A STUDY OF BIOMOLECULAR INTERACTIONS USING THREE BIOLOGICAL COMPLEXES TO EXPLORE STRUCTURE, DYNAMICS AND METHOD DEVELOPMENT

SHERRY NICOLE BOODRAM

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

AUGUST 2013

ABSTRACT

The protein, RDE-4 in *C. elegans*, served as a model for studying how double stranded RNA binding proteins that bind to dsRNA molecules are critical for RNAi cellular processes. NMR spectroscopy confirms that the RDE-4 construct has characteristic protein domains that bind to dsRNA and that RNA binding causes a significant global change of the protein structure. SAXS analysis indicates that the two binding domains in the RDE-4 protein do not interact with one another, but instead forms a continuous interface onto which long target dsRNA can bind. Gel shift assay experiments reveal that multiple RDE-4 molecules bind to a non-sequence specific RNA substrate with positive cooperativity. RNA binding occurs with micromolar affinity and a second binding event occurs with millimolar affinity.

The binding of *E. coli* dihydrofolate reductase (DHFR) to inhibitors methotrexate (MTX) and 1,4-Bis-{[*N*-(1-imino-1-guanidino-methyl)]sulfanylmethyl}-3,6-dimethyl-benzene (inhibitor **1**) has been studied to investigate the dynamics involved in the catalytic mechanism of DHFR. NMR relaxation methods show that in the presence of inhibitor **1**, the catalytic domain of DHFR binds **1** in the substrate-binding pocket and an occluded conformation is assumed. In both the DHFR:NADPH and DHFR:NADPH:**1** complexes, motion is exhibited on the microsecond-millisecond timescale. The heat capacity change of DHFR upon binding to **1** and MTX are 43 ± 10 cal/mol-K and -120 ± 109 cal/mol-K respectively. Differences in ΔC_p of DHFR

binding to inhibitor **1** compared to that of MTX indicate that the mode of binding to **1** is different from what is observed in the crystal structure of the complex.

An affinity electrophoresis method to screen for RNA-small molecule ligand interactions has been developed. This method is made quantifiable by cross-linking the ligand into the gel matrix and gauging binding by RNA mobility. The utility of this method is demonstrated using the known interaction between different aminoglycoside ligands with the *E. coli* ribosomal A-site RNA and with an RNA molecule containing a C-C mismatch. Average apparent dissociation constants are determined. This method allows an easy quantitative comparison between different nucleic acid molecules for a small molecule ligand.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Philip E. Johnson for his guidance, support, and continuous belief in me. I am so grateful for this opportunity that he has given to me. I would also like to thank my examining committee, which includes Professor Logan Donaldson, Professor Gerald Audette, Professor Kathi Hudak, and Professor Gaya Amarasinghe for their time and valuable suggestions. I wish to extend my gratitude to Professor Logan Donaldson for providing his expertise and assistance with the NMR experiments. I would also like to thank my colleagues in the Johnson lab and friends from the Chemistry Department who have made my experience as a graduate student memorable.

To my mother, Patricia, and father, Sinanan, whose love, guidance, encouragement, and unwavering faith led me to where I am now. And for that I am forever grateful. To my sisters, Natalie, Kate, and Sarah, and to my brother, Kevin, who traveled the academic journey along side me and understand the paradox of loving and hating school all at the same time. To my grandmother, Enid, for her gentle support. To my partner, Lucas McCann, for his shared interest in chemistry and for providing me with endless love, laughter, and support in all areas of my life. And lastly, to God for always watching over me and for having a plan for me that was greater than I even imagined possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
aa	amino acid
ABPP	activity based protein profiling
AE	affinity electrophoresis
APS	ammonium persulfate
ASA	apolar surface area
bp	base pair
CCMS	capture compound mass spectrometry
C_p	heat capacity
CPMG	Carr-Purcell-Meiboom-Gill
CSI	chemical shift index
DCR-1	Dicer-1
DGCR8	DiGeorge syndrome critical region gene 8
DHFR	Dihydrofolate Reductase
DNA	deoxyribonucleic acid
dsRBD	double stranded RNA-binding domain
dsRBP	double stranded RNA-binding protein
dsRNA	double stranded RNA
EMSA	electrophoretic mobility shift assay
endo-siRNA	endogenous small interfering RNA
exo-siRNA	exogenous small interfering RNA
FID	free induction decay
FPLC	fast protein liquid chromatography
GdnHCl	guanidine hydrochloride
hetNOE	heteronuclear overhauser effect
HRMS	high-resolution mass spectrometry
HSQC	heteronuclear single-quantum coherence
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infrared
ITC	isothermal titration calorimetry
miRNA	microRNA
mRNA	messenger RNA
MTX	methotrexate
MWCO	molecular weight cut-off
NADPH	nicotinamide adenine dinucleotide phosphate
ncRNA	non-coding RNA
NMR	nuclear magnetic resonance
PACE	polyacrylamide gel co-electrophoresis

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
piRNA	Piwi-interacting RNA
PSA	polar surface area
R_1	spin-lattice or longitudinal relaxation
R_2	spin-spin or transverse relaxation
rasiRNA	repeat associated small interfering RNA
RD	relaxation dispersion
RDE-4	RNAi defective-4
RF	radio frequency
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RRM	RNA recognition motif
rRNA	ribosomal RNA
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
siRNA	small interfering RNA
S/N	signal-to-noise
snRNA	small nuclear RNA
ssRNA	single stranded RNA
T_1	relaxation time of R_1
T_2	relaxation time of R_2
TEMED	tetramethylethylenediamine
tRNA	transfer RNA
TSMC	thymidylate synthase mRNA construct
UTR	untranslated region

STATEMENT OF COLLABORATIONS

I am the primary contributor of the experimental work presented in Chapter 4. The exception to this includes the affinity electrophoresis gel experiment described in section 4.3.1 on page 180 and shown in Figure 4.4, which was completed by Chul M. Cho. In addition, Lucas C. McCann completed the synthesis of the cross-linked aminoglycoside-acrylamide described in section 4.2.2. Both persons mentioned were colleagues in the Department of Chemistry at York University at the time the experiments were performed.

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

1.1 BIOMOLECULAR INTERACTIONS

Biomolecular interactions play a fundamental role in the majority of key biochemical processes. Biomolecules can form a variety of interactions including protein-protein, protein-nucleic acid, protein-small molecule, and nucleic acid-small molecule interactions. To gain a better understanding of the array of interactions, and more specifically the relationship between biomolecular interactions and cellular processes, it is important to investigate the mechanisms involved and the effects of these interactions. This knowledge gives rise to the function of biomolecules, which can lead to significant advances in diagnostic applications and in the treatment of diseases.

This chapter provides background information on the three projects presented in this dissertation as well as the details of the techniques used in each project.

