The Pennsylvania State University

The Graduate School

Eberly College of Science

TOWARD A MORE COMPLETE UNDERSTANDING OF MICROPROCESSING

A Thesis in

Chemistry

by

Joshua C. Kranick

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

May 2016

The thesis of Joshua C. Kranick was reviewed and approved* by the following:

Scott A. Showalter

Associate Professor of Chemistry

Thesis Advisor

Edward P. O'Brien

Assistant Professor of Chemistry

Christine D. Keating

Professor of Chemistry

Kenneth S. Feldman

Professor of Chemistry

Graduate Program Chair

*Signatures on file with the Graduate School.

Abstract

Gene silencing via micro-RNAs is a cell's main mechanism in maintaining cellular homeostasis. This is directly evidenced by the immense number of disease states directly correlated to the improper function or maturation of a micro-RNA. Initiation of the canonical micro-RNA maturation pathway begins via substrate selection and processing of a primary micro-RNA by the catalytic complex, Microprocessor, in the nucleus. The catalytic unit of this complex, Drosha, has no means to recognize substrate as its double stranded RNA binding domain has been empirically shown to not bind double-stranded RNA. Further analysis of this domain via human cell-based processing assays revealed that, while this domain is not utilized via its canonical nature, is imperative to Microprocessor function- deletion of the domain leads to complete ablation of primary-micro-RNA processing. In this thesis, I describe my work toward understanding the role of this domain utilizing classic biochemical and biophysical techniques.

Table of Contents

List of Figures

List of Abbreviations

a	Hill fit curve amplitude correction factor
Arg	Drosha arginine/ serine rich region
ATP	Adenosine triphosphate
b	Hill fit curve amplitude correction factor
bp	Base pair
CA	Alpha carbon
CB	Beta carbon
CED	Drosha central domain
CO	Carbonyl carbon
DEDX	DEAD box helicase domain
DGCR8	Digeorge Syndrome Critical Region 8
DNA	Deoxy ribonucleic acid
Drosha+	Drosha dsRBD and native C- terminus
ds44	44 base pair perfect Watson-Crick duplex
dsRBD	Double stranded RNA binding domain
dsRNA	Double stranded RNA
EMSA	Electrophoretic mobility shift assay
G2	DGCR8 C- terminal tail
HEK293T	Human embryonic kidney cell
HeliC	C- terminal helicase domain
hetNOE	Heteronuclear Nuclear Overhouser Effect
HSQC	Heteronuclear single-quantum coherence spectroscopy
I1317K	Drosha mutant replacing isoleucine with lysine at amino acid 1317
IDR	Intrinsically disordered region
INEPT	Insensitive nucleus enhanced by polarization transfer
IPTG	Isopropyl β-D-1- thiogalactopyranoside
K_d	Apparent dissociation constant
miRNA	micro-RNA

mRNA	Messenger RNA
n	Hill coefficient
NOE	Nuclear Overhouser Effect
nt	Nucleotide
OD_{600}	Optical density at 600 nm
PACT	PKR activating protein
PBD	Dicer protein binding domain
pre-miRNA	Precursor micro-RNA
pri-miRNA	Primary micro-RNA
Pro	Drosha proline rich region
R1	Longitudinal relaxation rate
R2	Transverse relaxation rate
Rhed	DGCR8 RNA- binding heme domain
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
S^2	Order parameter
ss-ds	Single strand- double strand junction
T1	Longitudinal relaxation time
T2	Transverse relaxation time
TOCSY	Total correlation spectroscopy
TRBP	HIV-I TAR RNA binding protein
WT	Wild type
φ	N_{i-1}, C_i, CA_i, N_i torsion angle
Ψ	CO_i, CA_i, N_i, CO_{i+1} torsion angle
τ_e	Effective correlation time for internal motion

Acknowledgements

For Kate, my guiding light.

Chapter 1: Introduction

I. General Introduction

miRNA Maturation and Function

Micro-RNAs (miRNAs) are a class of single-stranded, twenty to twenty-two nucleotide (nt) non-coding RNAs that function in RNA interference (RNAi). RNAi is one of the primary means through which eukaryotes post-transcriptionally regulate gene expression. In this cellular process, the mature miRNA binds its complementary messenger RNA (mRNA) in the RNA-Induced Silencing Complex (RISC) in which expression of that gene is downregulated via destabilization of the bound mRNA[1]. miRNA dysfunction has been implicated in a number of disease states, including a number of different cancers[2], several organ based diseases (i.e. heart disease and kidney disease)[3], mental disorders (including schizophrenia)[4], and innate immunity to viral infection[5]. It is of utmost importance to study the maturation of these miRNAs so we can better understand how these disease states arise.

miRNA maturation begins in the nucleus, where the Microprocessor complex cleaves a dsRNA hairpin loop out of a primary RNA transcript (pri-miRNA, Figure 1). The Microprocessor complex consists of one copy of its catalytic unit, the RNase III enzyme Drosha, and two copies its necessary cofactor, the dsRNA binding protein Digeorge Syndrome Critical Region 8 (DGCR8). In this complex, it is thought that Drosha recognizes the single strand-double strand (ss-ds) junction of the pri-miRNA and the DGCR8 dimer is hypothesized to clamp the apical stem-loop[6, 7]. Microprocessing yields

a ~60 nt precursor miRNA (pre-miRNA) that possesses a two-nt 3′ overhang necessary for shuttling to the cytosol via the Exportin-5 pathway[8].
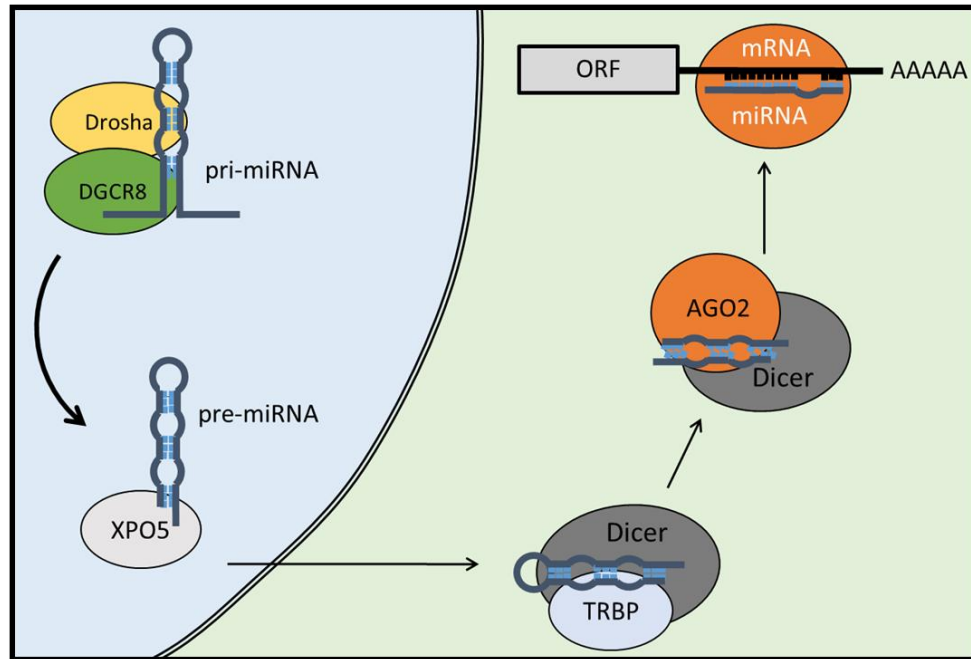
Figure 1: Cartoon diagram outlining the canonical miRNA maturation pathway. The nucleus of a eukaryotic cell is depicted in blue and the cytosol in grey. Though a number of different paths to mature miRNA exist, the maturation pathway represented in this figure is widely regarded as the canonical, direct pathway.

In the cytosol, the pre-miRNA is further processed by the RNase III enzyme Dicer, by which the stem-loop is cleaved to generate the miRNA-miRNA* duplex- the miRNA strand will become guide strand in RISC whereas the miRNA* strand will be discarded and degraded. Dicer has been shown to catalyze this processing event in the absence of its RNA-binding protein cofactors, which can be the HIV-I TAR RNA binding protein (TRBP) or the PKR activating protein (PACT), but its catalytic efficiency is significantly improved in the context of a Dicer-TRBP or Dicer-PACT complex. Specifically, the presence of TRBP or PACT in the dicing complex has been shown to affect what miRNA-

miRNA* duplex is generated by facilitating precise cut-site definition[9]. Following pre-miRNA cleavage, the pre-RISC complex binds an Argonaute protein which selects the guide strand of the miRNA-miRNA* duplex; the non-selected strand (miRNA*) is subsequently ejected and degraded[10, 11]. The sequence of the guide strand defines the mRNAs that will be targeted by RISC for destabilization and suppression of translation[12].
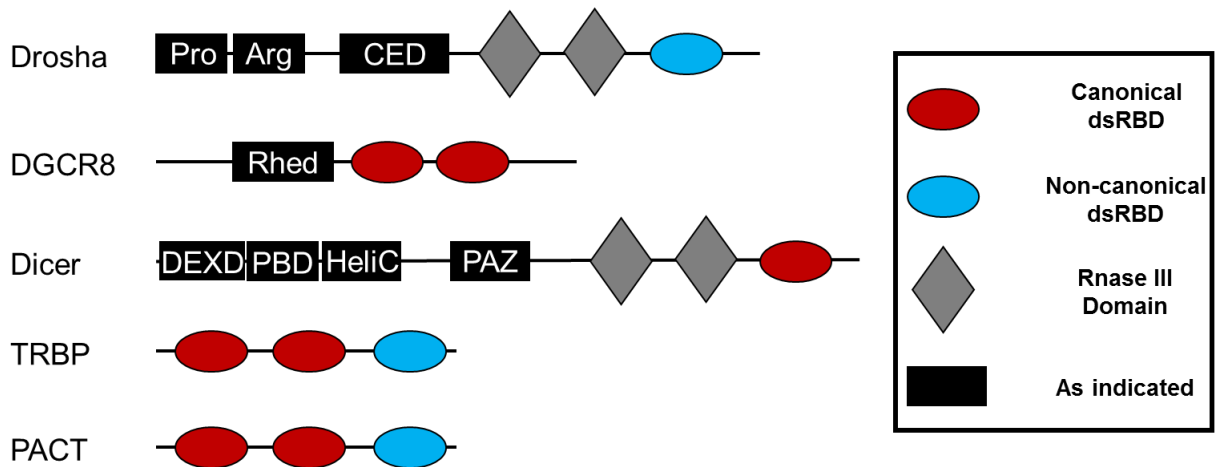
Figure 2: Cartoon diagram depicting the domain architecture of the major miRNA maturation proteins. Canonically functioning dsRBDs, non-canonically functioning dsRBDs and RNase III domains are represented as red ovals, blue ovals and grey diamonds, respectively. Domains integral to one maturation pathway component are as indicated in black boxes. They are abbreviated as Pro- proline rich region; Arg- arginine/ serine rich region; CED- central domain; Rhed- RNA- binding heme domain; DEXD- Deadbox helicase domain; PBD- protein binding domain; HeliC- C- terminal helicase domain; and the PAZ domain.

As with many eukaryotic proteins, the domain architecture of the proteins constituting the miRNA maturation pathway is complex (Figure 2). The RNase III enzymes Drosha and Dicer both contain a pair of RNase III domains, which comprise the catalytic centers of their respective processing complexes, and a C- terminal double stranded RNA binding domain (dsRBD). The N- terminal RNase III domain in each of these enzymes

cleaves the 3′ strand of the premature miRNA and the C- terminal RNase III domain cleaves the 5′ end[13]. The nuclear enzyme Drosha includes N- terminal Arginine/ Serine (Arg) and Proline (Pro) rich regions which have been shown to function in protein: protein and protein: RNA interactions in mouse Drosha[14], and a nuclear localization signal. Drosha also has a central domain (CED) which has no structure or function annotated to date, but is necessary for effective Microprocessor reconstitution *in vitro*[6].

Although Drosha and Dicer present the same C- terminal architecture of two RNase III domains followed by a dsRBD, the domain compositions of their N- terminal regulatory regions differ significantly. Dicer's N- terminal region contains a pair of Helicase domains responsible for unwinding the pre-miRNA and miRNA-miRNA* duplexes (DEDX and HeliC) and the TRBP/PACT protein binding domain (PBD). Between this region and the tandem RNase III domains is the PAZ domain – a molecular ruler that recognizes the two nt 3′ overhang of the pre-miRNA and positions the RNase III domains for effective processing[15]. Dicer's C- terminal dsRBD has been shown to enable effective Dicing *in vitro* in constructs that do not contain the PAZ domain[16]. Conversely, as introduced above, Drosha's dsRBD has no identified function to date.

The RNase III enzymes of bacterial organisms are usually comprised of a single catalytic domain and a single dsRBD at their catalytic regions. This one-to-one domain ratio suggests that the eukaryotic enzymes discussed above are depleted in dsRBDs compared to their prokaryotic progenitors and have evolved to split their function over multiple genes - obligating complex formation in their function[17]. As such, dsRBDs are the main functional domains of all three dsRNA binding cofactor proteins, DGCR8, TRBP and PACT; these domains have been the focus of intense study in the Showalter

4

laboratory. TRBP and PACT are comprised of three dsRBDs in tandem that are linked by short (~60 amino acid) intrinsically disordered regions (IDRs). Both N- terminal dsRBDs of TRBP and PACT function canonically and bind dsRNA without nucleotide-sequence specificity, whereas the C-terminal dsRBDs of these proteins have been shown empirically not to bind dsRNA *in vitro*[18]. Rather, these domains mediate the protein: protein interaction with Dicer via the enzyme's PBD[19]. Similarly, Drosha's dsRBD has been shown to not bind dsRNA *in vitro*[20]. The main focus of this thesis is to understand the function and role that this non-canonical dsRBD plays in Microprocessing.

Double-stranded RNA Binding Domains

Every major protein in the miRNA maturation pathway contains at least a single dsRBD. These domains are small protein domains (65-70 amino acids) that all adopt the canonical $\alpha-\beta-\beta-\beta-\alpha$ dsRBD fold[21] (Figure 3). While the general juxtaposition of the secondary structural elements is consistent throughout the domain family, a few deviations in structure have been evidenced - including an extended $\alpha 1$-$\beta 1$ loop in the dsRBD of Drosha and the presence of a third alpha helix C-terminal to the dsRBD of the yeast RNase III enzyme (Rnt1p) that folds back on and interacts with the dsRBD[22-25]. The $\alpha-\beta-\beta-\beta-\alpha$ architecture positions the dsRNA binding face – helix-1, the base of helix-2, and the $\beta 1$-$\beta 2$ loop – for effective substrate binding. It is important to note that while the general three-dimensional fold of this domain is strictly conserved, sequence is not, indicating that functional variability in this domain family is a possibility[26].
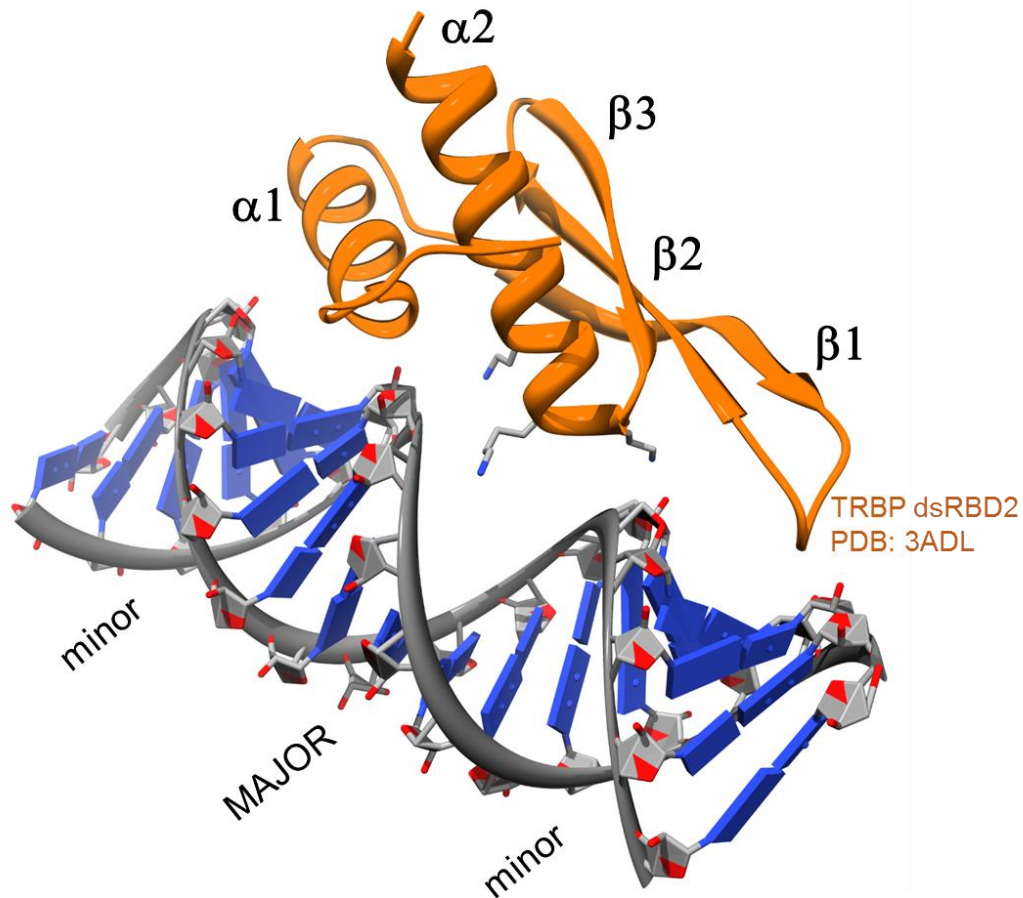
Figure 3: Ribbon diagram depiction of TRBP's dsRBD2 binding perfect Watson-Crick duplex RNA. This figure was generated using the Chimera PDB visualization software and PDB file 3ADL. The original crystal structure shows the dsRBD's KR- helix motif in contact with the junction of two coaxially stacked duplexes. This pair of duplexes was replaced by a single, perfect Watson-Crick throughout the dsRBD binding face by S.A.S.

