>124.8</td>
<td>7.97</td>
<td>175.7</td>
<td>56.30</td>
<td>32.81</td>
</tr>
<tr>
<td>Q583</td>
<td>127.3</td>
<td>7.89</td>
<td>180.5</td>
<td>57.29</td>
<td>30.10</td>
</tr>
</tbody>
</table>
Appendix B

Supporting Information: Chapter 4

Jim Cole performed the AUC experiment and analysis of Dicer-dsRBD with ds16.

Alan Benesi assisted in the NMR diffusion experiments.

Supporting Information Methods and Analysis

Analytical Ultracentrifugation (AUC):

Analytical ultracentrifugation was performed in the Cole lab using methods similar to those previously described.¹ ² Dicer-dsRBD and ds16 were buffer exchanged into 50 mM phosphate and 50 mM potassium chloride buffer, pH 6.0. The fitting gave a $K_d = 5.83 \mu\text{M}$, which agreed well with the EMSA value ($K_d = 8.94 \mu\text{M}$) considering a different buffer was used. The data indicates a one-to-one binding interaction between protein and RNA.

Dynamic Light Scattering (DLS):

Dicer-dsRBD oligomerization state was verified by dynamic light scattering (DLS) on the Viscotek 802 instrument on NMR sample. Measurements were taken every five seconds for one minute and fitted using the Vicsotek software. The software assumes an isotropic global protein and has not been calibrated; therefore the estimation on the molecular weight given is inaccurate. The data does show that at NMR concentration Dicer-dsRBD is mono-dispersed.
**NMR Diffusion:**

The improved diffusion-ordered 2D NMR spectroscopy (DOSY) incorporating bipolar-gradient pulses\(^3\) was used to determine the oligomerization state of an unlabeled sample of Dicer-dsRBD on a Bruker Avance III 850 MHz spectrometer. The data was fit in TopSpin version 2.1 patchlevel 6 in the T\(_1/T_2\) analysis menu to the following equation:

\[
I = I_0 e^{-D \sqrt{2} \gamma_H \text{Gradient} \delta_1 (\delta_1 - \delta_2/3) \times 10^4}
\]  

where \(I\) is the intensity, \(I_0\) is the initial intensity, \(D\) is the diffusion constant (m\(^2\)/s), \(\gamma_H\) is the gyromagnetic ratio of hydrogen (267.513 rad/sT), \(\delta_1\) is the first delta (2.0 m) and \(\delta_2\) is the second delta (20.0 m). This gave a diffusion constant of 1.542e-10 m\(^2\)/s, which corresponds with a molecular weight of 8 kDa.\(^4\) Together with the DLS data this indicates that Dicer-dsRBD is a monomer at the concentration used for NMR experiments.
Figure B.1 Representation of transcription construct to obtain pre-miR-16-11. Representation of the hammerhead (red) with pre-miR-16-1 (green) RNA construct used for transcription. The arrow represents the hammerhead cleavage site, which causes the release of mature pre-miR-16-1. The two cut sites (EcoR1 and Sph1) are for inserting the construct into pUC19. The inverted BsaI site is used to linearize the plasmid to avoid run on transcription.
Figure 6.2: AUC data of Dicer-dsRBD binding ds16. Plots of the AUC data of ds16 with varying concentrations of Dicer-dsRBD show a Svedberg value of 3.31 for the complex corresponding to a one-to-one complex. This data was used to get a $K_d = 5.83 \mu M$. 
Figure B.3: $^{15}$N-HSQC spectrum of Dicer-dsRBD. The representative $^1$H,$^{15}$N-HSQC spectrum of Dicer-dsRBD on the 500 MHz NMR spectrometer with TCI cryoprobe.
Figure B.4: Dicer DLS data at NMR concentration. Dicer DLS data reveals that at NMR concentration, Dicer is mainly mono-dispersed. The high calculated molecular weight of the peak (17.86 kDa) would indicate a dimer, but this is due to the instrument not being calibrated and Dicer being anisotropic as illustrated by the $^{15}$N spin relaxation.
Figure B.5: NMR diffusion of Dicer-dsRBD. Plot of the NMR diffusion data on Dicer-dsRBD reveals a diffusion constant of $1.542 \times 10^{-10} \text{ m}^2/\text{s}$, this value agrees with Dicer-dsRBD being a monomer in solution.
Table B.1: Chemical shifts (ppm) from the backbone assignment of Dicer-dsRBD.