1.2 PROTEIN-RNA INTERACTIONS

The recognition of ribonucleic acid (RNA) molecules by proteins is critical for understanding cellular processes such as gene expression, RNA processing, and viral replication. Since RNA molecules have fairly diverse structures, proteins that interact with RNA must be able to recognize an assortment of RNA [1, 2]. The non-

canonical Watson-Crick base pairing of single-stranded like structures of RNA allows for a range of geometries that facilitates greater hydrogen-bonding ability and base stacking capabilities. Even though the size, shape, and general binding interactions may vastly differ between RNA molecules, proteins can conform tight binding pockets for specific RNA with few contact points upon complex formation [3]. The structures found in RNA molecules are much more diverse than those of deoxyribonucleic (DNA) molecules, suggesting that protein-RNA interactions can be quite different than protein-DNA interactions [3, 4].

The three dimensional structure of an RNA molecule is helpful for understanding how the RNA molecule docks onto a protein molecule. If the three-dimensional structure of an RNA molecule is unknown, considering the type of RNA of interest can often allow for conformational predictions. Most often, RNA molecules that have extensive structures such as tRNA and rRNA are rigid whereas other RNA such as mRNA, non-protein coding RNA (ncRNA), and small nuclear RNA, (snRNA) are fairly flexible [1, 3].

Proteins that bind RNA have characteristic domains containing amino acid sequences that allow for RNA interactions. Comparing the similarity of domain sequences to the sequences of proteins with known RNA binding domains often makes the prediction of RNA-binding proteins possible, although this is not the only condition for which proteins need for RNA binding to occur. Information about the RNA-binding interface on the surfaces of proteins provides a means for making binding predictions. An easy approach to making predictions about protein-RNA

complex formation is to determine the electrostatic potential of the surface of the given protein and evaluate the opportunity of the protein binding RNA [3]. Since the RNA molecule carries a negative charge due to the phosphate backbone, it is logical to assume that the RNA interface on the protein surface has a positive electrostatic potential. It must be noted that the pH and ionic strength of the solvent used can alter the electrostatic potential.

RNA interfaces on protein surfaces are reported to contain preferred amino acids residues for binding. To compliment the negative charge of RNA, amino acids with positive charges such as arginine (Arg) and lysine (Lys) frequently appear in regions of the protein sequence that suggests an RNA interface. In addition, RNA molecules favour van der Waal contacts with amino acids that have aromatic side-chains that allow for stacking interactions with the RNA bases. Furthermore, the location of amino acid residues at the interface is spatially located so that its neighbour in the tertiary protein structure is also part of the same interface [3].

1.2.1. Double Stranded RNA-Binding Proteins (dsRBPs)

Proteins with conserved double stranded RNA-binding domains (dsRBDs) are a growing protein class found in eukaryotes, prokaryotes, and viral products that facilitate the interaction with double stranded RNA (dsRNA) [5]. These double stranded RNA binding proteins (dsRBPs) have been reported to have critical and diverse roles in cellular processes including the regulation of gene expression, RNA editing, RNA transport, RNA processing, and host defense in biological processes [5].

Examples include *S. cerevisiae* RNase III (rnt1p) that cleaves RNA substrates containing hairpins with an A/uGNN tetraloop cap (Figure 1.1) [6] and protein kinase PKR that is involved in dsRNA signaling, RNA interference (RNAi) and host defense against viral infection [5].

Figure 1.1: Solution structure of double stranded RNA binding domain of *S. cerevisiae* RNase III (rnt1p) (green) with the terminal RNA hairpin of snr47 precursor (PDB ID: 2LUP) [6].

The dsRBD is a small protein domain consisting of at least one αββββα motif of 65-70 amino acids. The discovery of dsRBDs as dsRNA binding domains was first recognized in 1992 as a conserved functional domain based on sequence similarity matches between the proteins Staufen in *Drosophila*, human TAR-RNA binding protein (TRBP), and *Xenopus leavis* RNA binding protein A (XlrbpA), along with characterization of dsRBDs in human protein kinase (PKR) and human TRBP [7]. The dsRBD domain is one of the most abundant RNA binding domains found in cellular systems aside from the RNA recognition motif (RRM) that is known to mostly bind single-stranded RNA (ssRNA). It is important to note that dsRNA binding is not limited to dsRBDs, rather other domains known to bind dsRNA include zinc fingers and SAM domains, and RRM domains [7]. The solution NMR structures of *E. coli* RNase III dsRBD and the third dsRBD of *Drosophila* Staufen were the first three-dimensional structures of dsRBDs to be solved [7].

Interestingly, proteins with dsRBDs have been found to interact with short strands of A-form helix dsRNA that are 11-16 base pairs of usually nonspecific nucleotide sequences [5]. Proteins that bind dsRNA have been found to bind sequence specific dsRNA targets *in vivo*, however, rather than being sequence specific, binding of dsRBDs to dsRNA is commonly believed to be dependent on the shape of dsRNA [7]. One reason being that A-form helix dsRNA differs form classic B-form dsDNA in that the major groove is shallow and broad and the minor groove is deep and narrow, allowing dsRBPs to bind nonspecifically to dsRNAs. The lack of

nucleotide specificity suggests that other helper proteins are needed to assists with target recognition [5].

1.3 PROTEIN-SMALL MOLECULE INTERACTIONS

The interaction of small molecules with proteins is a central component of many diverse biological processes. Such interactions are often seen in regulatory systems and catalytic cycles. One of the most recognized instances of a small molecule binding with a protein is the interaction of an O_2 molecule with hemoglobin to transport oxygen from the respiratory systems to the rest of the body (Figure 1.2) [8, 9]. The function of the small molecule moiety is highly selective for the target protein. For this reason, the small molecules are often used as bait for studying protein binding where the selectivity functional group of the molecule can be exploited to explore the complexity of binding modes and ligand orientation preferences known as "molecular docking" [10, 11]. Currently available methods for studying ligand-protein interactions, and more specifically molecular docking, largely involve computer algorithms as a virtual screen tool [11]. In doing so, it is possible to quickly and efficiently classify the binding of a vast set of small molecules with a target protein, which is extremely useful to the pharmaceutical industry [10, 11]. When the small molecule that is used as bait is a pharmaceutical compound, screening processes can lead to the discovery of drug targets [10]. For drug target discovery, it is important that other functionalities can be positioned on

parts of the small molecule, especially when a structure-activity relationship is unclear or unknown [10].