The function of a canonical dsRBD is as the domain family name suggests – to bind A- form dsRNA. To perform this function, most dsRBDs contact the dsRNA substrate at three locations. The first of these interaction regions is with the minor groove of the A- form dsRNA duplex via the hydrophilic residues constituting the solvent exposed face of the amphipathic helix-1 of the dsRBD. This contact is facilitated by hydrogen bonding

between the hydrophilic residues of helix-1 and the ribose sugars presented by the duplex in the minor groove. The second minor groove-based binding region is facilitated by insertion of the β1-β2 loop into the minor groove. In one of the few nucleotide specific contacts common to these complexes, this interaction often features hydrogen bonding of a histidine sidechain in this loop with the carbonyl oxygen of a guanine within the minor groove[25].
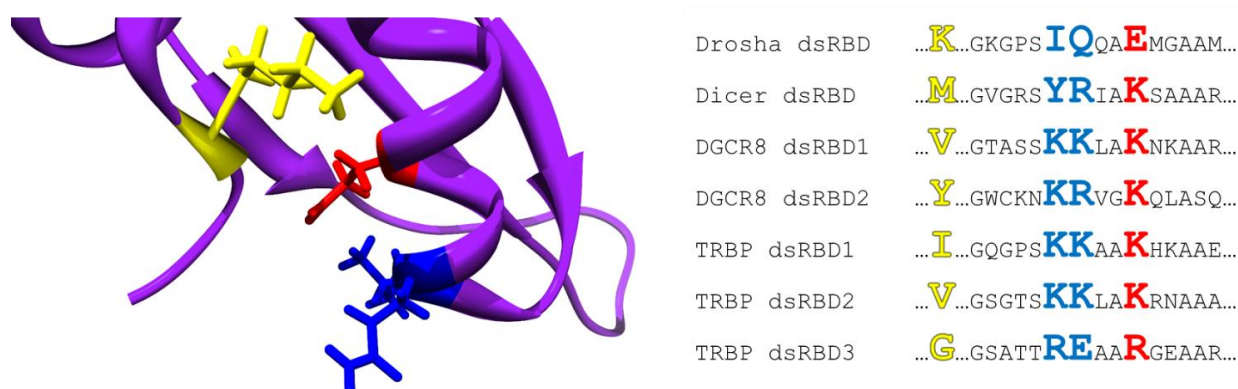
Figure 4: Ribbon diagram of KR helix motif of Drosha's dsRBD (PDB: 2KHX) with residues of interest highlighted and annotated in the adjacent sequence alignment with the corresponding color. The sequence alignment shows the KR helix motif region and an important amino acid position for through-space interaction with this region on helix-1.

While the minor groove interactions described above are important, strong thermodynamic evidence suggests that the most critical dsRBD-dsRNA interactions are mediated by the N- terminal end of helix-2, which projects basic amino acid sidechains into the major groove of the dsRNA duplex[25] (Figure 4). This interaction is driven by the chemistry of the KR helix motif at the base of helix-2, which is comprised of a KKxAK sequence element[21, 25]. The terminal amino groups of the lysine sidechains form charge-charge interactions with the non-bridging oxygens of the phosphate backbone. In some cases, the lysine positions of this motif may substitute arginine[25, 27]. As all of these

interaction regions take place where the sequence of the RNA cannot be read out, due to the fact that the minor groove presents the ribose moieties and underside of the nucleobases and the deep and narrow nature of the major groove sterically inhibits interaction with nucleotide faces, dsRBD: dsRNA interactions are non-sequence specific.

Canonically functioning dsRBDs bind to a single turn of A- form dsRNA duplex, utilizing the aforementioned binding mode, with low micromolar apparent dissociation constants (K_d)[20, 28-30]. Conveniently, these binding affinities establish the dsRBD concentration at which half of the potential saturating dsRNA lattice sites are occupied[30, 31]. dsRBDs exhibit a dsRNA lattice length dependence where binding becomes tighter as dsRNA lattice length increases. This phenomenon is amplified in multiple dsRBD containing systems, indicating that it may be physiologically significant whether these domains exist in isolation or in tandem[30, 31]. The notion that individual, canonical dsRBD function is not excessively variable makes study of non-dsRNA binding dsRBDs especially exciting, namely dsRBD3 of TRBP and Drosha's dsRBD.

Further inspection of the KR helix motif sequence of these domains illustrates why they are non-dsRNA binding. Both of these domains deviate from canonical KR helix motif sequence at the positions which drive dsRBD: dsRNA binding, resulting in a lack of positive charge density at the canonical dsRNA binding face. This lack of positive charge density results in dsRBD3 of TRBP and Drosha's dsRBD being unable to bind dsRNA *in vitro*[18, 20]. A recent biophysical study from the Doudna laboratory characterized the non-canonical function of dsRBD3 of TRBP in mediating protein: protein interaction with the RNase III enzyme Dicer. Such a definitive role of Drosha's dsRBD has yet to be realized[19]. Interestingly, the interface of the Dicer: TRBP interaction utilizes the β-sheet face of

dsRBD3 of TRBP and not the non-canonical KR- helix motif. This implies that the protein: protein interaction mediating function of the domain is not reliant on KR- helix motif chemistry, save for the inability of the domain to bind dsRNA. It is the position of this thesis that Drosha's dsRBD functions in a different, non-canonical manner in that the domain utilizes its unique KR- helix motif chemistry in an obligate role during Microprocessing.

The Microprocessor Complex

It is of utmost importance to understand the functional details of the Microprocessor complex as the substrate selection of this complex ultimately determines which primary transcript sequences become miRNAs. This interest is further compounded by the fact that Drosha, the catalytic unit of the Microprocessor complex, is the only obligate factor in the canonical miRNA maturation pathway that does not contain a dsRNA binding- competent dsRBD. This suggests that Drosha is unable to recognize pri-miRNA substrates, save for what substrate specificity may be imparted by its RNase III domains. Recognizing this, early works studying the Microprocessor complex tested hypotheses in which DGCR8, which is obliged to select the target pri-miRNA substrate, subsequently recruits Drosha for substrate processing[32].

In agreement with these early hypotheses, multiple groups have shown that neither recombinant DGCR8 nor Drosha reconstituted Microprocessing *in vitro,* but effective Microprocessing was recovered when both proteins were present[32, 33]. DGCR8's role was established to hinge on the function of its dsRNA binding 'core' fold - in which the tandem

9

dsRBDs of the cofactor form a pseudo-dimeric fold, sandwiching a short C- terminal helix between the β-sheet faces of each dsRBD[34]. This domain architecture positions the dsRNA binding faces of each dsRBD opposing one another, indicating that either the fold must rearrange for both domains to engage dsRNA or the pri-miRNA must have enough flexible elements to conform to DGCR8 core[29, 35].

This phenomenon, as well as reported *in vivo* and *in vitro* Microprocessing assays, suggest that pri-miRNA structural elements may contribute to Microprocessing[7, 35-37]. One region of interest is the basal ss-ds junction, which is hypothesized to be sensed by Drosha via an unknown mechanism[36, 37]. The deformability of the Drosha cut site and a secondary site one turn away are also hypothesized to play integral roles in the processing efficiency of the pri-miRNA[35]. Finally, a DGCR8 dimer is hypothesized to clamp the apical loop of the pri-miRNA via their RNA-binding heme domains[7]. Since dsRBDs do not bind with sequence specificity, it may make some sense that certain structural elements of a pri-miRNA substrate might play a role in pri-miRNA selection.

This Microprocessing model of DGCR8-driven substrate selection and subsequent recruitment of Drosha persisted for a decade until recent results from the Kim Lab challenged many of these hypotheses[6]. In this seminal paper, the authors conclude that the only protein regions needed to reconstitute Microprocessing *in vitro* are the central domain through the native C- terminus of Drosha and the C- terminal seventy-three residues of DGCR8[6]. Strikingly, these regions include no domains that have been demonstrated to recognize the pri-miRNA substrate and suggest Microprocessor assembly is substrate independent. Drosha's non-dsRNA binding dsRBD offers an interesting route of study in that the reason it cannot bind dsRNA is quite clear (Figure 4).

In this work, we assert that Drosha's dsRBD is imperative to proper *in vivo* Microprocessor function. This hypothesis will be established from both a biophysical and functional standpoint in an effort to elucidate the function of this domain (Figure 5). In this work, biophysical analyses will test characteristics of Drosha's dsRBD in isolation and full-length protein based Microprocessing assays will probe the domains function in this context. The combination will provide a deep and powerful understanding of this domains role in Microprocessing.
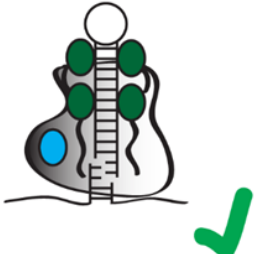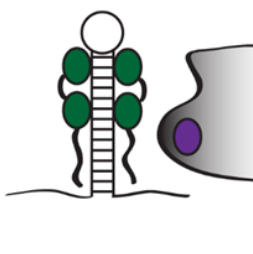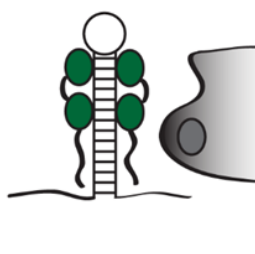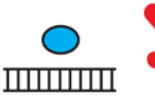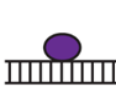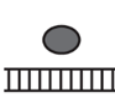
Figure 5: Graphical abstract depicting the experimental design and results of our studies involving the non-canonical dsRBD of Drosha. The blue oval in the left most frames depicts the non-canonical WT Drosha dsRBD. The purple oval in the middle frames represents the domain that results when a canonical dsRBD KR- helix motif is mutated to that of a canonical dsRBD. In the right most frames, the grey oval depicts the domain when all KR- helix motif residues are mutated to alanine, removing all amino acid side-chain chemistry at that region. Green checkmarks represent successful Microprocessing of model pri-miRNA substrate or dsRNA binding by the domain *in vitro*. Red 'X's' depict negative results of these same experiments.

II. Technical Introduction

Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSAs) are a class of native poly-acrylamide gel-based assays centered on the principle that larger complexes, be they any combination of DNA, RNA or proteins, are less mobile than smaller mass complexes or molecules[38]. To perform these experiments, a ligand is titrated into a constant amount of labelled macromolecule, usually to a state of saturation, and the resulting gel presents retardation of the complexes in the lanes that contain higher concentrations of ligand to macromolecule. The retardation of large complexes comes about as a result of the fact that they are less likely to travel through gel pores than the components of that complex or smaller variations thereof. The last major consideration is the visualization of the gel, which can be achieved by either a covalent label on the macromolecule such as a fluorophore or radiolabel, or SYPRO protein or CYBR nucleic acid post-stains.

In the case of the EMSAs in this thesis, the macromolecule will refer to a dsRNA molecule and the ligand will refer to the dsRBD. The visualization for these assays will be achieved through an internal radiolabel via a ^{32}P-γ-phosphate at the 5′ end of the model pri-miRNA, read out using a GE Typhoon-9410 imager. It is important to note that these experiments are performed such that our protein ligand is titrated into dsRNA macromolecule across a roughly two-log concentration range as this will allow the visualization of intermediate bound states. The first ligand concentration is in several order of magnitude excess of the macromolecule as a means to attain the apparent macroscopic binding constant of the complex- something that is unattainable had we performed these experiments using stoichiometric ratios of ligand and macromolecule

(Figure 6). Each of these gels contain at least a single negative control lane, in which only the macromolecule (i.e. the dsRNA) has been loaded in order to provide a visual reference for the bands present from the unbound dsRNA, alone. Accompanying each gel image is a fit to the general Hill Equation, generated in MatLab (MathWorks), plotting fraction bound vs protein ligand concentration (Equation 1).

$$(1) \qquad Fraction\ Bound = a\left(\frac{[protein]^n}{[protein]^n + K_d^n}\right) + b$$

Where [protein] refers to the protein concentration at a given point, n is the Hill coefficient, K_d is the dissociation constant, and a and b are corrections for curve amplitude and baseline, respectively[20, 31].
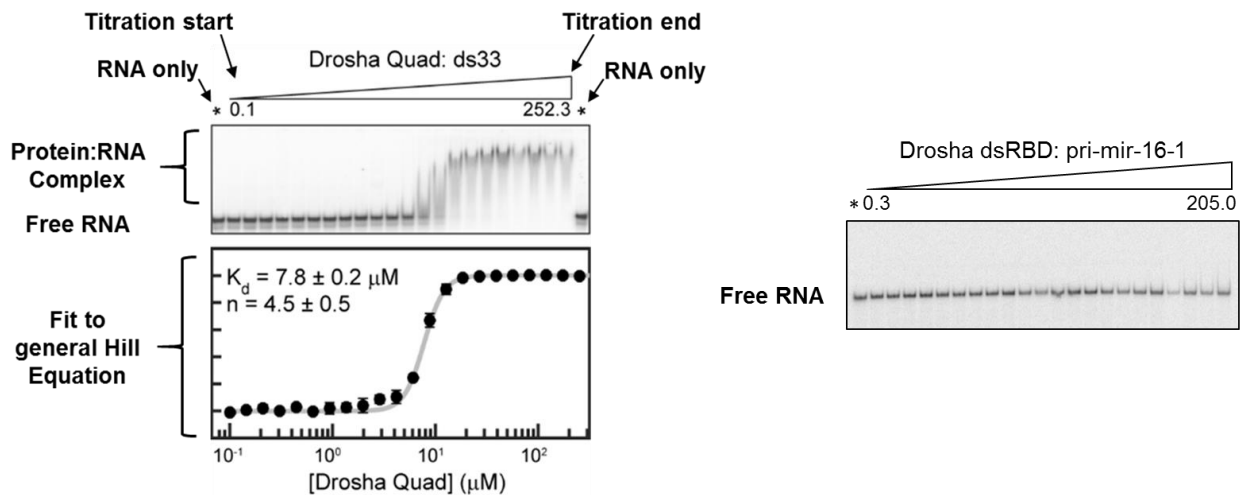
Figure 6: Sample EMSA data depicting the representative gel and fit to the general Hill equation. RNA only lanes will be depicted by 'RNA' or an asterisk, the titration lanes indicated with a wedge, and dissociation constant and Hill coefficient values annotated within the fit. Error bars are shown for one standard deviation in either direction, calculated from a minimum of two gels. Shown is a binding interaction (left, Drosha-Quad: ds33) and a non-binding interaction (right, Drosha-dsRBD: pri-mir-16-1).

Nuclear Magnetic Resonance Spectroscopy

All NMR experiments presented in this work are heteronuclear single-quantum coherence spectroscopy (HSQC) based. All NMR protein samples must be uniformly ^{15}N and if used for 3D experiments, ^{13}C labelled as these spin ½ nuclei are NMR active. In the basic ^{1}H-^{15}N HSQC experiment, a radiofrequency pulse is applied to all amide protons in the system, perturbing that set of nuclei and preparing the system for polarization transfer[39]. Polarization is then transferred to the heteronucleus, ^{15}N amide nitrogen, via the insensitive nucleus enhanced by polarization transfer (INEPT) method[40]. Chemical shift of the heteronucleus is indirectly labelled in the subsequent evolution period, followed by polarization transfer back to amide proton via the inverse of the initial INEPT. The proton chemical shift is then labelled via the radiofrequency detector on the NMR instrument.

The resulting HSQC spectrum yields a single signal for each amide N-H pair in the system (Figure 7). The cross-peaks are distributed in a 2D spectrum at the intersection of the proton and nitrogen chemical shifts for any given N-H pair. These chemical shifts are dependent on the chemical environment of the N-H pair and each HSQC spectrum is unique to a given protein construct. While little structural information exists solely in the HSQC spectrum, a few distinct characteristics important in this work can be identified. Generally, well dispersed signals indicate that each residue is in a unique chemical environment which typically indicates a folded structure. Likewise, collapse of signals to 8 ppm in the proton dimension indicates that residues in this regime are solvent exposed and not highly ordered[41, 42]. Lastly, multiple signals per residue indicates that the protein construct may be in a state of slow exchange or demonstrate

dynamics on a timescale slower than the NMR timescale. It is important to note that exceptions to these generalizations exist and in order to make more informed structural and dynamic inference, expansions of the basic HSQC must be employed[43, 44].
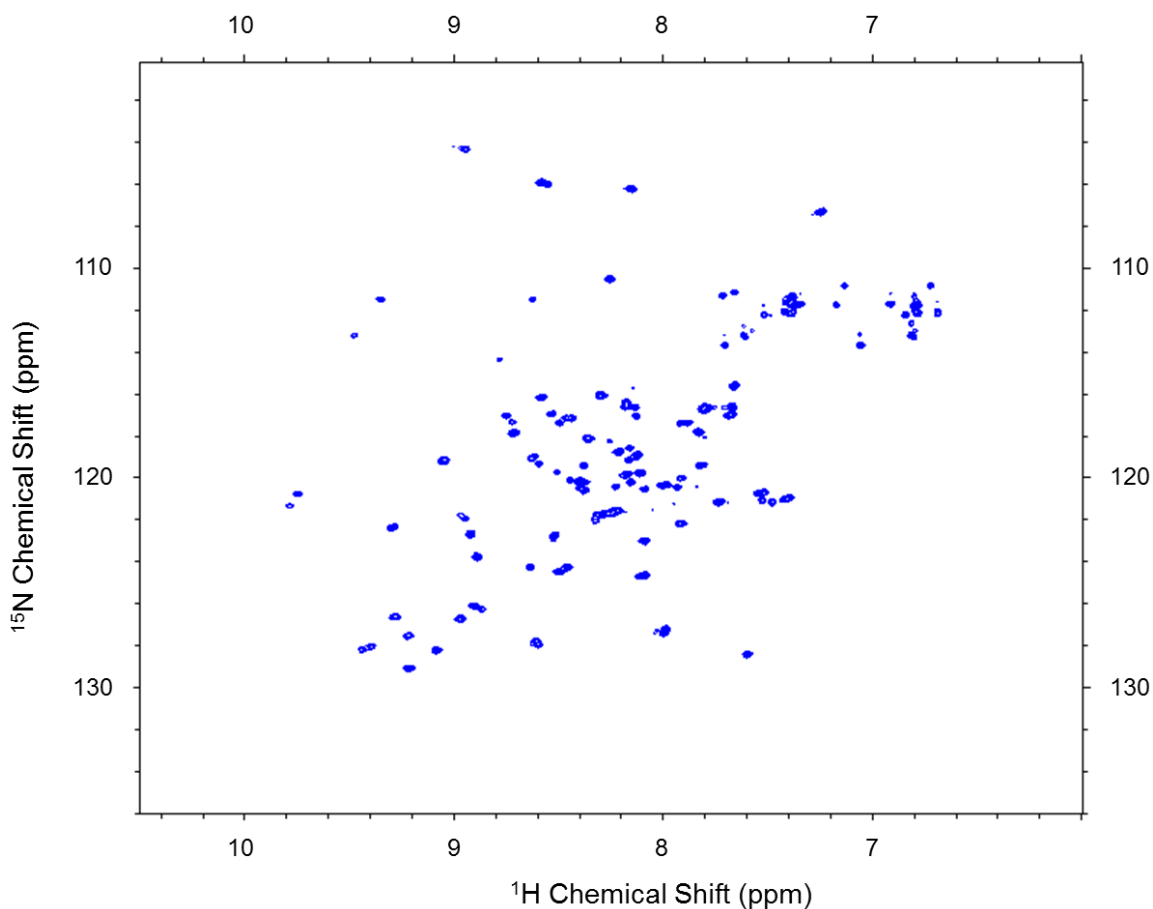
Figure 7: Representative HSQC for WT Drosha dsRBD. Note the high signal dispersion, which indicates that the domain is highly ordered. Each signal in this spectrum represents a single, unique N-H pair in the peptide.