<table>
<thead>
<tr>
<th>Residue</th>
<th>N</th>
<th>H</th>
<th>CO</th>
<th>CA</th>
<th>CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1850</td>
<td>---</td>
<td>---</td>
<td>177.2</td>
<td>63.24</td>
<td>32.13</td>
</tr>
<tr>
<td>R1851</td>
<td>120.8</td>
<td>8.59</td>
<td>176.9</td>
<td>57.10</td>
<td>29.99</td>
</tr>
<tr>
<td>S1852</td>
<td>114.4</td>
<td>7.99</td>
<td>174.4</td>
<td>62.23</td>
<td>69.59</td>
</tr>
<tr>
<td>P1853</td>
<td>---</td>
<td>---</td>
<td>177.8</td>
<td>64.70</td>
<td>31.95</td>
</tr>
<tr>
<td>V1854</td>
<td>117.0</td>
<td>7.45</td>
<td>177.1</td>
<td>67.05</td>
<td>31.80</td>
</tr>
<tr>
<td>R1855</td>
<td>118.0</td>
<td>7.10</td>
<td>178.5</td>
<td>58.70</td>
<td>29.48</td>
</tr>
<tr>
<td>E1856</td>
<td>118.5</td>
<td>8.20</td>
<td>178.8</td>
<td>59.76</td>
<td>29.59</td>
</tr>
<tr>
<td>L1857</td>
<td>119.3</td>
<td>8.18</td>
<td>177.8</td>
<td>58.06</td>
<td>41.54</td>
</tr>
<tr>
<td>L1858</td>
<td>118.1</td>
<td>8.11</td>
<td>178.9</td>
<td>57.06</td>
<td>41.24</td>
</tr>
<tr>
<td>E1859</td>
<td>117.6</td>
<td>7.84</td>
<td>178.5</td>
<td>58.38</td>
<td>29.58</td>
</tr>
<tr>
<td>M1860</td>
<td>118.0</td>
<td>7.42</td>
<td>176.5</td>
<td>57.98</td>
<td>34.82</td>
</tr>
<tr>
<td>E1861</td>
<td>118.0</td>
<td>8.05</td>
<td>173.2</td>
<td>58.17</td>
<td>29.20</td>
</tr>
<tr>
<td>P1862</td>
<td>---</td>
<td>---</td>
<td>179.9</td>
<td>65.66</td>
<td>31.77</td>
</tr>
<tr>
<td>E1863</td>
<td>114.4</td>
<td>8.59</td>
<td>177.5</td>
<td>55.82</td>
<td>30.69</td>
</tr>
<tr>
<td>T1864</td>
<td>107.5</td>
<td>7.41</td>
<td>173.6</td>
<td>61.08</td>
<td>69.74</td>
</tr>
<tr>
<td>A1865</td>
<td>123.8</td>
<td>7.55</td>
<td>175.8</td>
<td>51.07</td>
<td>19.36</td>
</tr>
<tr>
<td>K1866</td>
<td>122.4</td>
<td>8.69</td>
<td>175.6</td>
<td>54.47</td>
<td>34.67</td>
</tr>
<tr>
<td>F1867</td>
<td>125.2</td>
<td>9.24</td>
<td>175.9</td>
<td>57.99</td>
<td>40.48</td>
</tr>
<tr>
<td>S1868</td>
<td>121.3</td>
<td>9.10</td>
<td>172.1</td>
<td>57.91</td>
<td>63.38</td>
</tr>
<tr>
<td>P1869</td>
<td>---</td>
<td>---</td>
<td>176.7</td>
<td>62.57</td>
<td>32.02</td>
</tr>
<tr>
<td>A1870</td>
<td>124.0</td>
<td>8.56</td>
<td>177.3</td>
<td>53.05</td>
<td>19.70</td>
</tr>
<tr>
<td>E1871</td>
<td>122.3</td>
<td>8.67</td>
<td>175.1</td>
<td>54.36</td>
<td>32.75</td>
</tr>
<tr>
<td>R1872</td>
<td>124.1</td>
<td>8.68</td>
<td>177.3</td>
<td>55.37</td>
<td>31.15</td>
</tr>
<tr>
<td>T1873</td>
<td>115.6</td>
<td>8.51</td>
<td>177.0</td>
<td>60.76</td>
<td>70.81</td>
</tr>
<tr>
<td>Y1874</td>
<td>---</td>
<td>---</td>
<td>175.7</td>
<td>59.87</td>
<td>37.50</td>
</tr>
<tr>
<td>D1875</td>
<td>116.2</td>
<td>7.94</td>
<td>176.3</td>
<td>52.33</td>
<td>38.85</td>
</tr>
<tr>
<td>G1876</td>
<td>107.3</td>
<td>7.82</td>
<td>174.8</td>
<td>46.23</td>
<td>---</td>
</tr>
<tr>
<td>K1877</td>
<td>119.7</td>
<td>7.64</td>
<td>174.5</td>
<td>55.14</td>
<td>33.50</td>
</tr>
<tr>
<td>V1878</td>
<td>119.0</td>
<td>8.61</td>
<td>173.5</td>
<td>61.48</td>
<td>33.64</td>
</tr>
<tr>
<td>R1879</td>
<td>125.2</td>
<td>9.09</td>
<td>175.0</td>
<td>53.52</td>
<td>33.88</td>
</tr>
<tr>
<td>V1880</td>
<td>124.5</td>
<td>8.84</td>
<td>173.5</td>
<td>61.13</td>
<td>36.05</td>
</tr>
<tr>
<td>T1881</td>
<td>121.3</td>
<td>8.56</td>
<td>173.6</td>
<td>60.95</td>
<td>70.88</td>
</tr>
<tr>
<td>V1882</td>
<td>124.9</td>
<td>9.41</td>
<td>172.4</td>
<td>58.08</td>
<td>35.23</td>
</tr>
<tr>
<td>E1883</td>
<td>129.6</td>
<td>8.63</td>
<td>175.3</td>
<td>54.03</td>
<td>31.84</td>
</tr>
<tr>
<td>V1884</td>
<td>126.9</td>
<td>8.90</td>
<td>176.1</td>
<td>60.48</td>
<td>31.87</td>
</tr>
<tr>
<td>V1885</td>
<td>129.0</td>
<td>8.72</td>
<td>177.0</td>
<td>65.88</td>
<td>31.70</td>
</tr>
<tr>
<td>G1886</td>