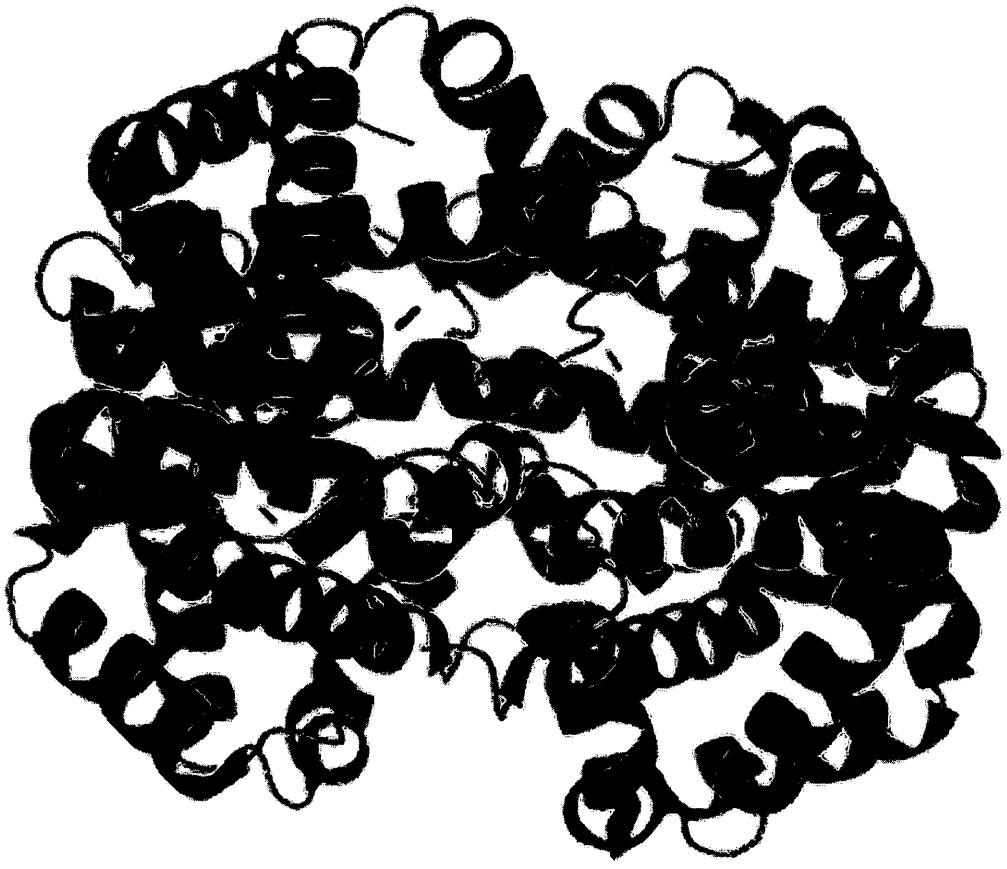


Figure 1.2: Three-dimensional structure of hemoglobin (blue) (PDB ID: 1GZX) in complex with 4 oxygen molecules (pink) [9].

Design strategies used for profiling small molecule-protein interactions often involve bi-functional probes that have an equilibrium-driven affinity with the target protein. One such strategy is the affinity pull-down approach that involves linking the target protein to a solid phase and subsequently allowing the probe to interact with the linked protein by passing the probe over the solid phase. The result is a low-affinity reversible interaction between the probe and target protein, which can be desirable in certain cases, however, unspecific binding proteins and buffer conditions may affect the recovery yield of the proteins of interest [10]. This strategy has been employed successfully to isolate various kinases in eukaryotic lysate by using kinase inhibitor beads [12, 13]. Activity based protein profiling (ABPP) is another type of probe strategy that instead uses small molecules that contain chemical groups that can directly react with amino acids in an enzyme active site by creating strong covalent bonds, thereby avoiding the problem of losing the small molecule-protein interaction during subsequent procedural steps [10]. ABPP is a common approach used for protein family profiling such as the cysteine hydrolase family of hydrolytic enzymes [14, 15]. In the strategy of Capture Compound Mass Spectrometry (CCMS), the small molecule probe is comprised of a central atom with three protruding functional units capable to selectively binding to target proteins in an equilibrium-driven manner. The established equilibrium is then "frozen" by a photo-activation process that results in a covalent bond crosslink that is strong enough to withstand further processing steps [10]. CCMS has been

used to select for protein families such as camp-binding protein [16] and GTPases [17].

The applications of small molecule-protein interaction profiling serves as a powerful tool as it benefits the field of drug discovery for treating diseases [18] and extends to the identification of natural products [19]. There is no doubt that understanding small molecules-protein interactions is necessary for uncovering the complexity of many biological processes.

1.4 RNA-SMALL MOLECULE INTERACTIONS

As the diverse responsibilities of RNA in biological pathways are becoming more evident, small molecules that selectively bind to RNA are also becoming more important for gaining insight into the many roles that RNA undertake, such transcriptional and translation regulation [20, 21], catalysis [22], and protein function [23]. Although protein targets form the basis of drug discovery, RNA targets are increasingly viewed as an alternative method for developing treatments for diseases [24]. Like DNA and proteins, RNA has many attractive features that allows for interaction with a small molecule target. However, RNA molecules are capable of adopting many complex three-dimensional structures because mismatched or unpaired bases in the nucleotide sequence can easily disrupt the regular A-form helix normally seen in nucleic acids [24]. It is this structural diversity that enables RNA to facilitate a wide range of biological pathways. It is not

surprising that RNA possesses the potential for therapeutic-based applications with antibacterial, antiviral and mRNA being the three most common RNA targets. For example, studies have shown that certain ribosomal RNA (rRNA) can bind with aminoglycoside antibiotics to initiate an antibacterial effect [25]. Also, several riboswitch RNA can form a complex with a small molecule metabolite to regulate translation (Figure 1.3) and even display antibacterial properties upon binding when targeted by an inhibitor of the normal riboswitch-ligand interaction [26, 27].

Figure 1.3: The SAM-I Riboswitch in complex with S-adenosylmethionine (pink) (PDB ID: 2YGH) [27].

The manner in which an RNA molecule is folded can indicate the possibility of an interaction with a small molecule. The structural hierarchy of RNA is such that the primary nucleotide sequence determines the secondary structure that subsequently dictates the tertiary structure. Since RNA molecules are single stranded, they require a definite tertiary structure in order to be functional. This scaffold is reinforced by a network of hydrogen bonding to create secondary structure elements such as hairpin loops, internal loops, bulges, and duplexes [2, 28]. Ligand-binding sites often occur at a single secondary structure element and secondary structures often provide pockets to secure small molecule interactions [24].

Selectively targeting RNA with small molecules is made possible with screening methods and RNA binding assays are available to determine the strength of binding affinity [24]. However, RNA biochemistry is still relatively new and there is a lot to be uncovered. An increase in RNA-small molecule research and advanced screening techniques are required to make detailed characterization of RNA-small molecule interactions achievable and to encourage the study of these complexes as a viable avenue for drug discovery.

1.5 RNA INTERFERENCE

This section provides an overview on the topic of RNA interference as it relates to the research project outlined in Chapter 2.