Expanding the HSQC to include a second indirect dimension, typically ^{13}C, is one of the more commonly utilized expansions in protein NMR. These experiments utilize the identical polarization transfer pathway as outlined earlier with additional coherence transfer from the nitrogen to the desired spin system and then back via the reverse path.

The resulting three-dimensional spectrum can be conceptualized as a data cube in which each cross-peak represents the amide N-H pair and the carbon or set of carbons in a given spin system. The most commonly implemented 3D experiments in protein NMR are the standard triple resonance experiments for protein backbone assignment[45-48].

These pulse sequences are run as two paired experiment sets - the HNCO/ HN(CA)CO and CBCA(CO)NH/ HNCACB. Each set of experiments contains a unidirectional (HNCO and CBCA(CO)NH) and a bidirectional (HN(CA)CO and HNCACB) pulse sequence in which backbone carbon resonances for residue i-1 and residues i-1 and i are obtained for a given N-H pair, respectively. This phenomenon is possible because polarization transfer between the amide nitrogen of residue i and the alpha carbons of residues i and i-1 occurs concurrently whereas concurrent polarization transfer from the amide nitrogen of residue i to the carbonyl carbon of residues i and i-1 is not possible due to vast differences in the coupling constants for N_i/CO_i and N_i/CO_{i-1}. The data obtained from these experiments are usually represented in strip plots that are used to 'daisy-chain' through the protein backbone and assign protein residues to HSQC signals (Figure 8). This task is made easier because any degeneracy in residue spin systems is usually broken by this combination of experiments and that mean CA and CB chemical shifts for all twenty amino acids are well documented (BMRB).
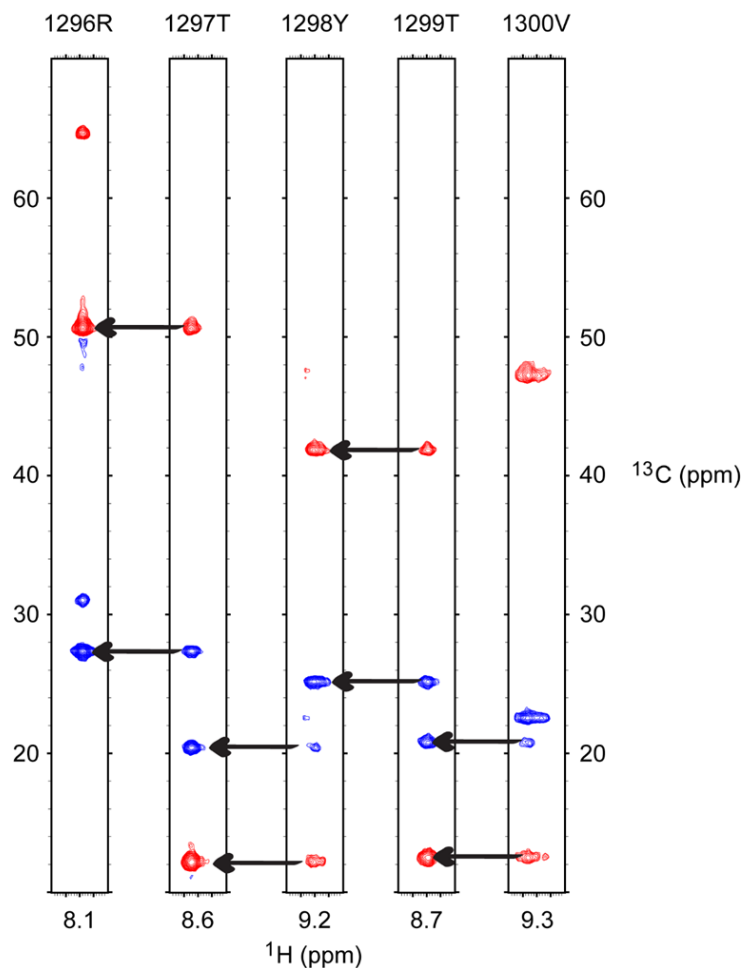
Figure 8: HNCACB strip plots showing a sample walk through the peptide backbone. One key feature of the HNCACB experiment is that CB is 180° out of phase, making for easy distinguishing of CA and CB. Each strip contains the CA and CB signals for the residue labelled above the strip and the preceding residue. Identification of self and preceding residue signals is performed using the unidirectional CBCACONH. Accuracy of this walk was confirmed with an analogous analysis of the HNCO/HNCACO walk.

Once all of the protein backbone nuclei are assigned, the chemical shift data can be used in a number of tasks that are essential for the determination of biomolecular structure and structure/function relationships. For example, in this thesis I will report data used to collect through-space distance constraints and dihedral angle constraints[49, 50] that enabled us to calculate the structure of the re-functionalized Drosha-dsRBD that is the

focus of Chapter II. Complete backbone resonance assignment also enables other analyses that are integral to determining the dynamics of the peptide at hand, including per-residue relaxation data[51, 52].

Adding dynamic information to the otherwise static structural picture is one of the main advantages of NMR spectroscopy. To achieve this, a number of NMR experiments have been designed to probe different aspects of peptide backbone relaxation via slight modifications to the basic 2D HSQC experiment. Depending on the pulse sequence modification, longitudinal relaxation (T1), transverse relaxation (T2) and heteronuclear NOE (hetNOE) data can be obtained. These chemical environment-dependent variables report on the recovery rate to thermal equilibrium after excitation, the recovery rate to magnetic equilibrium after excitation, and relaxation after polarization saturation, respectively[51, 52]. These parameters are utilized in a model-free analysis to calculate an order parameter (S^2) that directly reports on the flexibility of the N-H bond vector within the peptide structure[51, 53]. In the work presented, the chemical shift derived dihedral angles, through-space NOE interaction assignments and NMR relaxation data are utilized as input for solution structure calculation via the XPLOR- NIH software package[54].

Human 293T Cell-based pri-miRNA Processing Assays

In conjunction with the biophysical analyses described above, it is important to identify the functional manifestations of our biochemical manipulations to the dsRBD of Drosha. To achieve this, *in vitro* Microprocessing assays were performed utilizing full length Microprocessor components and a model pri-miRNA-16-1 substrate. The human

embryonic kidney cell line (HEK293T) was chosen as it has been optimized for effective recombinant protein expression and ease of cell line maintenance[55]. In preparation for the processing assay, liposome transfection is employed to introduce plasmid DNAs encoding the full length protein constructs of choice to the HEK293T cells[56]. Once the cells reach 100% confluence and the target proteins are expressed, the cells are harvested and lysed to yield a cell lysate containing the protein constructs encoded by the transfected DNA plasmids.

The processing assays presented in this thesis are performed by incubating end- or body-labelled model pri-miRNA-16-1 in a number of different protein construct-containing HEK293T cell lysates for an equivalent period of time. After this incubation period, the processing reactions are run on a denaturing gel and imaged to yield the extent to which the substrate was processed. The negative control in these experiments in which no DNA plasmid is transfected, labelled as 'mock', reflects the Microprocessing function of the endogenous Microprocessor components and the background of all samples. Knockdown of endogenous Drosha to eliminate background was attempted using RNAi, but was not tolerated by the cells (Durga Ghosh, personal communication). The positive control, in which the full- length WT Microprocessor components Drosha and DGCR8 are transfected, reflects the Microprocessing efficiency of the native complex. The other samples depicted in these experiments consist of varying Drosha constructs and WT full- length DGCR8 and probe the Microprocessing ability of different Drosha constructs. In these samples, abolishment to 'mock' levels indicates that the transfected Drosha construct cannot process the pri-miRNA substrate. The results of these

experiments, shown and discussed in Chapter II, proved integral in providing functional insight to our biophysical investigation of Drosha's dsRBD.

III. Thesis Overview

Chapter two outlines the work the author performed in probing the role of Drosha's dsRBD in Microprocessing. Traditional biophysical techniques and cell-based processing techniques were utilized in conjunction to tell a holistic story of the domain. Four point mutations to the KR helix motif region of Drosha's dsRBD were used to engineer a dsRNA binding-competent dsRBD, Drosha-Quad. EMSAs and NMR experiments confirmed that Drosha-Quad had the structure and dynamics of a canonical dsRBD and bound dsRNA via the canonical dsRBD mode. Processing assays utilizing full length Microprocessor proteins and model pri-miRNA-16-1 identified that Drosha's dsRBD is implicated in Microprocessing. Furthermore, the chemistry at the KR helix motif of this domain is the driving force of its role.

Chapter three focuses on the future directions that the author's work could potentially lead works in the Showalter Lab and others' studying the miRNA maturation pathway. These future directions focus on the C- terminal intrinsically disordered regions of the Microprocessor proteins, Drosha and DGCR8. EMSAs suggested that addition of the native C- terminus of Drosha to the WT Drosha dsRBD construct yields a dsRNA binding competent construct. Further NMR analysis revealed that presence of this C-terminal IDR stabilized the dsRBD fold and may interact with the dsRBD fold or contain some helical content. NMR analysis of DGCR8's native C- terminus revealed that this

domain may not be entirely disordered and it is quite likely that much of this IDR may be

helical in nature.

Chapter 2: Toward Understanding Drosha dsRBD's Role in Microprocessing

Chapter II Collaborators:

Durga Ghosh: Collaboration with pri-miRNA processing assays; performed Western Blot analysis.

Debashish Sahu: Collaboration with HSQC Assignment and NMR data collection; performed NMR structure and dynamics calculations and aided in generation of associated figures.

I. Introduction

Micro-RNAs (miRNAs) are a class of small (20-22 nt) non-coding RNAs known to primarily function in the cellular process of RNA silencing. Mature miRNAs have been shown to suppress genes necessary for tumor cell proliferation, innate immunity and cellular senescence. In order to execute its role in post-transcriptional gene regulation, a mature miRNA binds a complementary messenger RNA (mRNA) in the RNA induced silencing complex (RISC) and inhibits translation of the target mRNA[12].

The canonical miRNA maturation pathway begins with a primary transcript (pri-miRNA, often encoded within the introns of nascent mRNA transcripts) that is processed in the nucleus via the Microprocessor complex, which is comprised minimally of the

catalytic unit, Drosha, and its processing-dependent cofactor, DGCR8[6, 32]. After this initial processing, the precursor miRNA (pre-miRNA) is exported from the nucleus via the Exportin-5 pathway and further processed in the cytosol by a minimal complex of Dicer and TRBP[8]. This complex then associates with an Argonaut protein for transfer of the miRNA guide strand into RISC where regulation of complementary mRNAs occurs[11, 12].

Double-stranded RNA binding domains (dsRBDs) are present in each major protein in the miRNA maturation pathway, making them likely candidate domains for RNA substrate recognition. Each dsRBD fold is comprised of a strictly conserved $\alpha-\beta-\beta-\beta-\alpha$ structure; sequence is not necessarily conserved, although some consensus elements have been identified[26]. Both RNase III enzymes, Drosha and Dicer, contain a single C-terminal dsRBD, whereas DGCR8 and TRBP each contain multiple dsRBDs. The canonical binding face for dsRBDs contains a KR-helix motif beginning at the base of the second helix, but the presence of the motif is not strictly conserved throughout the pathway. For example, it is not present in dsRBD3 of TRBP and PACT or in Drosha's dsRBD (Figure 9).

The Kim group's recent identification of the heterotrimeric nature of the Microprocessor complex and implication of the C- terminus of DGCR8 in Microprocessing has brought a greater understanding to this process[6]; however, there are a number of questions still unanswered. While most of the DGCR8 regions have been functionally characterized, the central domain and dsRBD of Drosha remain functionally uncharacterized; yet both are obligate in Microprocessing. Recently, the Doudna group structurally-characterized the role of PACT/TRBP dsRBD3 in mediating interaction with Dicer, which has motivated us to investigate the role of Drosha's dsRBD in the

Microprocessor complex. These domains both exhibit divergence from the canonical KR

Helix motif, the extent of which is vastly greater in the case of Drosha's dsRBD.

A

```
                 α1                    β1           β2          β3              α2
           *                                                          **   *
Drosha   DPKSQLQQCCLT-LRTEGKEPDIPLYKTLQTVGPSHARTYTVAVYFKGERIGCGKGPSIQQAEMGAAMDALEKYNF
Dicer    VPRSPVRELLEM-EPE------TAKFSPAER---TYDGKVRVTVEVVGKGKFKGVGRSYRIAKSAAARRALRSLKA
DGCR8-1  SEVCILHEYMQR-VLKV-----RPVYNFFECENPSEP--FGASVTIDGVTYGSGTASSKKLAKNKAARATLEILIP
DGCR8-2  SPYQILHECLKRNHGMG-----DTSIKFEVVPGKNQKSEYVMACGK--HTVR-GWCKNKRVGKQLASQKILQLLH-
TRBP-1   TPISLLQEYGTR-IGK------TPVYDLLKAEGQAHQPNFTFRVTV-GDTSCTGQGPSKKAAKHKAAEVALKHLKG
TRBP-2   NPVGALQELVVQ-KGWR-----LPEYTVTQESGPAHRKEFTMTCRVERFIE-IGSGTSKKLAKRNAAAKMLLRVHT
TRBP-3   ACCRVLSELSEE-QAF------HVSYLDIEELSLSGLCQCLVELSTQPATVCHGSATTREAARGEAARRALQYLKI
Quad     DPISQLQQCCLT-LRTEGKEPDIPLYKTLQTVGPSHARTYTVAVYFKGERIGCGKGPSKKQAKMGAAMDALEKYNF
```

B

C

Figure 9: Homology of miRNA Maturation Pathway Components.

(A) Sequence alignment of miRNA maturation pathway dsRBDs. Gross secondary structure is annotated above and shaded boxes are used to annotate structures within individual sequences. Asterisks and bolded text are used to annotate the residues mutated to generate the Drosha-Quad mutant.

(B) Cartoon depicting the Microprocessor complex, modelled as such to reflect the recent results of the Kim Group. DGCR8 (dsRBDs green ovals, IDRs black wavy lines) recognizes the distal loop while the catalytic unit, Drosha (grey scale, dsRBD blue), recognizes and processes the apical stem.

(C) Cartoon depicting the pre-RISC Dicer/TRBP/pre-miRNA complex, modelled as such to reflect the recent results of the Doudna Group. TRBP's dsRBDs 1 and 2 (Red) interacts with the stem of the pre-miRNA while dsRBD3 (Blue) modulates the protein: protein interaction with the catalytic unit, Dicer (grey scale).

In the following chapter, I present my success in engineering a dsRNA binding Drosha dsRBD mutant, which was achieved via site directed mutagenesis at the KR-Helix motif region. I refer to the generated construct throughout as Drosha-Quad. Biophysical characterization of this engineered domain and cell-based pri-miRNA processing assays, probing the biological impact for pri-miRNA processing by full length Drosha possessing the "Quad" variation, are presented. These results demonstrate that avid dsRNA binding by the dsRBD of Drosha is incompatible with efficient pri-miRNA processing. Further, they demonstrate the chemistry of the KR- helix motif is the driving force for Drosha dsRBD's function in Microprocessing. Thus, these findings bring the field one step closer to assigning an affirmative role for Drosha dsRBD in Microprocessing and suggest that the chemistry at the canonical dsRNA binding face of the seemingly innocuous domain is indispensable.

II. Materials and Methods

Protein Preparation

Mutants were generated using QuickChange Lightning site-directed mutagenesis kit (Agilent Technologies) standard procedures from a WT Drosha dsRBD (1259-1337) sequence in pET-47b(+) vector. Upon sequence conformation, the mutant plasmid was transformed into BL21 competent Escherichia coli (E. coli) for overexpression. Cells were grown to $OD_{600} = 0.500$ at which point they were induced with 500 mM final concentration IPTG at 24 °C for 16 hours. Cells were lysed via sonication and the lysate was clarified via centrifugation at 4°C. Protein was purified via nickel-affinity column chromatography

as previously reported. Protein was buffer exchanged in to the desired buffer system, indicated in each method section below.

NMR Methods

Drosha dsRBD Quad backbone resonances were assigned using standard triple resonance experiments. Side chain aliphatic carbons and protons were assigned using triple resonance TOCSY- mixing based experiments. All triple resonance experiments were performed using 25-40% sparsity non-uniform sampling. Spectra were processed in Topsin 3 (Bruker) and analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Ensemble structures were generated using XPLOR-NIH.

All experiments were performed on a Bruker Avance III 600 MHz spectrometer equipped with a TCI cryoprobe for enhanced sensitivity. The sample conditions for all experiments were 840 uM protein in 100 mM Cacodylate pH 7.3, 100 mM KCl and 5 mM BME. The temperature of the spectrometer bore was 25°C.

EMSA Methods

For the variable length duplexes, the top strand RNA was 5′-end labelled with γ-^{32}P-ATP and mixed in a 3-fold molar excess of cold bottom strand. The duplex was purified from an 8% acrylamide native gel. The pri-miRNA 16-1 was similarly labeled with ^{32}P-ATP and renatured via heating at 90°C for 1 minute and snap cooling at 4°C for 5

min. Binding reactions were run in 50 mM cacodylate pH 6.0, 50 mM potassium chloride, 5% glycerol (v/v), 1mM dithiothreitol, 0.1 mg/ mL Bovine Serum Albumin and 0.1 mg/ mL herring sperm DNA for thirty minutes at room temperature. The reactions were loaded onto a 10% acrylamide native gel and run at 200 V for 3.5 hrs at 4°C.

Gels were imaged using a Typhoon-9410 imager and then quantified using the software ImageQuant (GE Healthcare Life Sciences). Boxes were drawn for the bound and unbound RNA state and the fraction bound was calculated using the ratio of the bound RNA intensity to the sum of this value and the intensity of the free RNA state. The points in the titration curves represent the mean fraction bound from two gels and the error indicated is the uncertainty of this value to one standard deviation. The curves were fit to the general Hill equation binding model using Matlab (MathWorks).