<td>115.5</td>
<td>8.89</td>
<td>174.7</td>
<td>45.30</td>
<td>---</td>
</tr>
<tr>
<td>K1887</td>
<td>119.0</td>
<td>8.63</td>
<td>176.0</td>
<td>56.73</td>
<td>35.14</td>
</tr>
<tr>
<td>G1888</td>
<td>106.7</td>
<td>8.02</td>
<td>170.5</td>
<td>45.21</td>
<td>---</td>
</tr>
<tr>
<td>K1889</td>
<td>117.4</td>
<td>7.47</td>
<td>175.3</td>
<td>54.89</td>
<td>35.32</td>
</tr>
<tr>
<td>F1890</td>
<td>121.4</td>
<td>8.94</td>
<td>174.0</td>
<td>56.59</td>
<td>41.96</td>
</tr>
<tr>
<td>K1891</td>
<td>123.0</td>
<td>8.81</td>
<td>175.5</td>
<td>55.59</td>
<td>35.67</td>
</tr>
<tr>
<td>G1892</td>
<td>109.2</td>
<td>8.96</td>
<td>171.2</td>
<td>44.31</td>
<td>---</td>
</tr>
<tr>
<td>V1893</td>
<td>119.7</td>
<td>8.10</td>
<td>175.6</td>
<td>59.59</td>
<td>35.58</td>
</tr>
<tr>
<td>G1894</td>
<td>108.6</td>
<td>8.80</td>
<td>173.3</td>
<td>45.15</td>
<td>---</td>
</tr>
<tr>
<td>R1895</td>
<td>118.7</td>
<td>7.98</td>
<td>176.7</td>
<td>62.17</td>
<td>32.76</td>
</tr>
<tr>
<td>S1896</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Y1897</td>
<td>---</td>
<td>---</td>
<td>175.9</td>
<td>61.32</td>
<td>38.02</td>
</tr>
<tr>
<td>Year</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>R1898</td>
<td>117.8</td>
<td>8.63</td>
<td>179.1</td>
<td>59.76</td>
<td>29.94</td>
</tr>
<tr>
<td>I1899</td>
<td>119.5</td>
<td>7.33</td>
<td>178.2</td>
<td>64.23</td>
<td>38.70</td>
</tr>
<tr>
<td>A1900</td>
<td>124.2</td>
<td>8.17</td>
<td>178.6</td>
<td>55.44</td>
<td>18.89</td>
</tr>
<tr>
<td>K1901</td>
<td>117.9</td>
<td>8.60</td>
<td>177.3</td>
<td>60.06</td>
<td>32.67</td>
</tr>
<tr>
<td>S1902</td>
<td>113.6</td>
<td>7.41</td>
<td>176.9</td>
<td>62.86</td>
<td>62.00</td>
</tr>
<tr>
<td>A1903</td>
<td>123.7</td>
<td>8.15</td>
<td>180.5</td>
<td>--</td>
<td>18.06</td>
</tr>
<tr>
<td>A1904</td>
<td>123.3</td>
<td>8.39</td>
<td>178.8</td>
<td>55.25</td>
<td>17.80</td>
</tr>
<tr>
<td>A1905</td>
<td>118.0</td>
<td>8.15</td>
<td>178.6</td>
<td>55.21</td>
<td>18.49</td>
</tr>
<tr>
<td>R1906</td>
<td>117.5</td>
<td>8.60</td>
<td>178.8</td>
<td>60.07</td>
<td>30.09</td>
</tr>
<tr>
<td>R1907</td>
<td>119.7</td>
<td>7.71</td>
<td>178.6</td>
<td>59.28</td>
<td>29.73</td>
</tr>
<tr>
<td>A1908</td>
<td>122.8</td>
<td>7.79</td>
<td>179.0</td>
<td>54.07</td>
<td>19.13</td>
</tr>
<tr>
<td>L1909</td>
<td>117.3</td>
<td>8.36</td>
<td>178.5</td>
<td>58.05</td>
<td>42.45</td>
</tr>
<tr>
<td>R1910</td>
<td>116.9</td>
<td>7.56</td>
<td>178.7</td>
<td>58.92</td>
<td>29.87</td>
</tr>
<tr>
<td>S1911</td>
<td>112.8</td>
<td>7.07</td>
<td>177.7</td>
<td>61.24</td>
<td>63.91</td>
</tr>
<tr>
<td>L1912</td>
<td>120.3</td>
<td>8.24</td>
<td>179.1</td>
<td>57.09</td>
<td>41.82</td>
</tr>
<tr>
<td>K1913</td>
<td>117.1</td>
<td>7.93</td>
<td>176.7</td>
<td>57.11</td>
<td>32.47</td>
</tr>
<tr>
<td>A1914</td>
<td>122.2</td>
<td>7.47</td>
<td>177.7</td>
<td>53.23</td>
<td>19.00</td>
</tr>
<tr>
<td>N1915</td>
<td>116.2</td>
<td>8.09</td>
<td>174.7</td>
<td>53.14</td>
<td>38.83</td>
</tr>
<tr>
<td>Q1916</td>
<td>120.8</td>
<td>7.96</td>
<td>173.8</td>
<td>53.68</td>
<td>28.91</td>
</tr>
<tr>
<td>P1917</td>
<td>---</td>
<td>---</td>
<td>176.7</td>
<td>62.99</td>
<td>32.08</td>
</tr>
<tr>
<td>Q1918</td>
<td>121.0</td>
<td>8.47</td>
<td>175.8</td>
<td>55.65</td>
<td>29.41</td>
</tr>
<tr>
<td>V1919</td>
<td>122.9</td>
<td>8.17</td>
<td>174.4</td>
<td>59.81</td>
<td>32.56</td>
</tr>
<tr>
<td>P1920</td>
<td>---</td>
<td>---</td>
<td>176.6</td>
<td>63.03</td>
<td>32.20</td>
</tr>
<tr>
<td>N1921</td>
<td>119.1</td>
<td>8.51</td>
<td>174.3</td>
<td>53.45</td>
<td>38.84</td>
</tr>
<tr>
<td>S1922</td>
<td>120.9</td>
<td>7.82</td>
<td>178.5</td>
<td>59.97</td>
<td>64.86</td>
</tr>
</tbody>
</table>
Supporting Information References:


Appendix C

Supporting Information: Chapter 6


*Figure C.6 to C.17 were made by Sushant Kumar and W. G. Noid.*

**Appendix C Summary**

Atomistic MD simulations reveal that binding FCP1 impacts the stability and structural organization of RAP74 and also provide new insight into the conformational flexibility of FCP1 in this complex. Appendix C Figures C.1 to C.4 provide extended analysis of the interaction of FCP1 residues Leu-953 and Leu-957 with their respective binding pockets in the FCP1 groove of RAP74. Included are analysis of residues Leu-960 and Met-961, which serve as controls to demonstrate the extent of solvent access and conformational transitions available to residues that do not make spatio-temporally stable contact with RAP74. The remainder of Appendix C section provides further quantitative characterization of the impact FCP1-RAP74 interactions have on the structure and dynamics of RAP74.

**C.1 Supporting Information Methods**

**C.1.1 Secondary Structure Definitions**

RAP74 subunit is a winged helix domain protein consisting of 4 helices – H1 (Thr-455 to Lys-466), H2 (Thr-469 to Lys-475); H2.5 (Gln-478 to Gly-483); and H3 (Ser-485 to Asn-501) – and 3 strands – S1 (Pro-467 to Met-468); S2 (Glu-503 to Ile-507); and S3 (Lys-510 to Ser-514). The N- and C- termini of H1, H2, H2.5, S2, and S3 were defined by the first and last
amino acids listed above, respectively. However, the N- and C-termini of H3 were defined by the midpoint of the alpha carbons for the first two (Ser-485 and Ser-486) and last two (Leu-500 and Asn-501) residues, respectively. These termini were used to calculate the end-to-end distance of each secondary structure and also the distances between different secondary structures.