The RNA-related discovery that has gained the most widespread recognition and use is that of RNA interference (RNAi), a process that occurs innately in cells as a sequence dependent cellular response to RNA [29-31]. The role of double stranded RNA (dsRNA) in gene regulation through RNAi has defined a new biological role for RNA that has generated much interest since its discovery. RNAi is conserved in all eukaryotes and is involved in many diverse biological events including embryo development, viral defense and cancer progression. The ability of RNAi to silence the activity of specific genes using short (~22 base pair) RNA molecules has directed the use of RNAi to being a therapeutic modality and has also found applications in proteomics research [32-35]. The significance of the discovery of RNAi was recognized in 2006 by the awarding of the Nobel Prize in Physiology or Medicine to Craig Mello and Andrew Fire for their work on RNAi that was published in 1998 [29].

RNAi is believed to have evolved as a cellular defense mechanism against foreign intruders such as RNA viruses [36]. When viral RNA replicate, it temporarily exists as a double strand that is processed into short dsRNA to trigger the cascading events of RNAi. The virus' genes are effectively silenced and infection by the virus is avoided. RNAi may have also evolved to prevent the mobilization of transposable

genetic elements called transpoons [37]. Damage by transpoons is a result of the ability of these genetic elements to translocate on several areas on a genome, causing genetic mutations that can lead to diseases such as cancer. Transpoons, like RNA viruses, can adopt a double stranded RNA form that, when processed to short dsRNA, initiates RNAi to halt potentially damaging translocation.

The different classes of short RNA molecules that trigger RNAi are expanding steadily. These groups include micro RNA (miRNA), small interfering RNA (siRNA), repeat-associated siRNAs (rasiRNA), endogenous siRNA (endo-siRNA), and Piwi-interacting RNA (piRNA) [34, 38]. These small noncoding RNA molecules work by suppressing the expression of genes targeted by their nucleotide sequence. These classes of short RNA molecules are differentiated by the origin of the trigger RNA that initiates RNAi and the diverse biological roles they play. The trigger RNA can have either an endogenous or exogenous origin (Figure 1.4).

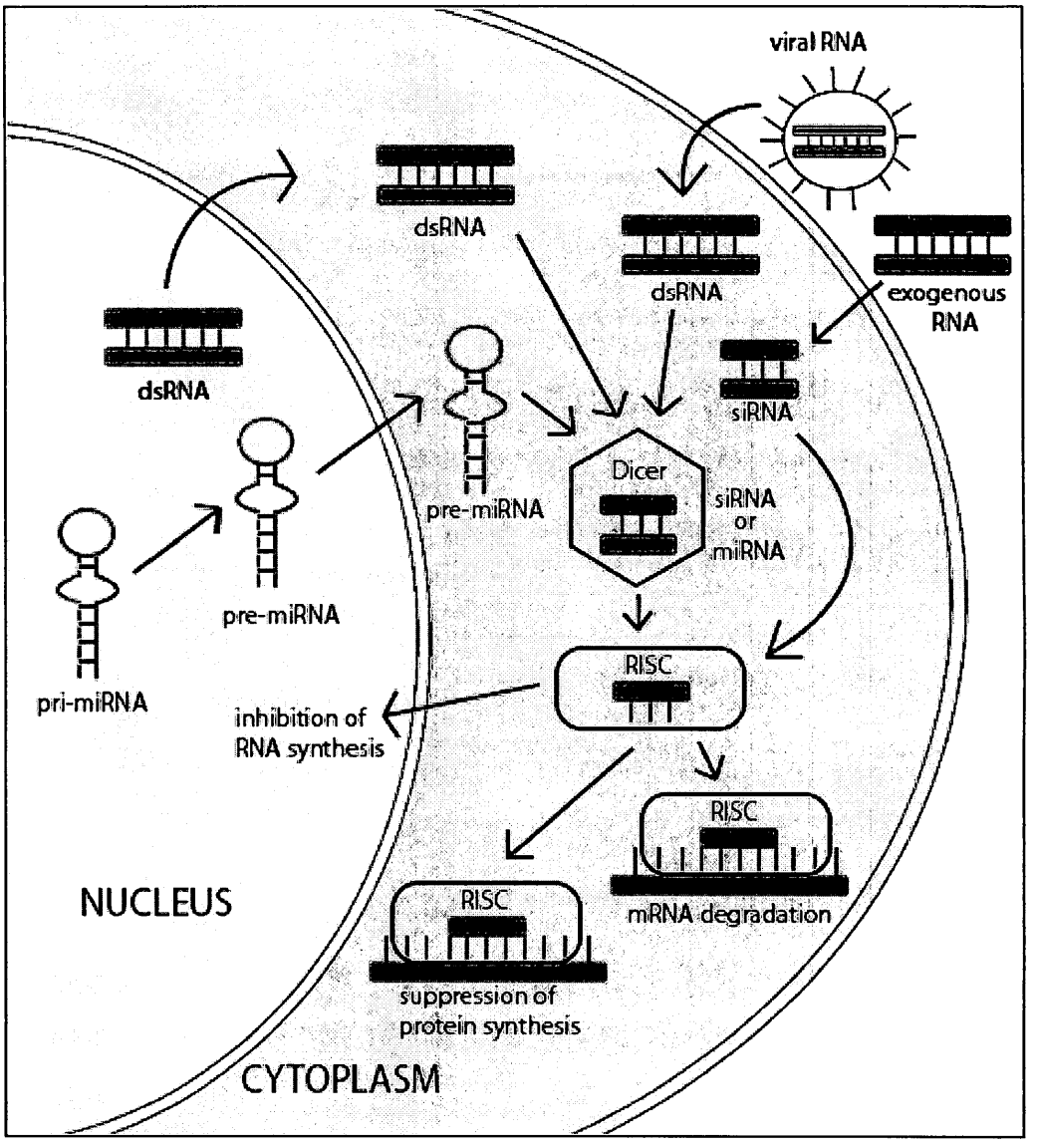


Figure 1.4: The processing pathways of miRNA and siRNA that are involved in RNAi. Short RNAs are differentiated by the origin of the trigger RNA. Adapted from [39].

Processing of siRNAs from endogenous or exogenous dsRNA precursors that are composed of perfectly base paired RNA results in short RNA strands that can bind to a complimentary sequence of target mRNA, causing the mRNA to be targeted for degradation. Mature miRNAs are structurally similar to siRNAs, however miRNAs must undergo extensive post-transcriptional modifications before entering the silencing pathway [40]. The miRNA originates as a long single stranded primary transcript that is transcribed by RNA polymerase II from a much longer RNA coding gene. This primary transcript, known as pri-miRNA, is processed in the nucleus to a 70-nucleotide stem loop structure called pre-miRNA by the processor complex. This complex consists of an RNase III endonuclease named Drosha and the essential dsRNA-binding protein cofactor called DGCR8 in humans or Pasha in invertebrates. The pre-miRNA is exported to the cytoplasm where Dicer, a second RNase III endonuclease, processes the dsRNA segment of the pre-miRNA to a short double strand miRNA duplex. Finally, an ATP-dependent helicase unwinds the short double strand miRNA duplex into two strands of mature single-stranded miRNA. One of the two strands, known as the guide strand or antisense strand, is incorporated into the RNA induced silencing complex (RISC) and then annealed to the target mRNA to perform regulatory functions. Although structurally similar to siRNAs, miRNAs regulate gene expression by binding to sequences in the 3' untranslated region (UTR) [41] of mRNAs by way of partial base pairing in mammals and perfect base pairing in plants, enabling the miRNA to regulate many mRNA with similar

sequences [34, 42, 43]. Binding of the 3' UTR mRNA leads to mRNA degradation or translational suppression.