Processing Assay Methods

All Drosha constructs and WT DGCR8 were overexpressed in HEK-293T cells. Plasmid was transfected into cells using Lipofectamine 2000 (Invitrogen) per the packaged instructions. Mock was similarly treated with Lipofectamine using the same procedure, but lacking plasmid DNA. Forty-eight hours later the cells were washed and subsequently harvested using phosphate-buffered saline and lysed via sonication using a buffer comprised of 20mM Tris, 100 mM KCl and 0.2 mM EDTA at pH 8.0. The lysate was cleared of cellular debris via centrifugation and combined with RNasin (Promega), 10 fmol 5′ γ-^{32}P end-labelled or uniformly α-^{32}P-UTP body-labelled pri-mir-16-1 and MgCl2 was added to a final concentration of 6.4 mM. The reactions were incubated at 37°C for

30 minutes and immediately thereafter loaded onto 10% polyacrylamide denaturing gels for analysis.

Structural calculations of Drosha-Quad mutant

The NMR data on Drosha-Quad was used as the primary inputs for the calculation of its tertiary structure. The chemical shifts of C', N, C_α, C_β, H_α and H_N of Drosha-Quad was used as inputs for the TALOS-N software to obtain φ and ψ dihedral constraints for the protein backbone. In addition to this data, we included the $^3J_{HNH\alpha}$ scalar couplings that report directly on the dihedral angles as described before. We also used a total of 455 NOEs including signals arising from amide-amide and amide-side chain proximity. All these values were used as inputs into the standard simulated annealing algorithm in XPLOR-NIH program, which starts from the extended structures folding iteratively into the final collapsed structure. This final structure is penalized for non-conformity to input data while favoring least amount of clashes and non-agreement with input data. The final structures were energy minimized to relax away the steric clashes while conforming the input restraints. The top ten structures were chosen with the least energy term used for the minimization and structure calculation.

Modelfree analysis

The dynamic properties of Drosha-Quad was examined using the Lipari-Szabo model-free analysis using the ModelFree 4.20 software. The diffusion tensor used in these calculations were done using the quadric analysis. The three dimensional

coordinates from the XPLOR structures calculated above from Drosha-Quad were used in this analysis for the determination of diffusion tensor. The model function 2 was chosen for the fitting of R_1, R_2 and NOE relaxations to determine S^2 and τ_e.

III. Results

KR Helix Motif Chemistry Drives dsRNA binding via dsRBDs

Our initial efforts focused on mutating the KR helix motif region directly as we hypothesized that it was possible to engineer dsRNA binding function into the domain by manipulation of this region. Recovery of the KR- helix motif (I1317K Q1318K E1321K) afforded a construct that was not soluble enough to assay. Upon expanding the sequence alignment to the entire domain, it was clear that the presence of a fourth lysine (K1262) directly adjacent on helix-1 was an anomaly in the domain family. Mutation of this residue to a β-branched aliphatic residue (K1262I), more representative of the dsRBD sequence motif, yielded a domain with dsRNA affinity indistinguishable from that of other dsRBDs in the miRNA maturation pathway (Figure 10, Table 1).

In addition to validating that we could engineer general dsRNA binding ability into the dsRBD of Drosha, it was important from a protein chemistry standpoint to validate the extent to which Drosha-Quad recapitulates canonical dsRBD function. As such, we screened Drosha-Quad against multiple lengths of canonical Watson-Crick dsRNA duplex (Figure 10). Using these EMSAs, we ascertained that Drosha-Quad binding affinity exhibits a dependence on substrate length, such that as dsRNA lattice length increases, binding becomes tighter. This trend is the same that we have observed in other canonical

dsRBDs in the miRNA maturation pathway. This work further reinforces that KR helix motif chemistry is the primary driving force of dsRBD:dsRNA interaction.
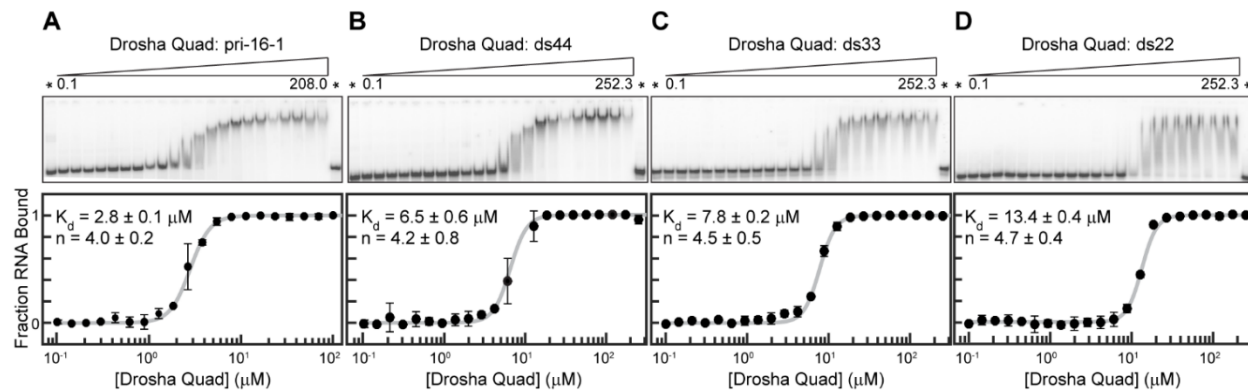
Figure 10: Electrophoretic Mobility Shift Assays of Drosha-Quad.

(A) Drosha-Quad titrated into pri-miRNA 16-1 analog, (B) ds44, (C) ds33, and (D) ds22 perfect Watson-Crick duplexes. Asterisks denote RNA only lanes and protein concentration ranges are indicated below the wedge. Fits to the general Hill Equation are presented below with error bars pertaining to one standard deviation in each positive and negative direction for 2 gels; dissociation constants and hill coefficient values with errors to one standard deviation are presented as insets.

Table 1. Best fit macroscopic binding affinities (K_d, μM) as determined by EMSA for dsRBD constructs.

	Drosha-Quad	Dicer dsRBD	DGCR8 dsRBD1	TRBP dsRBD1	TRBP dsRBD2
Native sequence[a]	2.8 ± 0.1	2.2 ± 0.1	9.7 ± 0.6	N/D[b]	N/D
ds44	6.5 ± 0.6	2.4 ± 0.1	5.9 ± 0.1	0.8 ± 0.05	0.8 ± 0.06
ds33	7.8 ± 0.2	4.9 ± 0.1	8.8 ± 0.2	0.9 ± 0.3	1.0 ± 0.08
ds22	13.4 ± 0.4	6.5 ± 0.1	21 ± 1.0	3.5 ± 0.2	1.7 ± 0.1

[a]pri-miRNA-16-1 for Drosha and DGCR8 constructs and pre-miRNA-16-1 for Dicer and TRBP constructs.

[b]Binding not determined.

Drosha-Quad is Structurally and Dynamically Homologous to WT Drosha dsRBD

Next, we solved the solution structure of Drosha-Quad as validation that our engineered mutant maintained the three-dimensional integrity of the domain. Analysis of the HSQC overlay, comparing WT Drosha dsRBD and Drosha-Quad, supports the conclusion that the overall structure of the domain does not change significantly due to the mutation process, as evidenced by the minimal peak movement and overall preservation of signal pattern (Figure 11). Using traditional three-dimensional NMR experiments in conjunction with the TALOS-N and XPLOR-NIH software packages, we solved the NMR solution structure of the mutant domain (Figure 11). The twenty best structures of Drosha-Quad reveal several interesting points: the domain maintains the canonical dsRBD $\alpha-\beta-\beta-\beta-\alpha$ fold, the proximity of amino acid position 1262 relative to the KR helix motif validates the necessity of that mutation, and the bundle is well constrained. The back-calculated NMR statistics confirms that the bundle is not over-constrained and correlates well with the empirical data (Table 2).

Next, we focused on analyzing the similarity between the Drosha-Quad and Drosha WT dsRBD structures. Strikingly, the lowest energy structures are nearly identical save for interactions between the α1-β1 loop and the C- terminal tail of the Drosha-Quad construct (Figure 12). The lowest energy structure overlay convincingly shows how structurally analogous these domains are- save for some variation in the helix-1 and α1-β1/ C- terminal tail regions (Figure 12). Thus, we are justified in concluding that these domains are, structurally, very similar and that Drosha-Quad is irrefutably a dsRBD.
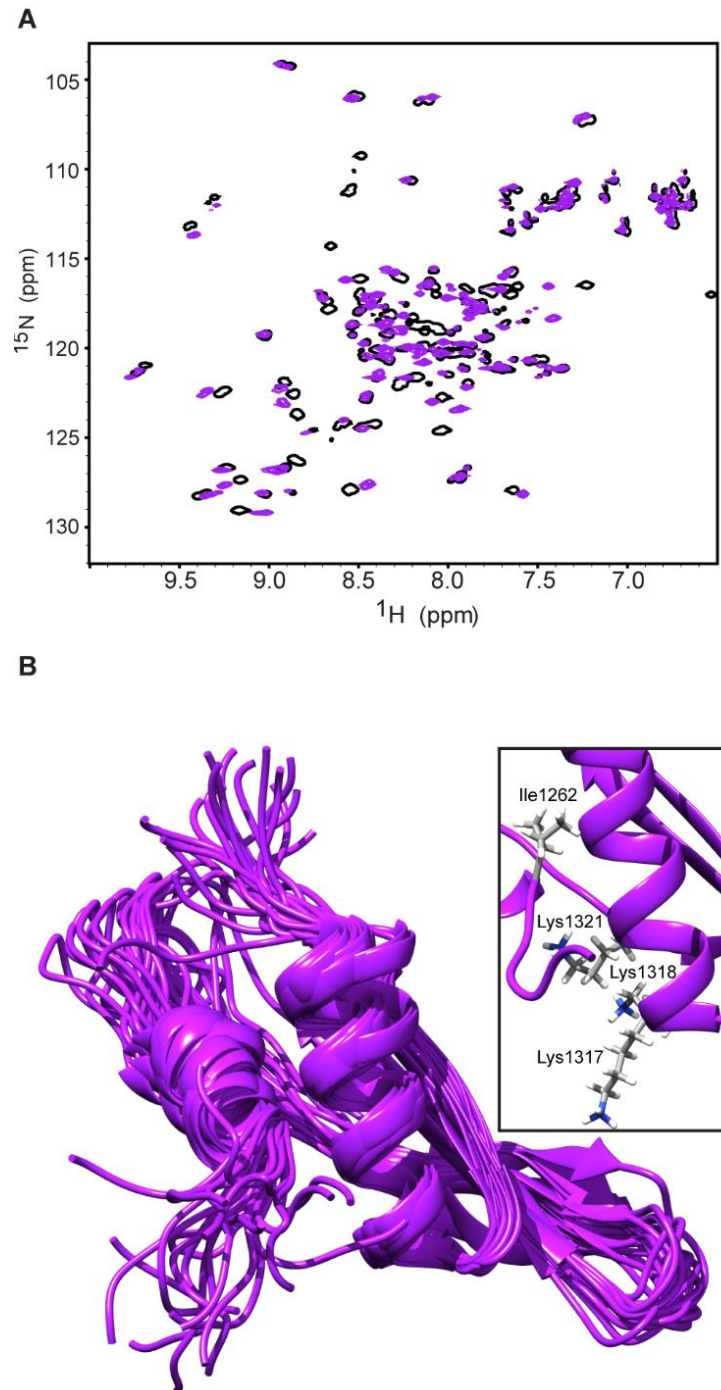
Figure 11: NMR analysis of Drosha-Quad.

(A) HSQC overlay of WT Drosha dsRBD (hollow black) and Drosha-Quad (purple).

(B) Twenty lowest energy structure bundle for Drosha-Quad. Inset depicts the KR Helix motif spatial region of the lowest energy Drosha-Quad structure with residues mutated from WT shown and annotated.

Table 2: NMR statistics of Drosha-Quad bundle.

NMR Distance and Dihedral Constraints	
Distance constraints	
Total NOE	455
Intraresidue	289
Inter-residue	166
Sequential ($|I - j| = 1$)	182
Medium range ($|I - j| \leq 4$)	398
Long range ($|I - j| \geq 5$)	57
Intermolecular	0
Hydrogen bonds	0
Total dihedral angle restraints	128
ϕ (TALOS)	64
ψ (TALOS)	64
χ_1 (J couplings)	0
J Couplings (Hz)	
$^3J_{HNHA}$	65
Structure Statistics	
Violations (mean ± SD)	
Distance constraints (Å)	0.075 ± 0.017
Dihedral angle constraints (°)	1.242 ± 0.493
J Couplings (Hz)	0.921 ± 0.125
Deviations from idealized geometry	
Bond lengths (Å)	0.004 ± 0.000
Bond angles (°)	0.513 ± 0.033
Impropers (°)	0.404 ± 0.046
Average pairwise rmsd (Å)	
Heavy	2.74
Backbone	2.01
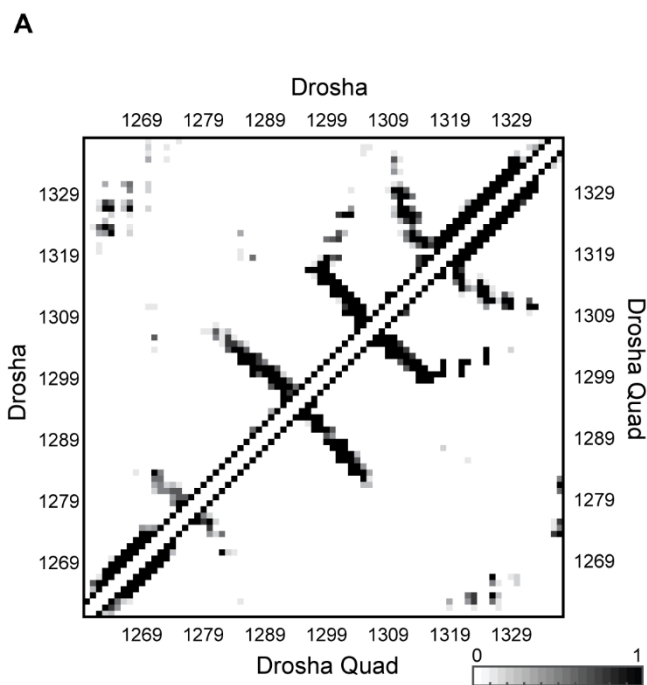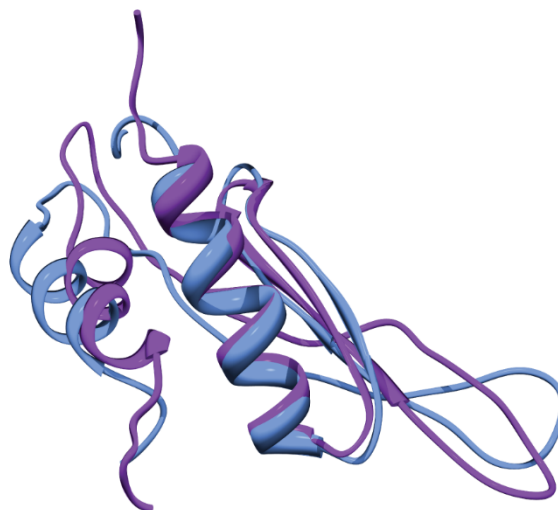
Figure 12: Comparative analysis of WT Drosha dsRBD and Drosha-Quad structure.

(A) CA-CA contact map comparing WT Drosha dsRBD to Drosha-Quad. Probabilities for CA-CA contact between residues are scaled 0 (White) to 1 (Black).

(B) Alignment of WT Drosha dsRBD (blue) and Drosha-Quad (purple) lowest energy structures.

While demonstrating the general structural comparison is advantageous, it is perhaps more rigorous to confirm that the structural dynamics of the Drosha-Quad construct are consistent with those determined by Wostenberg *et al.* in 2010 for the WT domain[20] (Figure 13). The calculated order parameters for WT Drosha dsRBD and Drosha-Quad are nearly identical, indicating that the overall structural dynamics of the two constructs are indistinguishable (Figure 14). Taking this more holistically, the rigid secondary structures in WT Drosha dsRBD remain rigid in Drosha-Quad and the dynamic loops and termini remain so as well (Figure 14). Confirmation that the overall structure and dynamics of Drosha-Quad were equivalent to that of the wild type domain allows us to further probe functional differences of the domain, understanding that the only difference between the two constructs was the chemistry at the KR Helix motif.
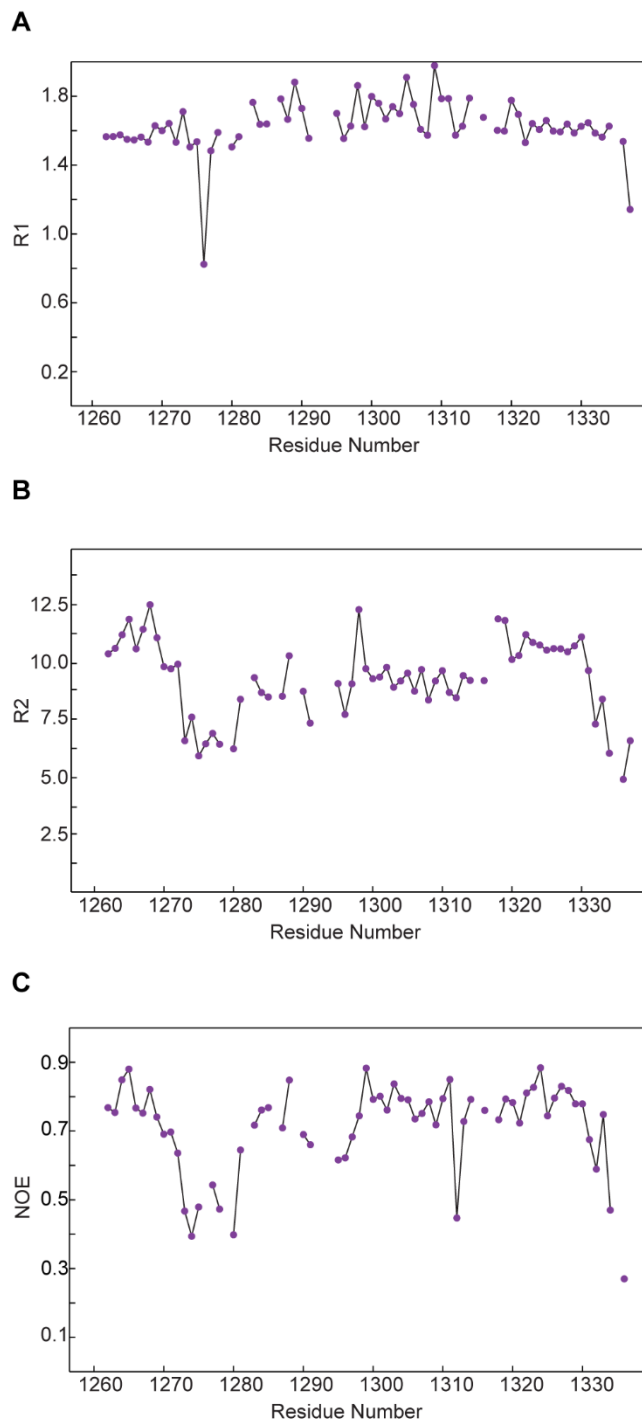
Figure 13: Per Residue Relaxation Data for Drosha-Quad.