C.1.2 Helix Dipole Calculation

A backbone dipole was defined for each residue by the backbone atoms - N, CA, H, HA and O. Because the AMBER force field defines a net charge for this atomic group, the backbone dipole moment was calculated after modifying the atomic charges by applying a uniform charge to neutralize the atomic group. We emphasize that these modified charges were only used for analysis purposes and not for the reported simulations. For each secondary structure element, a dipole vector was calculated by summing the dipoles for the corresponding residues.

C.1.3 Helix Orientation and Twist Vectors

The local axis director for each residue was calculated using the g_helixorient utility in the GROMACS software package. For each helix, the principal axis was defined by the vector sum of local axis directors for the corresponding residues. For each strand, the twist vector was similarly calculated as the analogous vector sum. In particular, the local twist vector for the ith residue was defined by the cross product of the displacement vector between residue i-1 and residue i+1 and the displacement vector between residue i and residue i+2.
Table C.1: Molecular Dynamics Determined Average Distances for Key Intermolecular Interactions

<table>
<thead>
<tr>
<th>RAP74 Residue and Atom</th>
<th>FCP1 Residue and Atom</th>
<th>Average Distance (Å)</th>
<th>Percent Less Than 4.0 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr470 C$_γ$</td>
<td>Leu953 C$_{δ1}$</td>
<td>5.83</td>
<td>0.9</td>
</tr>
<tr>
<td>Thr470 C$_γ$</td>
<td>Leu953 C$_{δ2}$</td>
<td>4.24</td>
<td>33.4</td>
</tr>
<tr>
<td>Thr470 O$_γ$</td>
<td>Glu954 O$_ε$</td>
<td>5.48</td>
<td>14.2</td>
</tr>
<tr>
<td>Thr470 O$_γ$</td>
<td>Glu954 O$_ε$</td>
<td>5.53</td>
<td>14.5</td>
</tr>
<tr>
<td>Lys471 N$_ζ$</td>
<td>Asp947 O$_δ$</td>
<td>4.77</td>
<td>32.9</td>
</tr>
<tr>
<td>Lys471 N$_ζ$</td>
<td>Asp947 O$_δ$</td>
<td>4.68</td>
<td>34.1</td>
</tr>
<tr>
<td>Leu474 C$_{δ1}$</td>
<td>Met949 C$_ε$</td>
<td>6.35</td>
<td>2.7</td>
</tr>
<tr>
<td>Leu474 C$_{δ2}$</td>
<td>Met949 C$_ε$</td>
<td>5.74</td>
<td>7.4</td>
</tr>
<tr>
<td>Leu474 C$_{δ1}$</td>
<td>Leu953 C$_{δ1}$</td>
<td>5.95</td>
<td>0.4</td>
</tr>
<tr>
<td>Leu474 C$_{δ1}$</td>
<td>Leu953 C$_{δ2}$</td>
<td>4.84</td>
<td>14.4</td>
</tr>
<tr>
<td>Leu474 C$_{δ2}$</td>
<td>Leu953 C$_{δ1}$</td>
<td>6.16</td>
<td>0.3</td>
</tr>
<tr>
<td>Leu474 C$_{δ2}$</td>
<td>Leu953 C$_{δ2}$</td>
<td>5.04</td>
<td>4.0</td>
</tr>
<tr>
<td>Ser486 C$_β$</td>
<td>Met949 C$_ε$</td>
<td>4.53</td>
<td>29.0</td>
</tr>
<tr>
<td>Val490 C$_γ$</td>
<td>Met949 C$_ε$</td>
<td>6.28</td>
<td>5.6</td>
</tr>
<tr>
<td>Val490 C$_γ$</td>
<td>Met949 C$_ε$</td>
<td>6.21</td>
<td>5.4</td>
</tr>
<tr>
<td>Asn491 N$_δ$</td>
<td>Glu956 O$_ε$</td>
<td>4.74</td>
<td>38.7</td>
</tr>
<tr>
<td>Asn491 N$_δ$</td>
<td>Glu956 O$_ε$</td>
<td>4.73</td>
<td>38.7</td>
</tr>
<tr>
<td>Leu493 C$_{δ1}$</td>
<td>Leu953 C$_{δ1}$</td>
<td>5.67</td>
<td>0.2</td>
</tr>
<tr>
<td>Leu493 C$_{δ1}$</td>
<td>Leu953 C$_{δ2}$</td>
<td>5.71</td>
<td>2.0</td>
</tr>
<tr>
<td>Leu493 C$_{δ2}$</td>
<td>Leu953 C$_{δ1}$</td>
<td>4.59</td>
<td>16.3</td>
</tr>
<tr>
<td>Leu493 C$_{δ2}$</td>
<td>Leu953 C$_{δ2}$</td>
<td>4.28</td>
<td>44.5</td>
</tr>
<tr>
<td>Leu497 C$_{δ1}$</td>
<td>Leu957 C$_{δ1}$</td>
<td>4.72</td>
<td>16.4</td>
</tr>
<tr>
<td>Leu497 C$_{δ1}$</td>
<td>Leu957 C$_{δ2}$</td>
<td>4.75</td>
<td>13.2</td>
</tr>
<tr>
<td>Leu497 C$_{δ2}$</td>
<td>Leu957 C$_{δ1}$</td>
<td>5.73</td>
<td>3.2</td>
</tr>
<tr>
<td>Leu497 C$_{δ2}$</td>
<td>Leu957 C$_{δ2}$</td>
<td>5.77</td>
<td>4.3</td>
</tr>
<tr>
<td>Lys498 N$_ζ$</td>
<td>Asp959 O$_δ$</td>
<td>6.63</td>
<td>14.7</td>
</tr>
<tr>
<td>Lys498 N$_ζ$</td>
<td>Asp959 O$_δ$</td>
<td>6.53</td>
<td>17.4</td>
</tr>
<tr>
<td>Phe513 C$_δ$</td>
<td>Leu957 C$_{δ1}$</td>
<td>5.07</td>
<td>20.6</td>
</tr>
<tr>
<td>Phe513 C$_δ$</td>
<td>Leu957 C$_{δ2}$</td>
<td>4.58</td>
<td>36.0</td>
</tr>
</tbody>
</table>
C.2 Interactions Stabilizing the RAP74-FCP1 Complex