Over 300 miRNAs are recognized in humans, however the exact function of the members of this group of RNA is still largely unclear [34]. Endogenous miRNAs are most important in translational repression, cellular developmental regulation and stem cell maintenance [44]. Studies have also shown that miRNA influences viral replication in mammals by having either a viral or anti-viral response [45]. Some viruses such as the Epstein-Barr virus (a member of the human Herpesvirus) and the Adenovirus, have been shown to encode miRNAs [46]. In the case of the Polyomavirus, Simian Virus 40 (SV40), the virus is able to potentially escape the immune response of the cytotoxic T cells by having its miRNA hybridize to a complimentary transcript encoding large T antigen [47]. Cellular miRNA can also have an alternate role that effectively inhibits or limits the replication of retroviruses. For example, replication of the Primate Foamy Virus type 1 (PFV-1), a complex retrovirus similar to Human Immunodeficiency Virus (HIV), is inhibited by cellular miR-32 that binds to the viral mRNAs [48].

In many organisms, including humans, miRNAs directly function as oncogenes or tumor supressors, with more than 50% of miRNA genes located in cancer-associated regions. Many miRNAs are directly implicated in human cancers, including lung, breast, brain, liver, colon, and bone marrow. Evidence for the contribution of miRNAs in cancers was initially reported in a 2002 molecular study characterizing the deletion of 13q14 in human chronic lymphocytic leukemia (CLL),

which is known to be the most prevalent form of adult leukemia in North America [49]. In cancers, overexpressed miRNAs such as mir-17-92 acts as oncogenes and negatively regulates genes that control tumor suppression, cell differentiation, or apoptosis [50]. The miRNAs with oncogenic roles are usually expressed at remarkably higher levels in cancer tissues compared to normal tissues. Underexpressed miRNAs can function as tumor suppressor genes that regulate oncogenes, cell differentiation, or apoptosis [51]. These miRNAs have decreased expression in cancerous cells and often inhibit the expression of a target oncogene by binding to the 3' UTR of the oncogene's mRNA to repress translation. For example, the let-7 miRNA negatively regulates expression of the RAS oncogene. In many human cancers, let-7 is often downregulated or deleted. Studies have reported that restoration of low levels of the let-7 miRNA in cancerous lung cells by the introduction of increased let-7 can effectively repress cell proliferation [52]. This ability to prevent tumorgenesis makes miRNAs, such as let-7, an appealing potential therapeutic [52, 53]. Depending on the type of cancer, miRNAs may have different expression profiles in cancerous cells compared with normal cells [54]. This suggests that in some instances, the same miRNA can function both as oncogenes or tumor suppressor depending on the type of cancer. Although the classification of miRNAs can be made by techniques such as expression profiling [54], better understanding of how miRNAs function in human oncogenesis on a molecular basis can potentially lead to early prognosis and possible treatments.

Despite their differences in origin, short RNA molecules are generated by the same general mechanism (Figure 1.4). In the initial stages of RNAi, the Dicer complex processes the trigger dsRNA into short (~22 base pair) interfering dsRNA (siRNA) molecules that is then incorporated into the RNA-induced silencing complex (RISC). Guided by the siRNA, the RISC then either degrades the mRNA molecules or represses translation based on the complementarity of the target mRNA to the siRNA [55].

1.6 PROTEIN NMR SPECTROSCOPY

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the few methods available for studying the three-dimensional structures of proteins at an atomic level and is still the only method for determining high-resolution three-dimensional protein structures in the solution state. Low-resolution protein structures are often achieved by the technique of small angle X-ray scattering (SAXS). The constitution of proteins is based on the unique connection of amino acids comprised of hydrogen, nitrogen, and carbon nuclei, among others, that can be detected by NMR spectroscopy when the nuclear spins go from its ground state to an excited state. This energy transition can be achieved by applying a static external magnetic field (B_0), a requirement that makes NMR spectroscopy a very attractive technique because it allows for a broad range of transition frequencies by simply using an oscillatory field (B_1) of radio-frequency (RF) electromagnetic radiation [56].

The energies associated with this type of two-energy level system results from the interaction of the external magnetic field and a nuclear magnetic dipole moment (μ) that arises from the spin angular momentum (S) of the nucleus. Nuclei that have a spin angular momentum include all nuclei with an odd mass number such as ^{1}H, ^{13}C, and ^{15}N because they have an unpaired proton. Nuclei with an even mass number and an odd charge such as ^{2}H and ^{14}N also have a spin angular momentum [57].

The spin angular momentum is a vector property that characterizes the nucleus as a spinning particle that produces a magnetic field described by a proportional magnetic dipole moment vector (Figure 1.5) [57]. In the absence of a magnetic field, all possible orientations of the magnetic dipole of the nuclear spins are present. However, the spin orientation can be made uniform when placed in a static magnetic field so that spins re-align to the direction of the z-axis defined by the magnetic field [57]. Note that the actual magnetic field felt by the nucleus is most often somewhat modified by surrounding electrons that shield the nucleus.