(A) Per residue longitudinal relaxation rates, (B) Transverse relaxation rates and (C) {^{1}H}-^{15}N NOE data collected at 600 MHz.
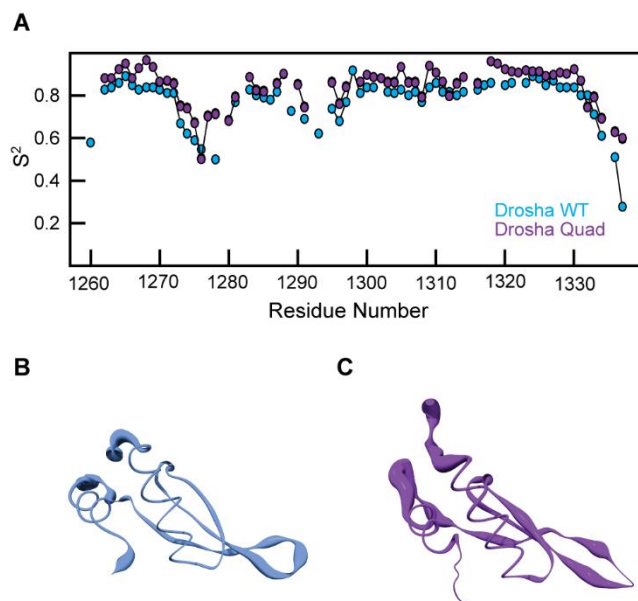
Figure 14: Structural dynamics comparison of WT Drosha dsRBD and Drosha-Quad.

(A) Per- residue order parameters for WT Drosha dsRBD and Drosha-Quad.

(B) Worm model plotting order parameter onto the ribbon diagram and WT Drosha dsRBD and (C) Drosha-Quad. Thicker regions depict residues with low order parameter values and thinner regions denote residues with order parameters closer to one. Unassigned residues and proline are scaled to be thinner than even the most rigid regions of the domains.

Drosha dsRBD KR Helix Motif Chemistry is a Key Component of Microprocessing

Next, we tested the effects that dsRNA binding competency would have on the processing ability of the full catalytic unit. Upon mutation to the binding competent domain, processing of our model pri-miRNA-16-1 substrate is abrogated to mock levels (Figure 15). Western Blot analyses of the transfected HEK 293T cells confirms that all mutants were successfully overexpressed (Figure 15). As an additional control, we mutated the three KR-Helix motif residues to alanine and were afforded an identical result (Figure 16). We confirmed, through EMSAs, that this mutant does not bind dsRNA (Figure 16). Interestingly, deviation from the WT Drosha dsRBD KR Helix motif chemistry yields

an identical result to deletion of the domain from the full length enzyme. These results

implicate some function of the KR Helix motif of Drosha's dsRBD in Microprocessing.
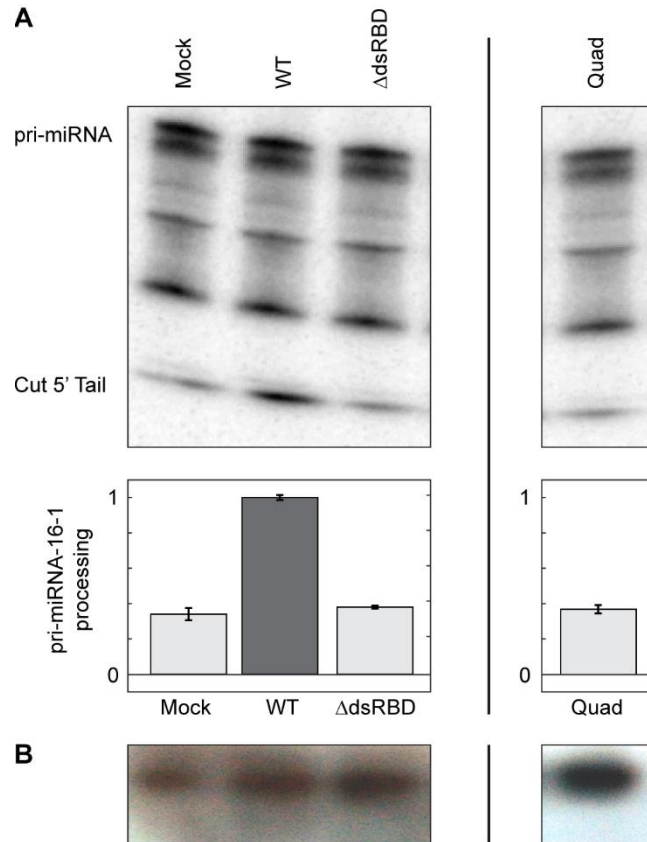
Figure 15: HEK 293T cell-based processing assays.

(A) 5′ end-labelled pri-miRNA 16-1 processing assay gel showing Microprocessing of the model pri-miRNA. The pri-miRNA and 5′ cut end bands are labelled. All samples were transfected with their respective full length Drosha plasmid full length DGCR8. The degradation bands present in mock and all samples result from other active RNases is the HEK 293T cell lysate.

(B) Western blot confirming over expression of full length Drosha mutants.
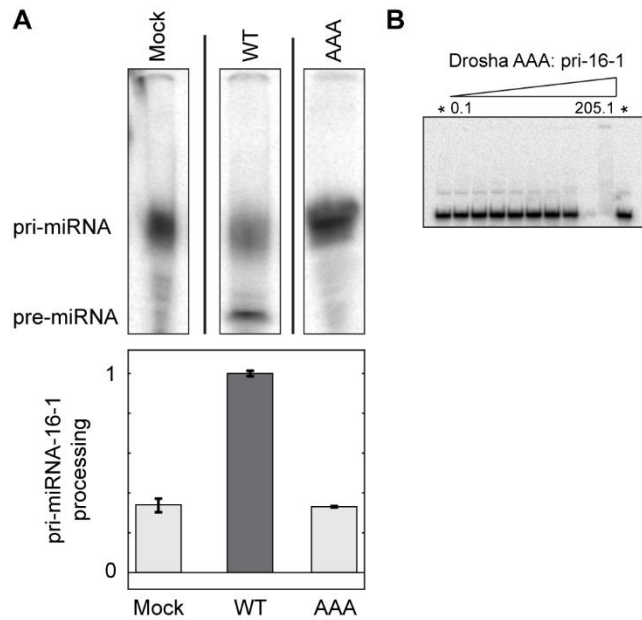
Figure 16: Triple Alanine KR Helix Motif Analysis.

(A) Uniformly α-^{32}P-ATP body-labelled pri-miRNA 16-1 processing assay gel showing Microprocessing of the model pri-miRNA. Bands of interest are labelled. All samples were transfected with their respective full length Drosha plasmid full length DGCR8.

(B) Representative EMSA gel indicating that the triple alanine is not dsRNA binding competent.

IV: Discussion

The recent success of the Doudna Lab biophysically characterizing the Dicer PBD: TRBP dsRBD3 interface brought renewed interest to understanding the function of non-canonical dsRBDs[19]. While the non-canonical dsRBD of Drosha and TRBP's dsRBD3 are similar in their lack of dsRNA-binding ability, the data presented herein suggests that Drosha dsRBD's obligate role in Microprocessing occurs via a very different mechanism than that of TRBP dsRBD3's role in pre-miRNA processing. The Doudna lab definitively showed that the Dicer PBD: TRBP dsRBD3 interaction occurs via the β-sheet face of

TRBP dsRBD3[19]. The biophysical and biochemical data presented in this work suggest that Drosha dsRBD's function is instead driven by the chemistry of the KR- helix motif.

Not only is this system remarkable from a biological function standpoint, it also intriguing from a protein chemistry point of view. While a number of interesting deviations from canonical dsRBD structure have been documented, including additional secondary structural elements and length variations of regions in the dsRBD fold, there are no examples of a domain utilizing the canonical dsRBD: dsRNA binding face for another function. Also, while it was not entirely surprising that the KR- helix motif chemistry could be manipulated to generate a dsRNA-binding competent domain, it was intriguing that the Drosha-Quad exhibited canonical dsRNA length dependence and maintained nearly identical structure and dynamics- such drastic point mutations in a dense, well ordered region of a small protein domain might be expected to cause a much greater perturbation to the system. This further confirms that KR- helix motif chemistry is the defining article in canonical dsRBD function, as well as the non-canonical function of Drosha's dsRBD.

The *in vitro* Microprocessing assays were integral in determining this key result- first in determining that Drosha's dsRBD is a necessary component of Microprocessing and second, that the chemistry of the KR- helix motif is obligate to the domain's function. Having solved the NMR solution structure of Drosha-Quad and validated that the overall structure and dynamic characteristics of Drosha-Quad were consistent with that of the WT dsRBD, we were justified in proceeding to test the consequences of KR- helix motif manipulation on Microprocessing. Where the Drosha-Quad processing result may point to dsRNA-binding competency as a reason for Microprocessing ablation, the processing results of the triple alanine mutant concretely demonstrated that this was not the case. In

fact, our data strongly supports the hypothesis that Drosha's dsRBD utilizes the canonical

dsRBD: dsRNA binding face in a novel, albeit yet to be determined, obligate role during

Microprocessing.
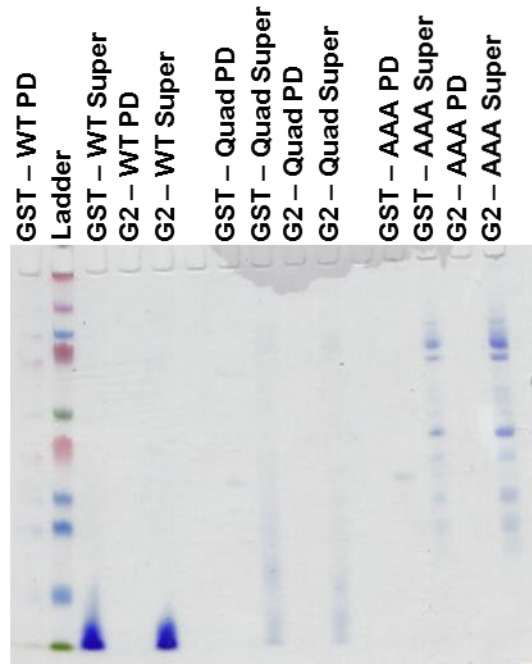
Figure 17: Preliminary pulldown experiment probing a potential G2 interaction with Drosha dsRBD (and KR- helix motif mutants). G2 was expressed as a GST fusion and GST- G2 and GST (negative control) were bound to glutathione-agarose beads and incubated with recombinant Drosha dsRBD constructs. Bands in the pulldown (PD) lanes at 10 kDa (bottom green ladder band) would indicate an interaction; bands in the supernatant (Super) lanes indicate presence of protein during the incubation period, but no binding interaction.

Further reinforcing the importance of this result, recent work from the Kim Lab has

challenged a number of previously employed DGCR8-centric Microprocessing models.

Their work suggested that Microprocessor assembly is non- substrate dependent and the

minimal Microprocessor components needed to reconstitute function *in vitro* is the C-

terminal region of Drosha, from the CED to the native C- terminus, and the 73- residue

C- terminal tail of DGCR8 (G2). While no function has been defined for the G2 construct, the identification of a precise, obligate region on both Microprocessor components is exceedingly powerful. Future experiments will need to challenge what roles the G2 region of DGCR8 and Drosha's dsRBD, in conjunction or otherwise, play in Microprocessing and Microprocessor assembly. A number of hypothesis can be realized based on these results, including one in which the G2 peptide activates Drosha's dsRBD for substrate binding or perhaps another in which a G2- Drosha dsRBD complex is the unit responsible for recognizing the basal ss-ds junction of the pri-miRNA substrate. We attempted a preliminary pulldown assay to see if the G2 and Drosha dsRBD constructs (and KR- helix motif mutants) interact *in vitro*, and our initial results suggest that there is not interaction (Figure 17). However, this may be an *in vitro* artifact and this potential interaction should be studied more thoroughly.

Chapter 3: Future Directions

I. Introduction

While studying the cooperatively folded domains of the major Microprocessor components is pertinent, recent work by the Kim Group focusing on the C- terminal tail of DGCR8 suggests that this is the only DGCR8 region necessary, in conjunction with Drosha, to support efficient Microprocessing *in vitro*[6]. While this observation provided novel insight into the miRNA maturation pathway, the notion that intrinsically disordered regions (IDRs) can play fundamental roles in biological processes is well established[57]. For the past several years, the Showalter lab has focused on studying IDR systems with interesting functional roles and structural propensities[58, 59]. These studies include the important pancreatic transcription factor PDX1, in which the majority of point mutations leading to phenotypic diabetes exist in the N- terminal and C- terminal IDRs and not the Homeodomain; RNA Polymerase II's C- terminal domain, the phosphorylation state and factor loading of which modulates function of the entire complex[60]; and FCP1, a factor that is partially helical in the unbound state and undergoes a disorder-to-order transition upon binding its binding partner, RAP74[59, 61].

In an effort to investigate if the C- terminal regions of Drosha or DGCR8 possessed any innate helicity, the helical propensity calculation program, Agadir, was used[62] (Figure 18). Inspection of the helical propensity plots reveals that both Microprocessor components possess C- termini with large runs of residues with high helical propensity. Along with the C- terminal IDRs of each protein, the dsRBDs adjacent to these IDRs were similarly plotted. In these plots, the two helices in each dsRBD are not predicted to

possess high helicity; this is a result of Agadir predicting helicity using sequence only, having no knowledge of the through-space interactions that form tertiary structures[62]. Thus, these observations may be physiologically interesting in that they may play some role in a protein: protein binding event as many analogous systems do[61]. The following experiments proposed are in an effort to probe the structure and function of these interesting C- termini.
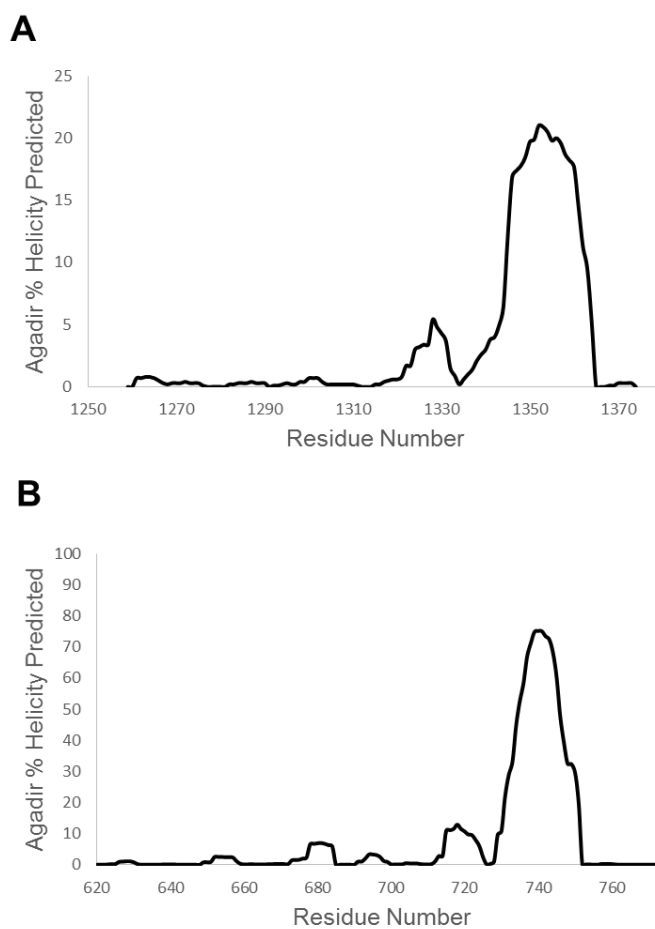
Figure 18: Calculated helical propensity for (A) the C- terminus of Drosha and (B) the C-terminus of DGCR8. Note the scale adjustment between the two plots.

II. Preliminary Results and Proposed Experiments for Drosha+

C- terminal to Drosha's lone dsRBD is a 37- residue IDR. This region, which is commonly left out of *in vitro* constructs, has never been structurally or functionally characterized[22, 32, 63, 64]. The first step taken in analyzing this domain was designing and subcloning a construct that contained Drosha's dsRBD fused to its native C- terminus, referred to throughout as Drosha+. Strikingly, when titrated into canonical Watson-Crick duplex the Drosha+ construct binds dsRNA, albeit with ten-fold weaker binding affinity than a canonical dsRBD (Figure 19). It is still unclear how this new binding function is achieved. For example, it is not clear from these experiments whether the IDR somehow imparts new function to the non-canonical dsRBD, or if the highly basic C- terminus can bind dsRNA in isolation. Nevertheless, these are important results that need to be further explored.