In the main text we describe the extensive dynamics of Met-949 in the shallow FCP1 binding groove on the surface of RAP74. Here we provide additional figures documenting the relative structural disorder of Leu-953 and Leu-957, which are the other two hydrophobic FCP1 residues defining the non-polar RAP74-FCP1 interface. Additional analysis is provided of Leu-960 and Met-961, which are solvent exposed, as a comparison that establishes a frame of reference for the extent of conformational restriction and exclusion from solvent experienced by the three resided providing critical interactions.
Figure C.1: Dynamic motion of Leu-953. The sidechain of Leu-953 is held tightly in the RAP74 binding groove. (A) Representative snapshots taken at 50 ns, 100 ns, 150 ns, and 200 ns zoomed in to show Leu-953 (colored by atom) from FCP1 (green ribbon) bound to the RAP74 binding groove (red van der Waals spheres). (B) Solvent accessible surface area of Leu-953 as a function of time. The remaining panels display Leu953 torsion angles $\phi$ (C), $\psi$ (D), $\chi_1$ (E), and $\chi_2$ (F) as a function of time.
Figure C.2: Dynamic motion of Leu-957. The sidechain of Leu-957 undergoes significant conformational dynamics, but does not escape from the RAP74 binding groove during the course of the simulation. (A) Representative snapshots taken at 50 ns, 100 ns, 150 ns, and 200 ns zoomed in to show Leu-957 (colored by atom) from FCP1 (green ribbon) bound to the RAP74 binding groove (red van der Waals spheres). (B) Solvent accessible surface area of Leu-957 as a function of time. The remaining panels display Leu957 torsion angles $\phi$ (C), $\psi$ (D), $\chi_1$ (E), and $\chi_2$ (F) as a function of time. Backbone torsion angles $\phi$ and $\psi$ show the melting out of helical structure in the final 50 ns of the simulation, although this transition does not result in release of the Leu-957 sidechain from the binding groove.
Figure C.3: Dynamic motion of Leu-960. The sidechain of Leu-960 is representative of a Leu sidechain that does not make strong interactions with RAP74. (A) Representative snapshots taken at 50 ns, 100 ns, 150 ns, and 200 ns zoomed in to show Leu-960 (colored by atom) from FCP1 (green ribbon) and RAP74 (red van der Waals spheres). (B) Solvent accessible surface area of Leu-960 as a function of time. The remaining panels display Leu-960 torsion angles $\phi$ (C), $\psi$ (D), $\chi_1$ (E), and $\chi_2$ (F) as a function of time.
**Figure C.4:** Dynamic motion of Met-961. The sidechain of Met-961 is representative of a Met sidechain that does not make strong interactions with RAP74. (A) Representative snapshots taken at 50 ns, 100 ns, 150 ns, and 200 ns zoomed in to show Met-961 (colored by atom) from FCP1 (green ribbon) and RAP74 (red van der Waals spheres). (B) Solvent accessible surface area of Met-961 as a function of time. The remaining panels display Met961 torsion angles $\phi$ (C), $\psi$ (D), $\chi_1$ (E), and $\chi_2$ (F) as a function of time.
C.3 FCP1 alters RAP74 structure and dynamics