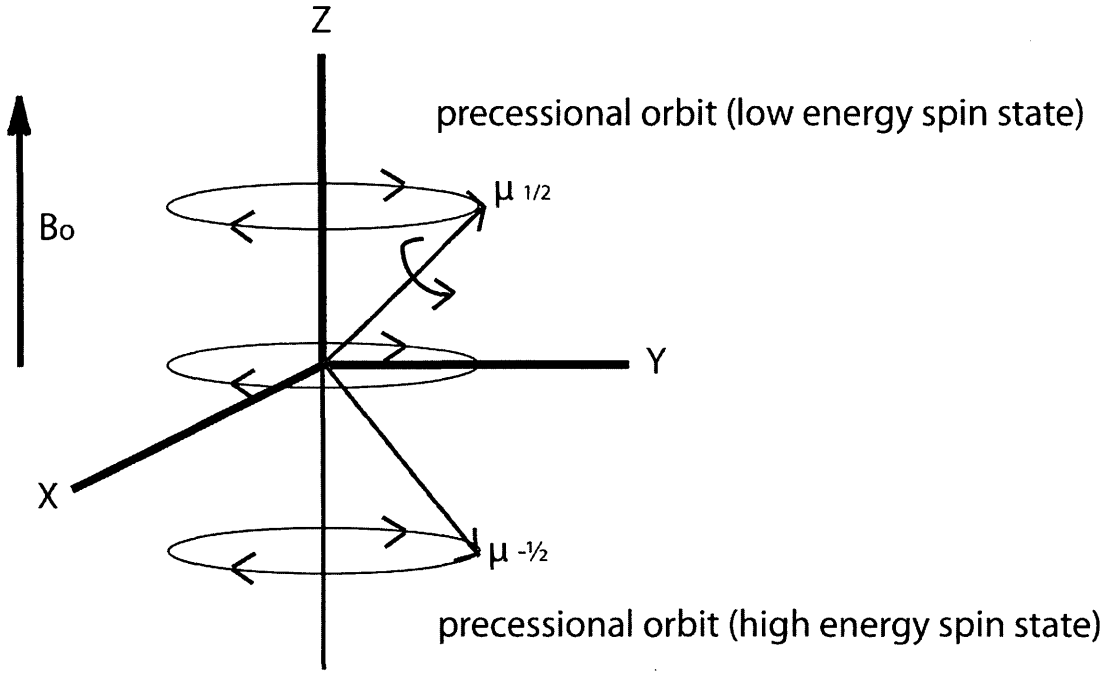


Figure 1.5: Orientation of the nuclear magnetic dipoles (μ) in the presence of a static magnetic field (B_o) along the z-axis. On average one half of the nuclear magnetic dipoles are pointed upward as ½ (low energy spin state) and the other half are pointed downward as -½ (high energy spin state). Since the magnet is spinning, the angular momentum is conserved and does not simply align with B_o. Instead, the nuclear magnetic dipoles precess around the magnetic field. Adapted from [57].

When a nucleus is placed in an external magnetic field, B_0, the nucleus acts as a rotating particle causing the magnetic field of the nuclei to precess around the axis of B_0. In actuality, all of the spins in a sample precess randomly around B_o at their specific Larmor frequency; however, slightly more of the nuclei are in the lower energy ½ spin state (Figure 1.6 A). This is called *Larmor precession* [56, 58]. The frequency property associated with this precession is proportional to the strength of the external magnetic field, meaning that higher frequencies are obtained with a stronger external magnetic field. To observe the magnetic moment of precessing nuclei, the nuclear magnetization must be made perpendicular to B_0, which can be achieved by applying a radio frequency pulse (B_1) that is the same frequency as the Larmor frequency and perpendicular to B_0 [57]. In other words, the bulk nuclear magnetization of the sample is tipped from the z-axis to the x-axis (Figure 1.6 B). After the pulse is applied, the rotating magnetic field induces an electric current that is measured by receiver coils in the circuit and digitized. Over time, the induced current declines because the perpendicular nuclear magnetization becomes out of phase and the magnetization returns the direction of the B_0 axis. This decay rate is referred to as the free induction decay, FID, and is a function of time. Converting the time domain to a frequency domain by Fourier transformation allows the FID to be observed as an NMR spectrum [58].

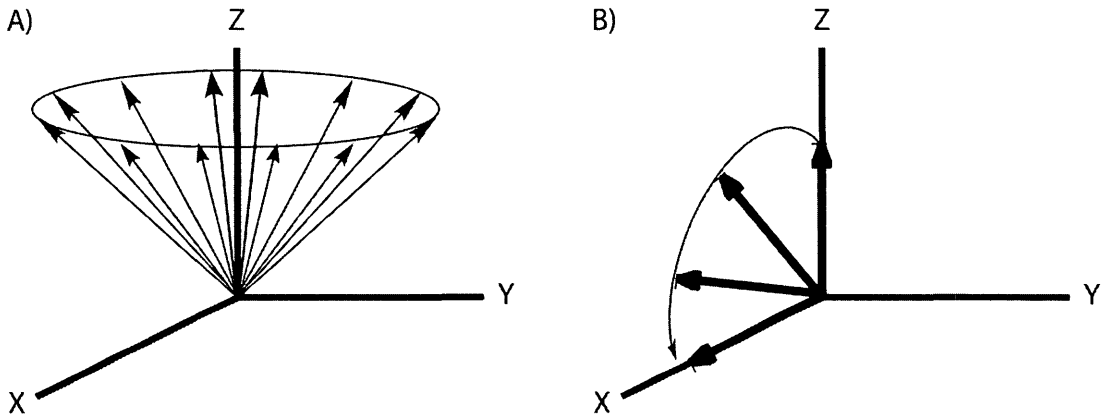


Figure 1.6: A) Collection of nuclear magnetic dipoles precessing around the axis of B_o. B) The bulk magnetization tipped from the z-axis to the x-axis, causing the bulk magnetization to remain in the z-x plane during the entire period of the pulse. Adapted from [57].

To study proteins using NMR, proteins are typically uniformly isotopically labeled with ^{15}N and ^{13}C nuclei by recombinant expression in minimal media that contain ^{15}N ammonium sulfate or ^{15}N ammonium chloride as the sole nitrogen source and ^{13}C glucose as the sole carbon source. For larger proteins, complete or partial ^{2}H labeling in D_2O or D_2O/H_2O mixture, a process called deuteration, may also be necessary. The downside of isotopically labeling is that it can be fairly expensive [59]. However, in doing so, various multinuclear experiments can be employed that allow for efficient magnetization transfer via J-coupling between

nuclei in the same or adjacent amino acid and thereby correlating the associated chemical shifts [60] (Figure 1.7).

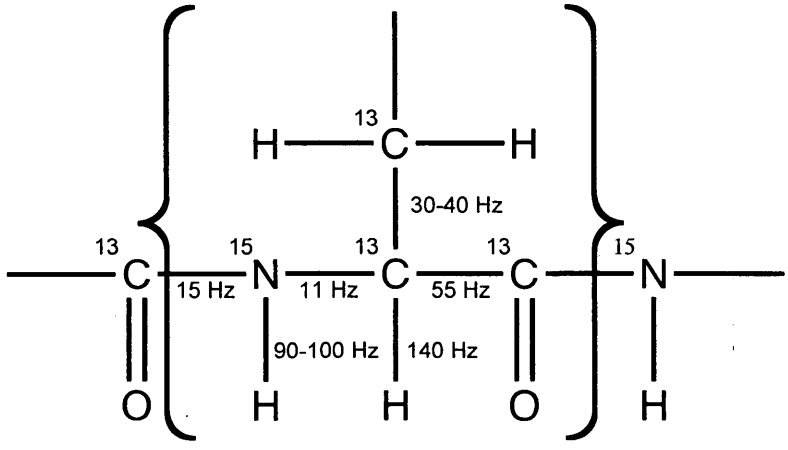


Figure 1.7: J-coupling constants between nuclei in an amino acid that allows for efficient magnetization transfer during multinuclear NMR experiments. Adapted from [57, 61].