One important experiment to perform in the near future is to generate a construct that contains only the native C- terminus and test its dsRNA binding competence. This will assess whether the interesting functionality is imparted by the native C- terminus alone or in conjunction with Drosha's dsRBD. It is important to note that if this 37- residue IDR formed a helix, it will generate a very positively charged helical face. Depending on the result of this experiment, it will be important to investigate the structure of the C-terminus in isolation or the Drosha+ construct. Preliminary NMR experiments have been performed on the Drosha+ construct and the results are quite exciting (Figure 20).
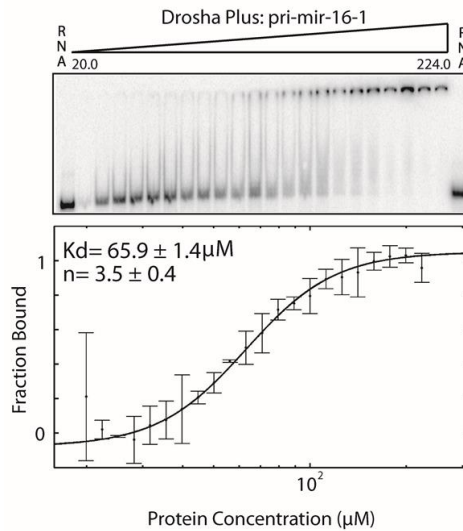
Figure 19: Representative EMSA for the Drosha+ construct binding to model pri-miRNA-16-1. RNA only lanes are indicated with 'RNA' and the calculated dissociation constant and Hill coefficient are annotated in the fit to the general Hill equation.
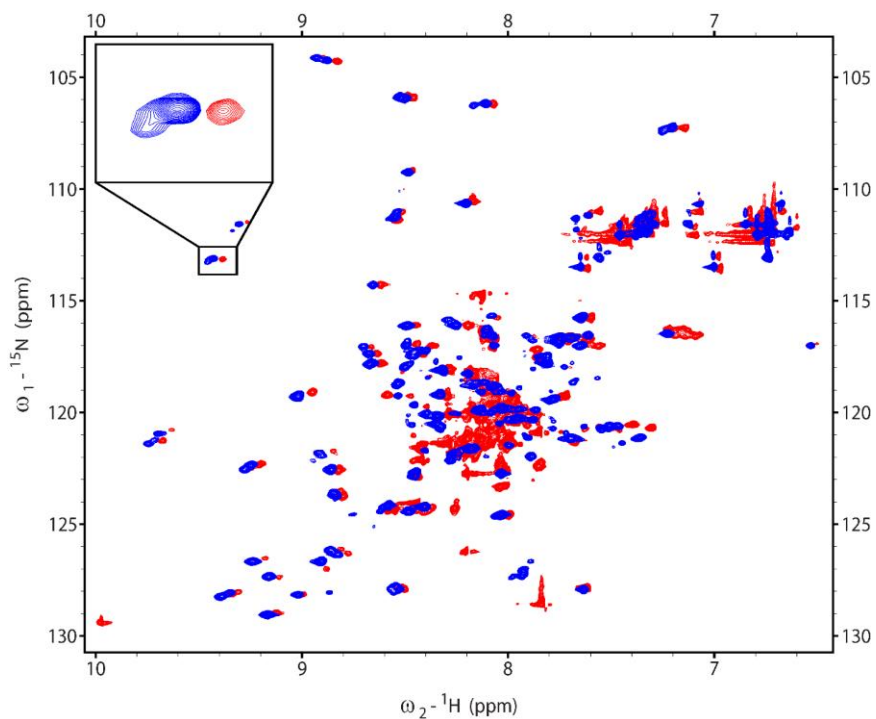
Figure 20: HSQC overlay of Drosha dsRBD (Blue) and Drosha+ (Red) constructs. The inset in the top left corner highlights the signal-multiplicity disparity between the constructs.

There are a few key observations that can be made from inspecting this HSQC overlay of Drosha+ and Drosha dsRBD. First, most of the residues contained in the Drosha+ and Drosha dsRBD have overlapping NMR signals- indicative that the dsRBD fold itself is not affected by the presence of the C- terminus. Next, the majority of signals pertaining to the native C- terminus of Drosha+ are clustered in the middle of the HSQC spectrum, thus indicating that this region is probably an IDR[41]. Lastly and most intriguingly, we observed signal doubling of multiple residues in the Drosha dsRBD construct[20]; in the Drosha+ construct, this phenomenon largely disappears. The presence of two signals per residue indicates that the Drosha dsRBD construct is in a state of slow conformational exchange. The fact that this does not exist in the Drosha+ case indicates that the presence of the C- terminal tail imparts some change to the dynamics of the dsRBD, leading to the hypothesis that this extension may interact stably with the dsRBD of Drosha.

To test this hypothesis, it is of utmost importance to assign the HSQC of Drosha+. In this effort, carbon detect NMR experiments will be used as there is minimal signal dispersion in the HSQC that will only be further exacerbated in the three dimensional proton detect experiments. The carbon detect experiments provide better signal dispersion as the carbonyl carbon signals are more dispersed than the amide proton signals in IDRs[41]. Once these residues are assigned, the same relaxation experiments performed on the Drosha dsRBD and Drosha-Quad constructs will be performed on Drosha+ and thereafter compared to the Drosha dsRBD case. Understanding where and how the presence of the C- terminus affects the dsRBD will influence a number of other

experiments- including solving the Drosha+ structure via the same methods utilized in the Drosha-Quad case.

III. Preliminary Results and Proposed Experiments of DGCR8 G2

DGCR8's C- terminus has recently been implicated in Microprocessing via work from Narry Kim's group[6]. From their analysis, they gathered that the only component of DGCR8 that was necessary, in conjunction with Drosha, for effective Microprocessor reconstitution *in vitro* was the region C- terminal to the adjacent dsRBDs. This striking functional result, paired with the Agadir prediction that this region has a near 90% helical propensity, lends this system to a myriad of important experiments. Thus far, I have subcloned a G2 construct for recombinant expression in E. coli and have successfully purified the construct and obtained an HSQC spectrum (Figure 21). Interestingly, this HSQC spectrum, via the relatively high level of signal dispersion and shift of signals up-field, suggests that the Agadir prediction is likely correct and much of this region is indeed helical. It is quite possible that this helix imparts the implicated function of this C- terminus in Microprocessing.
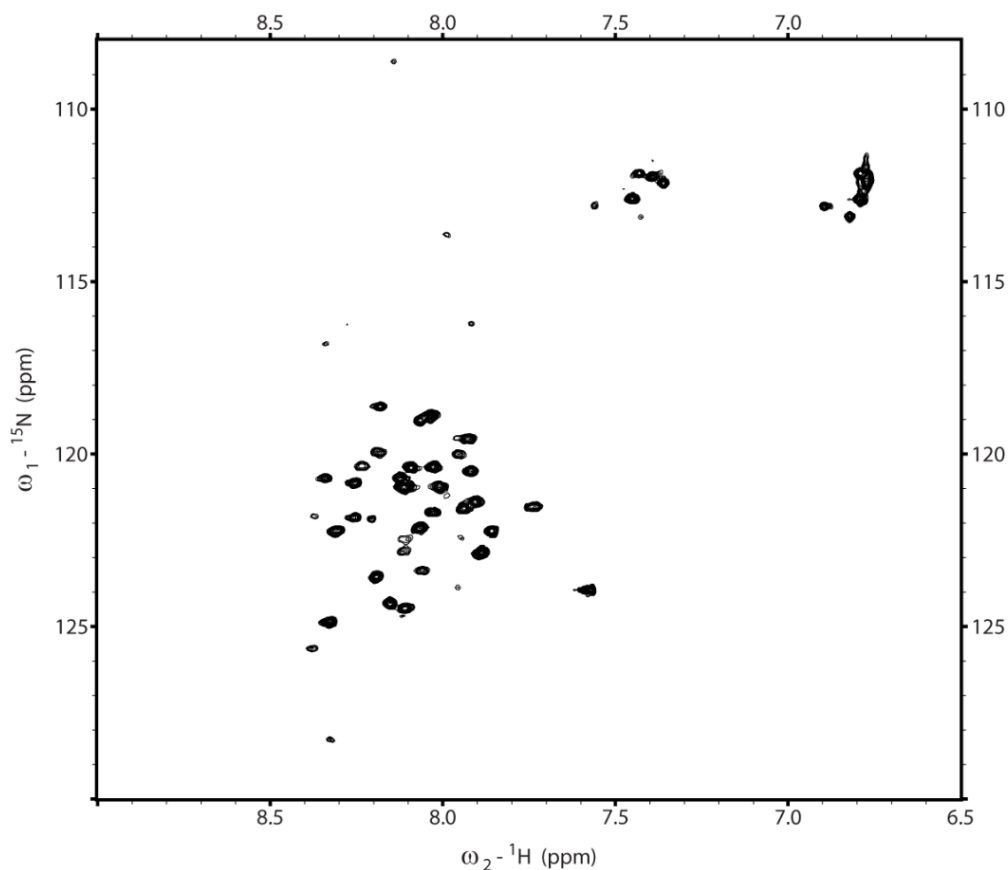
Figure 21: Representative HSQC of the G2 construct.

Going forward, it is of utmost importance to understand DGCR8 G2's role in Microprocessing. The first round of experiments I propose will attempt to ascertain the complementary Drosha interaction region via a series of deletion and mutation studies. Proceeding experiments will probe this interaction more deeply via point mutations targeted at breaking the G2-Drosha interaction as a means to better understand the interaction. Finally, using x-ray crystallographic techniques this complex should be structurally annotated. All of these experiments will challenge the hypothesis depicted (Figure 22). We hypothesize that Drosha forms a complex with two DGCR8s- an

interaction which is mediated by the Drosha+ region and DGCR8s' C- terminal tails. This Microprocessor complex formation is followed by substrate selection and subsequent Microprocessing. This hypothesis differs from preceding hypotheses[7, 32, 63, 64] in that: (1) Microprocessor formation is dsRNA independent and (2) functional roles for Drosha's dsRBD and C- terminal IDR are proposed.
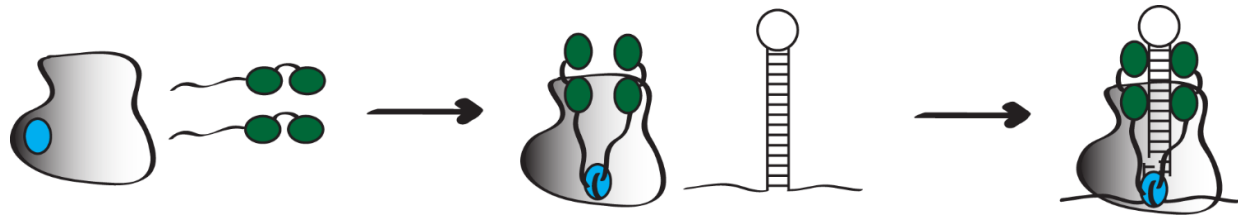
Figure 22: Cartoon diagram depicting our new model of Microprocessor assembly and Microprocessing. Drosha is depicted in grey-scale, its dsRBD in light blue and DGCR8's dsRBDs in green.

References

1.   Fabian, M. R.; Sonenberg, N.; Filipowicz, W., Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **2010,** *79*, 351-79.

2.   Sassen, S.; Miska, E. A.; Caldas, C., MicroRNA—implications for cancer. *Virchows Archiv* **2008,** *452* (1), 1-10.

3.   Zhang, C., MicroRNomics: a newly emerging approach for disease biology. *Physiological genomics* **2008,** *33* (2), 139-47.

4.   Maes, O. C.; Chertkow, H. M.; Wang, E.; Schipper, H. M., MicroRNA: Implications for Alzheimer Disease and other Human CNS Disorders. *Current Genomics* **2009,** *10* (3), 154-68.

5.   Ma, F.; Xu, S.; Liu, X.; Zhang, Q.; Xu, X.; Liu, M.; Hua, M.; Li, N.; Yao, H.; Cao, X., The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-gamma. *Nat Immunol* **2011,** *12* (9), 861-9.

6.   Nguyen, Tuan A.; Jo, Myung H.; Choi, Y.-G.; Park, J.; Kwon, S. C.; Hohng, S.; Kim, V. N.; Woo, J.-S., Functional Anatomy of the Human Microprocessor. *Cell* **2015,** *161* (6), 1374-1387.

7.   Quick-Cleveland, J.; Jacob, Jose P.; Weitz, Sara H.; Shoffner, G.; Senturia, R.; Guo, F., The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin. *Cell Reports* **2014,** *7* (6), 1994-2005.

8.   Zeng, Y.; Cullen, B. R., Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* **2004,** *32* (16), 4776-85.

9.   Lee, H. Y.; Doudna, J. A., TRBP alters human precursor microRNA processing in vitro. *RNA* **2012,** *18* (11), 2012-9.

10.   Khvorova, A.; Reynolds, A.; Jayasena, S. D., Functional siRNAs and miRNAs exhibit strand bias. *Cell* **2003,** *115* (2), 209-16.

11.   Matranga, C.; Tomari, Y.; Shin, C.; Bartel, D. P.; Zamore, P. D., Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **2005,** *123* (4), 607-20.

12.   Liu, J.; Carmell, M. A.; Rivas, F. V.; Marsden, C. G.; Thomson, J. M.; Song, J.-J.; Hammond, S. M.; Joshua-Tor, L.; Hannon, G. J., Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science* **2004,** *305* (5689), 1437-1441.

13.   Takeshita, D.; Zenno, S.; Lee, W. C.; Nagata, K.; Saigo, K.; Tanokura, M., Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human dicer. *J Mol Biol* **2007,** *374* (1), 106-20.

14.   Yeom, K.-H.; Lee, Y.; Han, J.; Suh, M. R.; Kim, V. N., Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Research* **2006,** *34* (16), 4622-4629.

15.   Macrae, I. J.; Zhou, K.; Li, F.; Repic, A.; Brooks, A. N.; Cande, W. Z.; Adams, P. D.; Doudna, J. A., Structural basis for double-stranded RNA processing by Dicer. *Science* **2006,** *311* (5758), 195-8.

16.   Ma, E.; Zhou, K.; Kidwell, M. A.; Doudna, J. A., Coordinated activities of human dicer domains in regulatory RNA processing. *J Mol Biol* **2012,** *422* (4), 466-76.

17.   Ian, J. M.; Jennifer, A. D., Ribonuclease revisited: structural insights into ribonuclease \{III\} family enzymes. *Current Opinion in Structural Biology* **2007,** *17* (1), 138 - 145.

18.   Yamashita, S.; Nagata, T.; Kawazoe, M.; Takemoto, C.; Kigawa, T.; Güntert, P.; Kobayashi, N.; Terada, T.; Shirouzu, M.; Wakiyama, M.; Muto, Y.; Yokoyama, S., Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Science : A Publication of the Protein Society* **2011,** *20* (1), 118-30.

19.   Wilson, R. C.; Tambe, A.; Kidwell, M. A.; Noland, C. L.; Schneider, C. P.; Doudna, J. A., Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Mol Cell* **2015,** *57* (3), 397-407.

20.   Wostenberg, C.; Quarles, K. A.; Showalter, S. A., Dynamic Origins of Differential RNA Binding Function in Two dsRBDs from the miRNA "Microprocessor" Complex. *Biochemistry* **2010,** *49* (50), 10728-10736.

21.   St Johnston, D.; Brown, N. H.; Gall, J. G.; Jantsch, M., A conserved double-stranded RNA-binding domain. *Proc Natl Acad Sci U S A* **1992,** *89* (22), 10979-83.

22.   Mueller, G.; Miller, M.; DeRose, E.; Ghosh, M.; London, R.; Hall, T., Solution structure of the Drosha double-stranded RNA-binding domain. *Silence* **2010,** *1* (1), 2.

23.   Leulliot, N.; Quevillon-Cheruel, S.; Graille, M.; van Tilbeurgh, H.; Leeper, T. C.; Godin, K. S.; Edwards, T. E.; Sigurdsson, S. T.; Rozenkrants, N.; Nagel, R. J.; Ares, M.; Varani, G., A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III. *EMBO J* **2004,** *23* (13), 2468-77.

24.   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. *Proc Natl Acad Sci U S A* **2004,** *101* (22), 8307-8312.

25.   Masliah, G.; Barraud, P.; Allain Féé, H. T., RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci* **2013,** *70* (11), 1875-95.

26.   Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B., The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol* **2004,** *5* (12), 1013-1023.

27.   Ryter, J. M.; Schultz, S. C., Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* **1998,** *17* (24), 7505-13.

28.   Wostenberg, C.; Lary, J. W.; Sahu, D.; Acevedo, R.; Quarles, K. A.; Cole, J. L.; Showalter, S. A., The Role of Human Dicer-dsRBD in Processing Small Regulatory RNAs. *PLoS ONE* **2012,** *7* (12), e51829.

29.   Quarles, K. A.; Chadalavada, D.; Showalter, S. A., Deformability in the cleavage site of primary microRNA is not sensed by the double-stranded RNA binding domains in the microprocessor component DGCR8. *Proteins* **2015,** *83* (6), 1165-79.

30.   Acevedo, R.; Orench-Rivera, N.; Quarles, K. A.; Showalter, S. A., Helical Defects in MicroRNA Influence Protein Binding by TAR RNA Binding Protein. *PLoS ONE* **2015,** *10* (1), e0116749.

31.   Parker, G. S.; Maity, T. S.; Bass, B. L., dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *J Mol Biol* **2008,** *384* (4), 967-79.

32.   Han, J.; Lee, Y.; Yeom, K. H.; Kim, Y. K.; Jin, H.; Kim, V. N., The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **2004,** *18* (24), 3016-27.

33.   Gregory, R. I.; Yan, K. P.; Amuthan, G.; Chendrimada, T.; Doratotaj, B.; Cooch, N.; Shiekhattar, R., The Microprocessor complex mediates the genesis of microRNAs. *Nature* **2004,** *432* (7014), 235-40.

34.   Sohn, S. Y.; Bae, W. J.; Kim, J. J.; Yeom, K. H.; Kim, V. N.; Cho, Y., Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* **2007,** *14* (9), 847-53.

35.   Quarles, K. A.; Sahu, D.; Havens, M. A.; Forsyth, E. R.; Wostenberg, C.; Hastings, M. L.; Showalter, S. A., Ensemble analysis of primary microRNA structure reveals an extensive capacity to deform near the Drosha cleavage site. *Biochemistry* **2013,** *52* (5), 795-807.

36.   Zeng, Y.; Yi, R.; Cullen, B. R., Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **2005,** *24* (1), 138-48.

37.   Chen, C. Z.; Li, L.; Lodish, H. F.; Bartel, D. P., MicroRNAs modulate hematopoietic lineage differentiation. *Science* **2004,** *303* (5654), 83-6.

38.   Garner, M. M.; Revzin, A., A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res* **1981,** *9* (13), 3047-60.

39.   Bodenhausen, G.; Ruben, D. J., Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **1980,** *69* (1), 185-189.

40.   Morris, G. A.; Freeman, R., Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **1979,** *101* (3), 760-762.

41.   Sahu, D.; Bastidas, M.; Showalter, S. A., Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Anal Biochem* **2014,** *449*, 17-25.

42.   Bastidas, M.; Gibbs, E. B.; Sahu, D.; Showalter, S. A., A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts in Magnetic Resonance Part A* **2015,** *44* (1), 54-66.

43.   Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E., Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry* **1994,** *33* (19), 5984-6003.

44.   Akke, M.; Brueschweiler, R.; Palmer, A. G., NMR order parameters and free energy: an analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k. *Journal of the American Chemical Society* **1993,** *115* (21), 9832-9833.

45.   Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A., Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **2011,** *213* (2), 423-41.

46.   Clubb, R. T.; Thanabal, V.; Wagner, G., A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N☐13C-labelled proteins. *Journal of Magnetic Resonance (1969)* **1992,** *97* (1), 213-217.

47.   Grzesiek, S.; Bax, A., Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **1992,** *114* (16), 6291-6293.

48.   Grzesiek, S.; Bax, A., An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **1992,** *99* (1), 201-207.

49.   Shen, Y.; Bax, A., Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR* **2013,** *56* (3), 227-41.

50.   Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A., TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **2009,** *44* (4), 213-23.

51.   Ishima, R.; Torchia, D. A., Protein dynamics from NMR. *Nat Struct Biol* **2000,** *7* (9), 740-3.

52.   Kay, L. E., Protein dynamics from NMR. *Nat Struct Biol* **1998,** *5 Suppl*, 513-7.

53.   Tjandra, N.; Szabo, A.; Bax, A., Protein Backbone Dynamics and 15N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects. *Journal of the American Chemical Society* **1996,** *118* (29), 6986-6991.