Atomistic MD simulations reveal that binding FCP1 impacts the stability and structural organization of RAP74 and also provide new insight into the conformational flexibility of FCP1 in this complex. Appendix C provides further quantitative characterization of the impact FCP1-RAP74 interactions have on the structure and dynamics of RAP74. Figure C.5 indicates that RAP74 accommodates FCP1 by exposing key residues in the hydrophobic groove, while also withdrawing residues on the exterior of this pocket. Figure C.6 demonstrates that the FCP1 helix dipole tends to partially align with the H2 helix and anti-align with H3. Figure C.7 demonstrates that this interaction does not significantly alter the structure or orientation of H2. Figure C.8 demonstrates that, upon binding FCP1, the hydrophobic groove of RAP74 expands and become more stable. Figures C.9 and C.10 suggest that, after binding FCP1, H2.5 helix approaches H3 at a right angle while aligning more closely with H2. Figure C.11 demonstrates that, after binding FCP1, H1 becomes more compact and aligns at an increasingly anti-parallel orientation with respect to H2. Figure C.12 indicates that FCP1 stabilizes the alignment of strands S2 and S3. Figures C.13 and C.14 demonstrate that the RAP74 binding pocket stabilizes the helical axis of FCP1, especially near the N-terminal of FCP1, while the C-terminal samples larger and more correlated fluctuations along its axis. Figure C.15 suggests that hydrophobic interactions involving Ala946 and Ala950 stabilize the N-terminal turn of FCP1 until the last 10ns. In contrast, Figure C.16 demonstrates that contacts involving Met-949 – Leu-953 and Leu-953 – Leu-957 remain very stable and anchor the C-terminal turn of FCP1. Figure C.17 demonstrates that the FCP1 dipole samples significant fluctuations and rapidly decays during the last 50ns of the simulation.
Figure C.5: Changes in Solvent Accessible Surface Area (SASA) of RAP74 upon binding FCP1. The black curve identifies the regions of RAP74 that are in direct contact with FCP1 by presenting the number of intermolecular contacts formed by each residue in the crystal structure. (The data in the black curve has been multiplied by a factor of four so that both curves fit on the same scale.) The red curve presents the difference in per-residue SASA between holo and apoRAP74 and shows that RAP74 residues directly contacting FCP1 become more surface exposed (to either solvent or FCP1) in the complex. Reorganization of RAP74 also results in a significant number of residues becoming increasingly buried (negative difference), meaning that they are less accessible to both solvent and FCP1; these residues are almost never in direct contact with FCP1. Of particular note is H1, which tilts to become increasingly antiparallel with H2, resulting in an oscillatory set of changes in per-residue surface exposure.
Figure C.6: Alignment of FCP1 in the RAP74 binding pocket.  

(a) Time traces of the cosine of the angles formed by the backbone dipole vectors of the FCP1 helix with the H2 helix (black curve) and the H3 helix (red curve).  

(b) Probability distributions for the corresponding time traces. FCP1 tends to be partially aligned with the H2 helix and anti-aligned with the H3 helix.
Figure C.7: Characterization of RAP74 H2 in the apo and holo state. Column 1 presents time traces calculated from simulations of apo (black) and holo (red) RAP74 for a1) the H2 end-to-end distance; b1) the H2 backbone dipole moment; c1) the angle formed between the principal axes of H2 and H3. For each observable, column 2 presents corresponding distributions calculated from the time traces. The structure, fluctuations, and orientation of H2 remain relatively unchanged by the binding of FCP1.
Figure C.8: Packing of the RAP74 hydrophobic groove in the apo and holo state. Column 1 presents time traces calculated from simulations of apo (black) and holo (red) RAP74 for distances between (a1) the N-termini of H2 (469) and H3 (485-486); (b1) the C-terminal of H2 (475) and the N-terminal of H3 (485-486); (c1) the C-termini of H2 (475) and H3 (500-501); (d1) the N-terminal of H2 (469) and the C-terminal of H3 (500-501). Column 2 presents distribution functions calculated from the corresponding time traces. In each case, distances are calculated between the centers of mass defined by the alpha carbons of the indicated residues. The binding pocket appears to slightly open and become more stable upon binding FCP1.
Figure C.9: Changes in the location and fluctuations of RAP74 H2.5 in the apo and holo state. Column 1 presents time traces calculated from simulations of apo (black) and holo (red) RAP74 for distances between a) the C-terminal of H2 (475) helix and the N-terminal of H2.5 (478); and b) the C-terminal of H2.5 (483) and the N-terminal of H3 (485-486). Column 2 presents distribution functions calculated from the corresponding time traces. In each case, distances are calculated between the centers of mass defined by the alpha carbons of the indicated residues. Prior to binding FCP1, H2.5 samples conformations that are more distant from H3.
**Figure C.10**: Orientation of RAP74 secondary structures in the apo and holo state. Column 1 presents time traces calculated from simulations of apo (black) and holo (red) RAP74 for the cosine of the angle formed by the principal axis vectors of 