Typically, structure determination of protein in the size range of 5-25 kDa is relatively facile, unless there are issues of aggregation and reduced solubility. With larger proteins, magnetization transfer between nuclei is decreased due to rapid signal decay by dipolar interactions and chemical anisotropy. Due to the increased protein size, molecular tumbling reduces the signal-to-noise (S/N) ratio in three-dimensional experiments by increasing magnetic field fluctuations that are a function of nuclear gyromagnetic ratios in the spin system [62]. Linewidth is also

increased by poor protein relaxation, which causes a reduction in resolution. For these reasons, structure determination by NMR is limited to proteins with molecular weights less than about 40-60 kDa [58].

1.7 TRIPLE RESONANCE EXPERIMENTS

Detailed information about protein backbone resonances can be obtained by performing triple resonance experiments that involve ^{15}N, ^{13}C and ^{1}H spins to generate three-dimensional spectra. This method is highly reliable, sensitive, simple to apply, and imperative for assigning resonances that are used to determine a protein's structure with confidence, especially when a protein is fairly large [57]. Large proteins (> 150 residues) often produce poor quality spectra because tumbling is slow due to the size of the protein. In such cases, multiple complimentary types of triple resonance experiments are needed to obtain sufficient information for making full assignments [63]. For very large proteins (> 250 resides), deuteration is usually necessary [57].

A large number of triple resonance pulse sequences have been created that allows for magnetization transfer between nuclei [63]. The magnetization pathway is reflected in the nomenclature for these experiments where the nuclei are listed by the frequency order in the pulse sequence. Spins are enclosed in parentheses when magnetization passes through spins with no frequency labeling. Amide protons are listed in the experiment name first when magnetization begins on that proton and

returns on the same proton for detection. This type of transfer is known as an 'out-and-back' experiment [63].

Complete sequential connectivities can be obtained using triple resonance experiments. To begin this strategy, the amide group (N_HH_N) chemical shifts of a spin system are correlated with the inter- and intra- residues chemical shifts of the backbone or sidechain atoms. The term *matching atom* is often referred to the atom that has shared correlations with an atom or atoms in it's own spin system and the spin system of a preceding residue [57]. For example, in the HNCA experiment, the amide nitrogen is coupled to the alpha carbon (Cα) in the residue (i) in which it is involved as well as to the Cα in the preceding residue (i-1) (Figure 1.8). Therefore, in this experiment Cα is the matching atom since both of these Cα nuclei are observed in the spectrum, with the Cα in its own residue showing up as a more intense peak due to a direct N-Cα coupling [64].

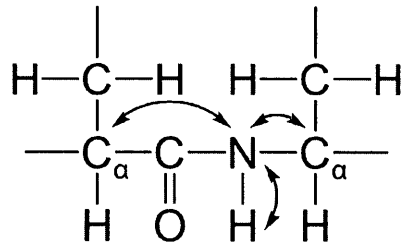


Figure 1.8: Correlations in an HNCA triple resonance experiment. In this experiment, Cα is the matching atom and both Cα nuclei appear in the spectrum. Adapted from [57].

Once the matching atom is identified, the amide of the residue *i* can be used to determine the frequency of the amide of the preceding residue. This process can be continued until all the backbone connectivities are obtained. Although this appears to be a very simple process, ambiguities can arise if a matching atom has a shift that is not easily distinguished in the protein sequence. Triple resonance experiments commonly used that are sufficient for protein structure determination include the HNCA, HNCACB, CBCACONH, HNCA, HNCACO and HNCO experiments (Figure 1.9) [57, 63].

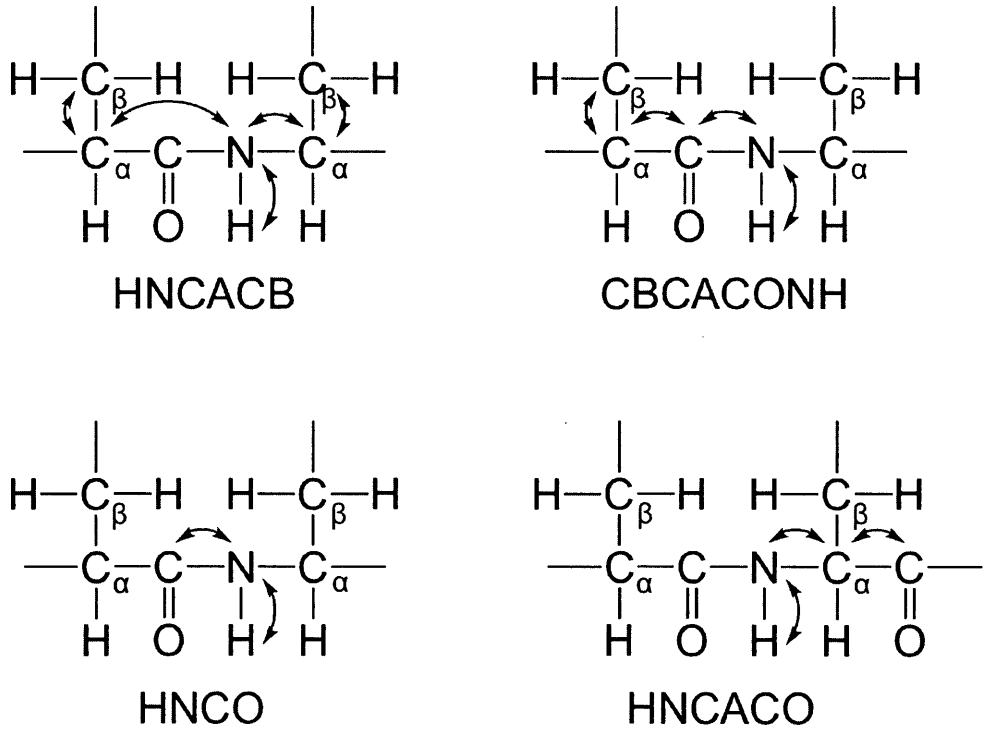


Figure 1.9: Correlations in common triple resonance experiments. Adapted from [57].

1.8 NMR RELAXATION IN PROTEINS

Following irradiation, the nuclear spins return to equilibrium along the z-axis of the external magnetic field in a process called *relaxation*. The resulting relaxation process is significant because it is largely responsible for the signal detected from the observed nuclei. In proteins, the relaxation rate constants for nuclei takes into consideration the large size of the macromolecule and thus is dependent upon various factors including internal motions, overall rotational correlation times, the geometrics of the nuclei itself, as well as the mechanistic pathway of the relaxation process [65]. Since these factors can vary considerably for individual macromolecules due to their size, relaxation behaviours and pathways can greatly differ. For instance, in larger proteins, the magnetization relaxes faster, which means that the detection time is less causing the resulting peaks to be broader and often much weaker.