54.   Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **2003,** *160* (1), 65-73.

55.   Thomas, P.; Smart, T. G., HEK293 cell line: a vehicle for the expression of recombinant proteins. *Journal of pharmacological and toxicological methods* **2005,** *51* (3), 187-200.

56.   Felgner, P. L.; Gadek, T. R.; Holm, M.; Roman, R.; Chan, H. W.; Wenz, M.; Northrop, J. P.; Ringold, G. M.; Danielsen, M., Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc Natl Acad Sci U S A* **1987,** *84* (21), 7413-7.

57.   Tompa, P., Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **2012,** *37* (12), 509-16.

58.   Bastidas, M.; Showalter, S. A., Thermodynamic and structural determinants of differential Pdx1 binding to elements from the insulin and IAPP promoters. *J Mol Biol* **2013,** *425* (18), 3360-77.

59.   Lawrence, C. W.; Bonny, A.; Showalter, S. A., The disordered C-terminus of the RNA polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochem Biophys Res Commun* **2011,** *410* (3), 461-5.

60.   Phatnani, H. P.; Greenleaf, A. L., Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **2006,** *20* (21), 2922-2936.

61.   Kamada, K.; De; Roeder, R. G.; Burley, S. K., Crystal structure of the C-terminal domain of the RAP74 subunit. *Proc Natl Acad Sci U S A* **2001,** *98* (6), 3115-20.

62.   Lacroix, E.; Viguera, A. R.; Serrano, L., Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J Mol Biol* **1998,** *284* (1), 173-91.

63.   Han, J.; Lee, Y.; Yeom, K.-H.; Nam, J.-W.; Heo, I.; Rhee, J.-K.; Sohn, S. Y.; Cho, Y.; Zhang, B.-T.; Kim, V. N., Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* **2006,** *125* (5), 887-901.

64.   Han, J.; Pedersen, J. S.; Kwon, S. C.; Belair, C. D.; Kim, Y.-K.; Yeom, K.-H.; Yang, W.-Y.; Haussler, D.; Blelloch, R.; Kim, V. N., Posttranscriptional Crossregulation between Drosha and DGCR8. *Cell* **2009,** *136* (1), 75-84.

Appendix: Drosha-Quad Chemical Shift Data

Atom ID	Residue ID	Amino Acid	Atom ID	Atom Type	Isotope ID	Chemical Shift (ppm)	Chemical Shift Error (ppm)
1	6	PRO	C	C	13	174.849	0.001
2	6	PRO	CA	C	13	62.494	0.116
3	6	PRO	CB	C	13	30.061	0.177
4	6	PRO	CD	C	13	48.165	0
5	6	PRO	CG	C	13	25.389	0
6	6	PRO	HA	H	1	4.233	0.015
7	6	PRO	HB	H	1	2.224	0.001
8	6	PRO	HD	H	1	3.89	0
9	6	PRO	HG3	H	1	1.827	0
10	6	PRO	HG2	H	1	2.021	0.018
11	7	ILE	C	C	13	177.294	0.006
12	7	ILE	CA	C	13	62.546	0.118
13	7	ILE	CB	C	13	34.984	0.166
14	7	ILE	CD1	C	13	11.591	0
15	7	ILE	CG1	C	13	26.844	0
16	7	ILE	CG2	C	13	14.195	0
17	7	ILE	HA	H	1	3.556	0.007
18	7	ILE	HB	H	1	1.959	0.017
19	7	ILE	HD11	H	1	0.599	0.007
20	7	ILE	HG12	H	1	1.459	0.025
21	7	ILE	HG21	H	1	0.78	0.006
22	7	ILE	HN	H	1	7.42	0.007
23	7	ILE	N	N	15	118.341	0.04
24	8	SER	C	C	13	174.216	0.006
25	8	SER	CA	C	13	58.686	0.142
26	8	SER	CB	C	13	59.913	0.435
27	8	SER	HA	H	1	4.14	0.009
28	8	SER	HB	H	1	3.796	0.022
29	8	SER	HN	H	1	7.867	0.005
30	8	SER	N	N	15	118.043	0.043
31	9	GLN	C	C	13	175.75	0.013
32	9	GLN	CA	C	13	56.418	0.113
33	9	GLN	CB	C	13	26.763	1.306
34	9	GLN	CG	C	13	31.217	0
35	9	GLN	HA	H	1	3.894	0.001
36	9	GLN	HB	H	1	1.934	0.016
37	9	GLN	HG	H	1	2.25	0.006

38	9	GLN	HN	H	1	8.384	0.004
39	9	GLN	N	N	15	121.528	0.099
40	10	LEU	C	C	13	174.336	0.006
41	10	LEU	CA	C	13	55.816	0.074
42	10	LEU	CB	C	13	38.168	0.068
43	10	LEU	CD1	C	13	22.236	0
44	10	LEU	CD2	C	13	21.164	0
45	10	LEU	CG	C	13	24.05	0
46	10	LEU	HA	H	1	3.903	0.012
47	10	LEU	HB	H	1	1.53	0
48	10	LEU	HD11	H	1	0.806	0.01
49	10	LEU	HN	H	1	8.357	0.005
50	10	LEU	N	N	15	120.026	0.039
51	11	GLN	C	C	13	175.196	0.009
52	11	GLN	CA	C	13	57.729	0.169
53	11	GLN	CB	C	13	20.954	9.335
54	11	GLN	CG	C	13	31.594	0
55	11	GLN	HA	H	1	3.591	0.004
56	11	GLN	HN	H	1	7.445	0.004
57	11	GLN	N	N	15	116.55	0.053
58	12	GLN	C	C	13	176.463	0.002
59	12	GLN	CA	C	13	56.111	0.093
60	12	GLN	CB	C	13	25.406	0.057
61	12	GLN	CG	C	13	31.359	0
62	12	GLN	HA	H	1	3.94	0
63	12	GLN	HB	H	1	2.123	0.057
64	12	GLN	HG	H	1	2.388	0.049
65	12	GLN	HN	H	1	8.159	0.005
66	12	GLN	N	N	15	116.948	0.055
67	13	CYS	C	C	13	174.884	0.004
68	13	CYS	CA	C	13	55.688	0.15
69	13	CYS	CB	C	13	36.819	0.061
70	13	CYS	HA	H	1	4.246	0.035
71	13	CYS	HB	H	1	2.939	0.011
72	13	CYS	HN	H	1	8.255	0.003
73	13	CYS	N	N	15	119.104	0.036
74	14	CYS	C	C	13	174.825	0.126
75	14	CYS	CA	C	13	59.649	0.13
76	14	CYS	CB	C	13	30.799	0.143
77	14	CYS	HA	H	1	3.872	0.001
78	14	CYS	HB	H	1	2.863	0.003
79	14	CYS	HN	H	1	8.472	0.005
80	14	CYS	N	N	15	119.738	0.032

82	15	LEU	C	C	13	176.334	0.003
83	15	LEU	CA	C	13	55.008	0.172
84	15	LEU	CB	C	13	39.276	0.206
85	15	LEU	CD1	C	13	22.508	0.039
86	15	LEU	CD2	C	13	21.074	0
87	15	LEU	CG	C	13	24.636	0
88	15	LEU	HA	H	1	4.164	0.019
89	15	LEU	HB	H	1	1.814	0.018
90	15	LEU	HD11	H	1	0.839	0.028
91	15	LEU	HG	H	1	1.609	0.01
92	15	LEU	HN	H	1	7.875	0
93	15	LEU	N	N	15	120.206	0.026
94	16	THR	C	C	13	172.543	0.003
95	16	THR	CA	C	13	61.556	0.162
96	16	THR	CB	C	13	67.074	0.253
97	16	THR	CG	C	13	19.069	0
98	16	THR	HA	H	1	4.188	0.012
99	16	THR	HB	H	1	3.92	0
100	16	THR	HG	H	1	1.32	0.006
101	16	THR	HN	H	1	7.636	0.004
102	16	THR	N	N	15	111.015	0.032
103	17	LEU	C	C	13	174.215	0.003
104	17	LEU	CA	C	13	52.269	0.113
105	17	LEU	CB	C	13	39.018	0.141
106	17	LEU	CD1	C	13	20.678	0
107	17	LEU	CD2	C	13	18.833	0
108	17	LEU	CG	C	13	23.628	0
109	17	LEU	HA	H	1	4.366	0.008
110	17	LEU	HB2	H	1	1.744	0.002
111	17	LEU	HD11	H	1	0.898	0
112	17	LEU	HG	H	1	1.502	0.197
113	17	LEU	HN	H	1	7.364	0.005
114	17	LEU	N	N	15	121.097	0.026
115	18	ARG	C	C	13	173.362	0.004
116	18	ARG	CA	C	13	53.709	0.069
117	18	ARG	CB	C	13	28.038	0.103
118	18	ARG	CD	C	13	41.199	0
119	18	ARG	CG	C	13	26.052	0
120	18	ARG	HA	H	1	4.371	0.013
121	18	ARG	HB2	H	1	1.851	0.002
122	18	ARG	HD2	H	1	3.104	0.022
123	18	ARG	HG2	H	1	1.588	0.008
124	18	ARG	HN	H	1	7.555	0.016

125	18	ARG	N	N	15	120.709	0.03
126	19	THR	C	C	13	171.58	0.005
127	19	THR	CA	C	13	58.936	0.195
128	19	THR	CB	C	13	67.672	0.23
129	19	THR	CG	C	13	18.956	0
130	19	THR	HA	H	1	4.387	0.005
131	19	THR	HG	H	1	1.144	0.023
132	19	THR	HN	H	1	8.54	0.004
133	19	THR	N	N	15	118.668	0.03
134	20	GLU	C	C	13	174.446	0.011
135	20	GLU	CA	C	13	54.949	0.134
136	20	GLU	CB	C	13	26.872	0.03
137	20	GLU	CG	C	13	33.61	0
138	20	GLU	HA	H	1	4.075	0.012
139	20	GLU	HG	H	1	2.196	0
140	20	GLU	HN	H	1	8.586	0.004
141	20	GLU	N	N	15	124.039	0.038
142	21	GLY	C	C	13	171.105	0.001
143	21	GLY	CA	C	13	42.662	0.129
144	21	GLY	HA2	H	1	3.733	0
145	21	GLY	HA3	H	1	4.046	0.007
146	21	GLY	HN	H	1	8.552	0.005
147	21	GLY	N	N	15	111.169	0.037
148	22	LYS	C	C	13	173.525	0.004
149	22	LYS	CA	C	13	52.416	0.109
150	22	LYS	CB	C	13	31.336	0.058
151	22	LYS	CD	C	13	25.839	0
152	22	LYS	CE	C	13	39.669	0
153	22	LYS	CG	C	13	21.849	0
154	22	LYS	HA	H	1	4.453	0.012
155	22	LYS	HB	H	1	1.733	0
156	22	LYS	HD	H	1	1.599	0
157	22	LYS	HE	H	1	2.93	0
158	22	LYS	HG	H	1	1.316	0
159	22	LYS	HN	H	1	7.798	0.004
160	22	LYS	N	N	15	119.376	0.048
161	23	GLU	C	C	13	172.069	0
162	23	GLU	CA	C	13	51.739	0
163	23	GLU	CB	C	13	26.486	0
164	23	GLU	N	N	15	122.631	0.052
165	24	PRO	C	C	13	173.547	0.003
166	24	PRO	CA	C	13	60.521	0.199
167	24	PRO	CB	C	13	29.522	0.172

168	24	PRO	CD	C	13	48.156	0
169	24	PRO	CG	C	13	24.731	0
170	24	PRO	HA	H	1	4.402	0.007
171	24	PRO	HB2	H	1	1.871	0.01
172	24	PRO	HB3	H	1	2.042	0
173	24	PRO	HD	H	1	3.688	0
174	25	ASP	C	C	13	171.891	0.006
175	25	ASP	CA	C	13	53.718	3.286
176	25	ASP	CB	C	13	40.677	0.211
177	25	ASP	HA	H	1	4.657	0.002
178	25	ASP	HB3	H	1	2.638	0
179	25	ASP	HB2	H	1	2.287	0.01
180	25	ASP	HN	H	1	8.232	0.004
181	25	ASP	N	N	15	121.674	0.026
182	26	ILE	C	C	13	171.291	0
183	26	ILE	CA	C	13	55.315	0
184	26	ILE	CB	C	13	37.365	0
185	26	ILE	HN	H	1	8.108	0.002
186	26	ILE	N	N	15	116.402	0.032
187	27	PRO	C	C	13	172.234	0.003
188	27	PRO	CA	C	13	60.676	0.194
189	27	PRO	CB	C	13	29.51	0.162
190	27	PRO	CD	C	13	48.353	0
191	27	PRO	CG	C	13	26.057	0
192	27	PRO	HA	H	1	4.723	0.001
193	27	PRO	HB2	H	1	1.999	0
194	27	PRO	HD2	H	1	3.688	0
195	27	PRO	HG2	H	1	1.755	0
196	28	LEU	C	C	13	172.463	0.004
197	28	LEU	CA	C	13	50.999	0.037
198	28	LEU	CB	C	13	42.433	0.126
199	28	LEU	CD1	C	13	21.813	0
200	28	LEU	CD2	C	13	21.011	0
201	28	LEU	CG	C	13	24.366	0
202	28	LEU	HA	H	1	4.712	0.009
203	28	LEU	HB2	H	1	1.564	0.017
204	28	LEU	HD11	H	1	0.818	0.01
205	28	LEU	HG	H	1	1.596	0.008
206	28	LEU	HN	H	1	8.477	0.006
207	28	LEU	N	N	15	124.504	0.035
208	29	TYR	C	C	13	173.614	0.003
209	29	TYR	CA	C	13	54.959	0.122
210	29	TYR	CB	C	13	37.328	0.101

211	29	TYR	HA	H	1	5.285	0.016
212	29	TYR	HB3	H	1	3.471	0
213	29	TYR	HB2	H	1	2.797	0.008
214	29	TYR	HN	H	1	8.907	0.004
215	29	TYR	N	N	15	123.137	0.04
216	30	LYS	C	C	13	172.5	0
217	30	LYS	CA	C	13	52.399	0
218	30	LYS	CB	C	13	34.086	0
219	30	LYS	HN	H	1	9.344	0.021
220	30	LYS	N	N	15	122.426	0.037
221	31	THR	CA	C	13	60.787	0.01
222	31	THR	CG	C	13	20.082	0
223	31	THR	HA	H	1	4.51	0.006
224	31	THR	HB	H	1	3.911	0.013
225	31	THR	HG1	H	1	1.22	0.02
226	32	LEU	C	C	13	174.368	0.005
227	32	LEU	CA	C	13	52.683	0.128
228	32	LEU	CB	C	13	40.207	0.063
229	32	LEU	CD	C	13	19.864	0
230	32	LEU	CG	C	13	23.197	0
231	32	LEU	HA	H	1	4.44	0.016
232	32	LEU	HB2	H	1	1.637	0.006
233	32	LEU	HD11	H	1	0.883	0.021
234	32	LEU	HG	H	1	1.508	0.008
235	32	LEU	HN	H	1	9.019	0.007
236	32	LEU	N	N	15	129.198	0.049
237	33	GLN	C	C	13	171.615	0.02
238	33	GLN	CA	C	13	53.264	0.158
239	33	GLN	CB	C	13	28.844	0.114
240	33	GLN	CG	C	13	31.205	0
241	33	GLN	HG	H	1	2.228	0
242	33	GLN	HN	H	1	7.689	0.01
243	33	GLN	N	N	15	115.992	0.039
244	34	THR	C	C	13	170.622	0.013
245	34	THR	CA	C	13	59.144	0.069
246	34	THR	CB	C	13	67.805	0.131
247	34	THR	CG2	C	13	18.915	0
248	34	THR	HA	H	1	4.489	0.008
249	34	THR	HG1	H	1	0.597	0.002
250	34	THR	HN	H	1	8.345	0.006
251	34	THR	N	N	15	119.608	0.027
252	35	VAL	C	C	13	172.242	0.007
253	35	VAL	CA	C	13	57.697	0.136

254	35	VAL	CB	C	13	32.804	0.125
255	35	VAL	CG1	C	13	17.159	0
256	35	VAL	CG2	C	13	18.595	0
257	35	VAL	HA	H	1	4.492	0.003
258	35	VAL	HB	H	1	1.934	0.011
259	35	VAL	HG11	H	1	0.754	0.006
260	35	VAL	HN	H	1	8.134	0.004
261	35	VAL	N	N	15	120.795	0.022
262	36	GLY	C	C	13	169.453	0
263	36	GLY	CA	C	13	41.094	0
264	36	GLY	HN	H	1	8.229	0.005
265	36	GLY	N	N	15	110.608	0.042
266	39	HIS	C	C	13	172.14	0
267	39	HIS	CA	C	13	53.584	0.219
268	39	HIS	CB	C	13	27.736	0.061
269	39	HIS	HA	H	1	4.631	0.041
270	39	HIS	HB2	H	1	3.3	0
271	39	HIS	HB3	H	1	2.885	0
272	40	ALA	C	C	13	171.693	0.001
273	40	ALA	CA	C	13	48.979	0.006
274	40	ALA	CB	C	13	15.335	0.075
275	40	ALA	HA	H	1	4.422	0.008
276	40	ALA	HB1	H	1	0.896	0.007
277	40	ALA	HN	H	1	7.582	0.003
278	40	ALA	N	N	15	128.155	0.031
280	41	ARG	C	C	13	174.236	0.007
281	41	ARG	CA	C	13	52.715	0.123
282	41	ARG	CB	C	13	29.506	0.168
283	41	ARG	CD	C	13	40.879	0
284	41	ARG	CG	C	13	23.465	0
285	41	ARG	HA	H	1	4.467	0.008
286	41	ARG	HB2	H	1	1.609	0.01
287	41	ARG	HD2	H	1	2.664	0.007
288	41	ARG	HG2	H	1	1.234	0.007
289	41	ARG	HN	H	1	8.09	0.007
290	41	ARG	N	N	15	122.997	0.033
291	42	THR	C	C	13	171.192	0.01
292	42	THR	CA	C	13	59.75	0.231
293	42	THR	CB	C	13	67.917	0.118
294	42	THR	CG	C	13	18.884	0
295	42	THR	HA	H	1	4.448	0.008
296	42	THR	HB	H	1	3.696	0.006
297	42	THR	HG1	H	1	1.02	0.073