- **a1)** H2 and H2.5 helices; 
- **b1)** H2 and H3 helices; 
- **c1)** H2.5 and H3 helices. Column 2 presents distribution functions calculated from the corresponding time traces. Upon binding FCP1, H2.5 demonstrates increased tendency to align parallel with H2 and perpendicular with H3. In contrast, the angle formed by the axes of H2 and H3 appears unchanged.
Figure C.11: Changes in H1 structure and orientation resulting from binding FCP1. Column 1 presents time traces calculated from simulations of apo (black) and holo (red) RAP74 for a1) the H1 end-to-end distance (measured between the alpha carbons of residues 455 and 466); and b1) the cosine of the angle formed between the principal axes of H1 and H2. Column 2 presents distribution functions calculated from the corresponding time traces. Upon binding FCP1, H1 tends to adopt more compact conformation with a more antiparallel orientation with respect to the H2.
Figure C.12: Orientation of secondary structures in RAP74 before and after binding FCP1. Column 1 presents time traces calculated from simulations of apo (black) and holo (red) RAP74 for the cosine of the angles formed by a1) the principal axes of H2 and S2; b1) the twist vectors of S1 and S2; c1) twist vectors of S1 and S3. Column 2 presents distribution functions calculated from the corresponding time traces. FCP1 stabilizes the alignment of S2 with respect to H2 and also dramatically rigidifies the interaction between S1 and S3.
Figure C.13: Angle of successive helix directors along the N-terminal region of FCP1. Column 1 presents time traces calculated from simulations of the RAP74-FCP1 complex for the cosine of the angle formed by the local helix directors at the i and i+4 alpha carbons of FCP1 including (a1) Ser-944 and Glu-948; (b1) Glu-945 and Met-949; (c1) Ala-946 and Ala-950; (d1) Asp-947 and Lys-951; and (e1) Glu-948 and Ala-952. Column 2 presents distribution functions calculated from the corresponding time traces. The RAP74 binding pocket maintains the helical axis along the N-terminal of FCP1.
Figure C.14: Angle of successive helix directors along the C-terminal region of FCP1. Column 1 presents time traces calculated from simulations of the RAP74-FCP1 complex for the cosine of the angle formed by the local helix directors at the i and i+4 alpha carbons of FCP1 including a1) Met-949 and Leu-953; b1) Ala-950 and Glu-954; c1) Lys-951 and Ala-955; d1) Ala-952 and Glu-956; and e1) Leu-953 and Leu-957. Column 2 presents distribution functions calculated from the corresponding time traces. The RAP74 binding pocket maintains the helical axis of FCP1, although the C-terminal region samples somewhat larger and more correlated fluctuations than the N-terminal region.
Figure C.15: Distance between successive turns along the N-terminal of FCP1. Column 1 presents time traces calculated from simulations of the RAP74-FCP1 complex for the distances between i and i+4 alpha carbons of FCP1 including a1) Ser-944 and Glu-948; b1) Glu-945 and Met-949; c1) Ala-946 and Ala-950; d1) Asp-947 and Lys-951; and e1) Glu-948 and Ala-952. Column 2 presents distribution functions calculated from the corresponding time traces. Although the flanking Glu-945 – Met-949, Asp-947 – Lys-951, and Glu-948 – Ala-952 pairs sample correlated fluctuations to larger distances, the Ala-946 – Ala-950 pair samples much smaller fluctuations and appears to stabilize the N-terminal turn of FCP1 until the last ten nanoseconds of the simulation.
Figure C.16: Distance between successive turns along the C-terminal of FCP1. Column 1 presents time traces calculated from simulations of the RAP74-FCP1 complex for the distances between i and i+4 alpha carbons of FCP1 including a1) Met-949 and Leu-953; b1) Ala-950 and Glu-954; c1) Lys-951 and Ala-955; d1) Ala-952 and Glu-956; and e1) Leu-953 and Leu-957. Column 2 presents distribution functions calculated from the corresponding time traces. The Met-949 – Leu-953 and Leu-953 – Leu-957 contacts remain very stable during the simulation and anchor the C-terminal turn of FCP1.
Figure C.17: Dipole moment for the FCP1 backbone. Panel a presents the time trace calculated from simulations of the RAP74-FCP1 complex for the magnitude of the FCP1 backbone dipole. Panel b presents the corresponding distribution function. The FCP1 backbone dipole samples significant fluctuations and precipitously drops during the last 50 ns of the simulation.
VITA

Christopher William Wostenberg

EDUCATION

Ph. D., Chemistry: The Pennsylvania State University, University Park, PA (08/2012)
B.S., Biochemistry with honors: California State University, Long Beach, CA (12/2006)
B.S., Mathematics with honors: California State University, Long Beach, CA (12/2006)
B.A., Chemistry: California State University, Long Beach, CA (12/2006)

PUBLICATIONS


PRESENTATIONS