298	42	THR	HN	H	1	8.572	0.008
299	42	THR	N	N	15	116.202	0.028
300	43	TYR	C	C	13	172.16	0.014
301	43	TYR	CA	C	13	54.921	0.114
302	43	TYR	CB	C	13	38.22	0.07
303	43	TYR	HA	H	1	4.719	0.023
304	43	TYR	HB2	H	1	2.492	0
305	43	TYR	HB3	H	1	2.593	0.021
306	43	TYR	HN	H	1	9.255	0.006
307	43	TYR	N	N	15	127.687	0.042
308	44	THR	C	C	13	171.33	0.014
309	44	THR	CA	C	13	59.299	0.082
310	44	THR	CB	C	13	67.619	0.138
311	44	THR	HA	H	1	5.162	0.004
312	44	THR	HG1	H	1	1.187	0.015
313	44	THR	HN	H	1	8.699	0.012
314	44	THR	N	N	15	117.178	0.043
315	45	VAL	C	C	13	169.635	0.005
316	45	VAL	CA	C	13	57.558	0.204
317	45	VAL	CB	C	13	32.829	0.095
318	45	VAL	CG1	C	13	18.666	0
319	45	VAL	CG2	C	13	20.235	0
320	45	VAL	HA	H	1	4.177	0
321	45	VAL	HB	H	1	1.527	0.004
322	45	VAL	HG11	H	1	0.905	0.013
323	45	VAL	HN	H	1	9.331	0.006
324	45	VAL	N	N	15	128.199	0.038
325	46	ALA	C	C	13	171.973	0.01
326	46	ALA	CA	C	13	46.865	0.014
327	46	ALA	CB	C	13	21.198	0.02
328	46	ALA	HA	H	1	5.26	0.013
329	46	ALA	HB1	H	1	1.332	0
330	46	ALA	HN	H	1	8.961	0.005
331	46	ALA	N	N	15	126.929	0.057
332	47	VAL	C	C	13	170.285	0.007
333	47	VAL	CA	C	13	55.384	0.171
334	47	VAL	CB	C	13	32.098	0.216
335	47	VAL	CG1	C	13	17.157	0
336	47	VAL	CG2	C	13	20.007	0
337	47	VAL	HA	H	1	4.961	0.019
338	47	VAL	HB	H	1	1.312	0.009
339	47	VAL	HG11	H	1	0.062	0.048
340	47	VAL	HG21	H	1	0.408	0.007

341	47	VAL	HN	H	1	8.248	0.006
342	47	VAL	N	N	15	118.914	0.041
343	48	TYR	C	C	13	172.504	0.004
344	48	TYR	CA	C	13	53.61	0.079
345	48	TYR	CB	C	13	40.339	0.203
346	48	TYR	HA	H	1	5.217	0.004
347	48	TYR	HB2	H	1	2.678	0.005
348	48	TYR	HN	H	1	9.274	0.008
349	48	TYR	N	N	15	126.815	0.03
350	49	PHE	C	C	13	172.39	0.011
351	49	PHE	CA	C	13	54.128	0.086
352	49	PHE	CB	C	13	40.24	0.189
353	49	PHE	HA	H	1	4.766	0.027
354	49	PHE	HB2	H	1	2.712	0.006
355	49	PHE	HB3	H	1	2.628	0
356	49	PHE	HN	H	1	9.024	0.007
357	49	PHE	N	N	15	119.234	0.018
358	50	LYS	C	C	13	173.659	0.004
359	50	LYS	CA	C	13	54.906	0.084
360	50	LYS	CB	C	13	26.726	0.146
361	50	LYS	CD	C	13	19.726	0
362	50	LYS	CE	C	13	39.384	0
363	50	LYS	CG	C	13	18.599	0
364	50	LYS	HA	H	1	3.502	0.015
365	50	LYS	HB2	H	1	1.336	0.012
366	50	LYS	HE2	H	1	2.759	0.026
367	50	LYS	HN	H	1	9.035	0.025
368	50	LYS	N	N	15	128.064	0.039
369	51	GLY	C	C	13	170.97	0.004
370	51	GLY	CA	C	13	42.553	0.082
371	51	GLY	HA2	H	1	4.086	0.007
372	51	GLY	HA3	H	1	4.335	0
373	51	GLY	HN	H	1	8.893	0.005
374	51	GLY	N	N	15	104.286	0.039
375	52	GLU	C	C	13	172.477	0.006
376	52	GLU	CA	C	13	51.709	0.204
377	52	GLU	CB	C	13	29.248	0.117
378	52	GLU	CG	C	13	31.954	0
379	52	GLU	HA	H	1	4.617	0.055
380	52	GLU	HB2	H	1	2.04	0.027
381	52	GLU	HG2	H	1	2.221	0
382	52	GLU	HN	H	1	7.71	0.005
383	52	GLU	N	N	15	121.087	0.041

384	53	ARG	C	C	13	173.977	0.004
385	53	ARG	CA	C	13	54.25	0.044
386	53	ARG	CB	C	13	27.373	0.013
387	53	ARG	CD	C	13	41.328	0
388	53	ARG	HD2	H	1	2.941	0
389	53	ARG	HN	H	1	8.919	0.011
390	53	ARG	N	N	15	126.72	0.036
391	54	ILE	C	C	13	172.161	0.009
392	54	ILE	CA	C	13	58.807	0.126
393	54	ILE	CB	C	13	37.207	0.125
394	54	ILE	CD1	C	13	12.548	0
395	54	ILE	CG1	C	13	29.182	0
396	54	ILE	CG2	C	13	15.832	0
397	54	ILE	HA	H	1	4.549	0.018
398	54	ILE	HB	H	1	1.913	0
399	54	ILE	HD11	H	1	0.443	0.019
400	54	ILE	HG12	H	1	0.865	0.009
401	54	ILE	HG21	H	1	0.662	0.004
402	54	ILE	HN	H	1	8.92	0.005
403	54	ILE	N	N	15	122.215	0.052
404	55	GLY	C	C	13	169.649	0.022
405	55	GLY	CA	C	13	42.712	0.25
406	55	GLY	HA2	H	1	3.774	0.016
407	55	GLY	HN	H	1	7.233	0.004
408	55	GLY	N	N	15	107.034	0.032
409	56	CYS	C	C	13	169.55	0.029
410	56	CYS	CA	C	13	53.638	0.621
411	56	CYS	CB	C	13	40.291	3.935
412	56	CYS	HA	H	1	5.264	0.017
413	56	CYS	HB2	H	1	3.108	0
414	56	CYS	HB3	H	1	2.897	0.018
415	56	CYS	HN	H	1	9.739	0.011
416	56	CYS	N	N	15	121.361	0.058
417	57	GLY	C	C	13	168.036	0.009
418	57	GLY	CA	C	13	41.978	0.032
419	57	GLY	HA2	H	1	3.863	0.022
420	57	GLY	HA3	H	1	4.501	0.005
421	57	GLY	HN	H	1	9.408	0.006
422	57	GLY	N	N	15	113.662	0.019
423	58	LYS	C	C	13	174.185	0.004
424	58	LYS	CA	C	13	51.418	0.088
425	58	LYS	CB	C	13	34.533	0.091
426	58	LYS	CD	C	13	26.895	0

427	58	LYS	CE	C	13	39.664	0
428	58	LYS	CG	C	13	21.926	0
429	58	LYS	HA	H	1	5.981	0.018
430	58	LYS	HB2	H	1	1.633	0.018
431	58	LYS	HD2	H	1	1.523	0.007
432	58	LYS	HE2	H	1	2.817	0
433	58	LYS	HG2	H	1	1.335	0.012
434	58	LYS	HN	H	1	8.427	0.004
435	58	LYS	N	N	15	117.565	0.032
436	59	GLY	C	C	13	168.361	0
437	59	GLY	CA	C	13	43.281	0
438	59	GLY	HN	H	1	8.534	0.003
439	59	GLY	N	N	15	106.06	0.025
440	60	PRO	C	C	13	172.019	0.012
441	60	PRO	CA	C	13	61.161	0.19
442	60	PRO	CB	C	13	29.185	0.049
443	60	PRO	CD	C	13	47.511	0
444	60	PRO	CG	C	13	23.585	0
445	60	PRO	HA	H	1	5.242	0.008
446	60	PRO	HB	H	1	2.268	0.017
447	60	PRO	HD2	H	1	3.657	0.024
448	60	PRO	HD3	H	1	3.428	0
449	60	PRO	HG2	H	1	2.022	0.06
450	61	SER	C	C	13	170.857	0
451	61	SER	CA	C	13	53.387	0
452	61	SER	CB	C	13	63.795	0
453	61	SER	HN	H	1	7.289	0.006
454	61	SER	N	N	15	110.749	0.017
455	62	LYS	C	C	13	175.308	0.001
456	62	LYS	CA	C	13	57.915	0.179
457	62	LYS	CB	C	13	28.989	0.312
458	62	LYS	CD	C	13	23.303	0.346
459	62	LYS	CE	C	13	39.311	0.055
460	62	LYS	CG	C	13	22.274	0.166
461	62	LYS	HA	H	1	3.155	0.02
462	62	LYS	HB2	H	1	1.524	0.034
463	62	LYS	HD2	H	1	1.227	0.036
464	62	LYS	HE2	H	1	2.796	0.016
465	62	LYS	HG2	H	1	0.824	0.009
466	63	LYS	C	C	13	175.856	0.003
467	63	LYS	CA	C	13	57.097	0.109
468	63	LYS	CB	C	13	29.746	0.294
469	63	LYS	CD	C	13	26.211	0.081

470	63	LYS	CE	C	13	39.627	0.002
471	63	LYS	CG	C	13	22.105	0.068
472	63	LYS	HA	H	1	3.709	0.018
473	63	LYS	HB2	H	1	1.692	0.045
474	63	LYS	HD2	H	1	1.557	0.006
475	63	LYS	HE2	H	1	2.862	0.032
476	63	LYS	HG2	H	1	1.238	0.034
477	63	LYS	HN	H	1	8.246	0.005
478	63	LYS	N	N	15	118.188	0.039
479	64	GLN	C	C	13	176.277	0.009
480	64	GLN	CA	C	13	56.24	0.057
481	64	GLN	CB	C	13	26.506	0.171
482	64	GLN	CG	C	13	32.492	0.112
483	64	GLN	HA	H	1	3.927	0.028
484	64	GLN	HB2	H	1	2.23	0.024
485	64	GLN	HG2	H	1	2.401	0.019
486	64	GLN	HN	H	1	7.701	0.005
487	64	GLN	N	N	15	116.733	0.031
488	65	ALA	C	C	13	175.75	0.007
489	65	ALA	CA	C	13	53.009	0.019
490	65	ALA	CB	C	13	16.089	0.039
491	65	ALA	HA	H	1	3.726	0.015
492	65	ALA	HB2	H	1	1.355	0.028
493	65	ALA	HN	H	1	7.933	0.007
494	65	ALA	N	N	15	123.381	0.042
495	66	LYS	C	C	13	175.275	0.005
496	66	LYS	CA	C	13	57.658	0.166
497	66	LYS	CB	C	13	30.268	0.232
498	66	LYS	CE	C	13	38.504	0
499	66	LYS	CG	C	13	23.432	0.023
500	66	LYS	HA	H	1	3.626	0.026
501	66	LYS	HB2	H	1	2.046	0.039
502	66	LYS	HD2	H	1	1.702	0
503	66	LYS	HG2	H	1	1.136	0.025
504	66	LYS	HN	H	1	8.353	0.006
505	66	LYS	N	N	15	118.715	0.025
506	67	MET	C	C	13	175.66	0.004
507	67	MET	CA	C	13	56.845	0.136
508	67	MET	CB	C	13	29.95	0.242
509	67	MET	CG	C	13	29.416	0.387
510	67	MET	HA	H	1	3.821	0.015
511	67	MET	HB2	H	1	1.984	0.006
512	67	MET	HB3	H	1	2.512	0

513	67	MET	HE1	H	1	1.646	0.006
514	67	MET	HG2	H	1	2.683	0.007
515	67	MET	HG3	H	1	2.456	0
516	67	MET	HN	H	1	8.011	0.003
517	67	MET	N	N	15	117.211	0.04
518	68	GLY	C	C	13	173.545	0.007
519	68	GLY	CA	C	13	44.643	0.181
520	68	GLY	HA2	H	1	3.896	0.005
521	68	GLY	HN	H	1	8.094	0.011
522	68	GLY	N	N	15	105.985	0.016
523	69	ALA	C	C	13	175.84	0.01
524	69	ALA	CA	C	13	52.512	0.012
525	69	ALA	CB	C	13	15.069	0.108
526	69	ALA	HA	H	1	3.841	0.021
527	69	ALA	HB1	H	1	1.436	0.042
528	69	ALA	HN	H	1	8.437	0.004
529	69	ALA	N	N	15	127.579	0.041
530	70	ALA	C	C	13	176.847	0.006
531	70	ALA	CA	C	13	52.759	0.053
532	70	ALA	CB	C	13	15.297	0.054
533	70	ALA	HA	H	1	4.003	0.003
534	70	ALA	HB1	H	1	1.531	0.002
535	70	ALA	HN	H	1	8.219	0.01
536	70	ALA	N	N	15	119.968	0.032
537	71	MET	C	C	13	175.596	0.007
538	71	MET	CA	C	13	56.613	0.107
539	71	MET	CB	C	13	30.407	0.143
540	71	MET	CE	C	13	15.427	0
541	71	MET	CG	C	13	29.618	0
542	71	MET	HA	H	1	4.01	0.011
543	71	MET	HB2	H	1	2.185	0.022
544	71	MET	HE1	H	1	1.036	0
545	71	MET	HG3	H	1	2.556	0.015
546	71	MET	HG2	H	1	2.724	0.006
547	71	MET	HN	H	1	8.294	0.004
548	71	MET	N	N	15	115.754	0.019
549	72	ASP	C	C	13	175.052	0.005
550	72	ASP	CA	C	13	54.995	0.019
551	72	ASP	CB	C	13	40.461	0.06
552	72	ASP	HA	H	1	4.389	0.004
553	72	ASP	HB2	H	1	2.867	0.004
554	72	ASP	HN	H	1	8.045	0.009
555	72	ASP	N	N	15	120.287	0.02

556	73	ALA	C	C	13	177.251	0.004
557	73	ALA	CA	C	13	52.041	0.037
558	73	ALA	CB	C	13	16.473	0.042
559	73	ALA	HA	H	1	4.188	0.008
560	73	ALA	HB2	H	1	1.183	0.004
561	73	ALA	HN	H	1	8.028	0.005
562	73	ALA	N	N	15	119.64	0.059
563	74	LEU	C	C	13	176.643	0.005
564	74	LEU	CA	C	13	55.519	0.155
565	74	LEU	CB	C	13	39.698	0.123
566	74	LEU	CD	C	13	21.46	0
567	74	LEU	CG	C	13	23.106	0
568	74	LEU	HA	H	1	3.992	0.007
569	74	LEU	HB	H	1	1.815	0.009
570	74	LEU	HD11	H	1	0.756	0.015
571	74	LEU	HD21	H	1	0.858	0.007
572	74	LEU	HG	H	1	1.464	0.011
573	74	LEU	HN	H	1	8.397	0.007
574	74	LEU	N	N	15	117.01	0.022
575	75	GLU	C	C	13	175.976	0.004
576	75	GLU	CA	C	13	55.916	0.071
577	75	GLU	CB	C	13	27.425	0.065
578	75	GLU	CG	C	13	34.215	0
579	75	GLU	HA	H	1	4.109	0.016
580	75	GLU	HB2	H	1	2.222	0
581	75	GLU	HG2	H	1	2.396	0.001
582	75	GLU	HN	H	1	7.812	0.011
583	75	GLU	N	N	15	117.709	0.027
584	76	LYS	C	C	13	174.882	0.002
585	76	LYS	CA	C	13	55.187	0.132
586	76	LYS	CB	C	13	31.27	0.172
587	76	LYS	CD	C	13	29.325	0
588	76	LYS	CE	C	13	40.074	0
589	76	LYS	CG	C	13	25.982	0
590	76	LYS	HA	H	1	4.14	0
591	76	LYS	HB	H	1	1.767	0.01
592	76	LYS	HD2	H	1	1.524	0.011
593	76	LYS	HE2	H	1	2.896	0
594	76	LYS	HG2	H	1	1.352	0
595	76	LYS	HN	H	1	7.892	0.004
596	76	LYS	N	N	15	117.44	0.039
597	77	TYR	C	C	13	172.041	0.005
598	77	TYR	CA	C	13	56.878	0.111

599	77	TYR	CB	C	13	35.852	0.047
600	77	TYR	HA	H	1	4.044	0.01
601	77	TYR	HB2	H	1	2.554	0.013
602	77	TYR	HN	H	1	8.149	0.008
603	77	TYR	N	N	15	120.225	0.021
604	78	ASN	C	C	13	171.116	0.006
605	78	ASN	CA	C	13	49.866	0.16
606	78	ASN	CB	C	13	35.302	0.159
607	78	ASN	HA	H	1	4.537	0.006
608	78	ASN	HB2	H	1	2.539	0.032
609	78	ASN	HN	H	1	7.436	0.009
610	78	ASN	N	N	15	121.088	0.052
611	79	PHE	C	C	13	171.22	0
612	79	PHE	CA	C	13	54.195	0
613	79	PHE	CB	C	13	35.969	0
614	79	PHE	HN	H	1	7.901	0.003
615	79	PHE	N	N	15	122.211	0.03
616	80	PRO	C	C	13	173.721	0.004
617	80	PRO	CA	C	13	60.634	0.157
618	80	PRO	CB	C	13	29.356	0.111
619	80	PRO	CD	C	13	48.246	0
620	80	PRO	CG	C	13	24.811	0
621	80	PRO	HA	H	1	4.367	0.009
622	80	PRO	HB	H	1	2.196	0
623	80	PRO	HD3	H	1	3.586	0
624	80	PRO	HD2	H	1	3.763	0.006
625	80	PRO	HG	H	1	1.898	0.013
626	81	GLN	C	C	13	172.475	0.006
627	81	GLN	CA	C	13	53.216	0.155
628	81	GLN	CB	C	13	26.823	0.184
629	81	GLN	CG	C	13	31.497	0
630	81	GLN	HA	H	1	4.232	0.009
631	81	GLN	HB2	H	1	2.061	0.004
632	81	GLN	HB3	H	1	1.905	0.003
633	81	GLN	HG2	H	1	2.352	0.044
634	81	GLN	HN	H	1	8.347	0.004
635	81	GLN	N	N	15	120.432	0.031
636	82	MET	C	C	13	178.029	0
637	82	MET	CA	C	13	54.275	0
638	82	MET	CB	C	13	31.111	0
639	82	MET	HN	H	1	7.933	0.003
640	82	MET	N	N	15	127.154	0.053