To understand the relaxation process, the components of the nuclear magnetization as it behaves in the magnetic field must be considered. The spin-lattice or longitudinal relaxation, R_1, describes the mechanism by which the bulk nuclear spin magnetization (M_z) returns to equilibrium along the z-axis after being affected by a radio frequency. This mechanism is characterized by an associated relaxation time, T_1. For proteins T_1 is around one second and is around tens of seconds for small molecules. The relaxation rate can be related to the relaxation time by the equation $R_1 = 1/T_1$. The gain and loss of magnetization in the z-direction

contributes to line broadening due to the inherent lifetime of the spin states. T_1 is associated with the peak width at half height by the relationship in Eq. (1) [66]

$$\nu_{1/2} = 1/\pi T_1 = R_1/\pi \quad (1)$$

In the longitudinal relaxation process, energy is transferred from the excited state to the surrounding lattice, meaning that oscillation of the local magnetic fields in the lattice at the Larmor frequency allows for a feasible relaxation pathway [66]. Therefore, factors such as viscosity, temperature, molecular size, molecular structure, and internal forces such as Van der Waals attractions and hydrogen bonding can all affect the spin-lattice relaxation rates. Overall, T_1 is largely affected by the frequency of the nuclei (ω) as well as the time it takes the molecule to rotate by one radian, known as the global correlation time (τ_m) [58]. Two nuclear spins close to each in space can be considered dipoles and experience relaxation by dipole-dipole coupling, which is highly dependent on the global correlation time and can therefore significantly contribute to T_1 relaxation [67]. That is, the more rapidly a molecule tumbles, the less effective the dipole-dipole relaxation, and the longer the T_1. The contribution of the dipole-dipole relaxation mechanism to T_1 is related by Eq. (2)

$$R_{1(DD)} = 1/T_{1(DD)} = \gamma^2_X \gamma^2_H (h/2\pi)^2 \tau_c \Sigma (1/r_{H\text{-}X}^6) \quad (2)$$

where γ_X and γ_H are gyromagnetic ratios of ^1H and X nuclei respectively and $r_{H\text{-}X}$ is the internuclear distance [57, 66].

T_1 can be measured by performing an inversion-recovery experiment where a 180° radio-frequency pulse is used to flip the magnetization into the negative z-axis and then allowed to relax back into the positive z-axis (Figure 1.10). The time, τ, it takes for the spin to return from π (180°) to $\pi/2$ (90°) is measured. The experiment is cycled many times and τ is varied so that the resulting signal has different signs and intensities that can be plotted to observe the exponential growth of the peak intensity. The plotted peak intensities show the exponential signal growth from a short τ (negative) to long τ (positive) until the intensity becomes constant at long τ. The time at which the intensity becomes constant is identified as T_1 and indicates the time it takes for the magnetization to equilibrate along the z-axis [67].

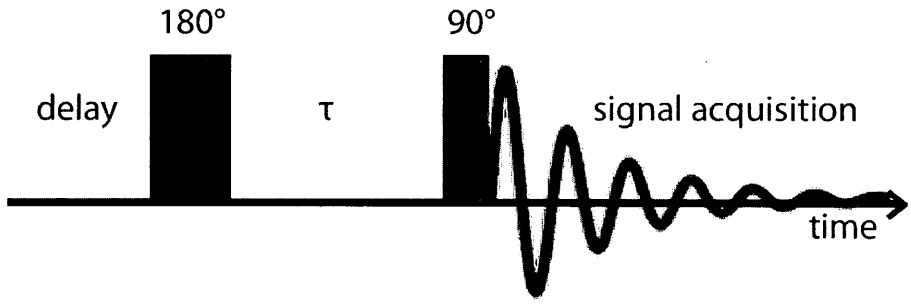


Figure 1.10: Inversion recovery experiment used to determine T_1. The spin system is first inverted with a 180° pulse, allowed to evolve during delay time τ, then given a 90° pulse, and finally a signal is acquired. The time between the 180° pulse and the 90° pulse is the inversion time. Adapted from [67].

Another type of relaxation process, known as spin-spin or transverse relaxation, R_2, describes the dephasing or loss of coherence of magnetization in the xy-plane as the spins precess around B_0. This process is a function of time, T_2, and has an associated relaxation rate of $R_2 = 1/T_2$ [67]. The loss of coherence is a result of a spin having different Larmor frequencies than its neighbouring spins so that eventually the spin decreases in magnetization (M_{xy}) and decays to zero at a time constant T_2. The loss of magnetization is such that T_1 can be equal to T_2, but T_2 cannot be greater than T_1 (i.e. $T_1 \geq T_2$). This is possible when magnetization in the xy-plane decreases faster than the increase of magnetization along the z-axis so that there is a loss of signal before the spins equilibrate [57]. During transverse relaxation, there is no involvement of the z-component of magnetization, energy is not lost to the surrounding lattice and the energy of the transitions does not change. Instead, energy is exchanged between the high and low spins and the level of spin ordering decreases such that there is an increase in entropy of the spin system and no enthalpy change. In macromolecules, the transverse relaxation rates are subject to molecular motions that occur in the microsecond timescale [67], [57].

T_2 can also be measured using Carr-Purcell-Meiboom-Gill (CPMG) experiments to enhance sensitivity of detection by measuring T_2 at constant time periods to obtain a complete dispersion profile in a short acquiring time. CPMG relaxation experiments use a spin-echo pulse sequence that has a 180° refocusing pulse to permit detection of conformational exchange times in the microsecond to millisecond timescales [57].

The heteronuclear NOE (hetNOE) is used to gain information about the molecular motion properties of two unlike spins that are coupled together such as ^{15}N-H and ^{13}C-H bond vectors. The hetNOE involves saturating the attached protons (^{1}H) and observing the ^{15}N signal intensities of the amides on the fast timescale (pico-nanoseconds). HetNOE experiments compare ^{15}N signals with and without proton saturation [68]. Protein regions that undergo fast motions in this timescale include flexible terminal ends and exposed surface loops that have below average hetNOE signals, which can be easily differentiated from rigid sections of a protein's structure. Negative signals are observed for unfolded regions that experience very fast motions [69].

1.9 PROTEIN DYNAMICS BY NMR

In addition to structure elucidation, NMR spectroscopy is a very useful technique for studying protein dynamics to provide insight into function, local folding and unfolding, drug binding and interactions with other bio-molecules such as DNA and RNA. Furthermore, NMR can be used as a tool for gaining information about the location and energetics of conformational transformations associated with these interactions [70].

NMR dynamics can be divided into two categories: fast timescales and slow timescales. The motions of proteins affect NMR parameters in such a way that motions are either faster or slower than the time it takes for one molecule to rotate

an average of one radian, defined as the rotational correlation time. Depending on the dynamics timescale, various types of motions are observable by NMR (Figure 1.11). Since NMR is sensitive to motions on a much slower timescale of 10^{-2} to 10^{2} seconds, NMR relaxation experiments measures protein dynamics on fast and slow timescales [71]. Fast timescales describe local flexibility in the picosecond (ps) to nanosecond (ns) range. Slow timescales represent collective motions and occur in microsecond (μs) to millisecond timescales (ms). Most often slower timescales are measured more directly and faster timescales are limited by the rotational correlation time. Depending on the timescales observed, the NMR experiments applied could be tailored to acquire data in the most appropriate manner.

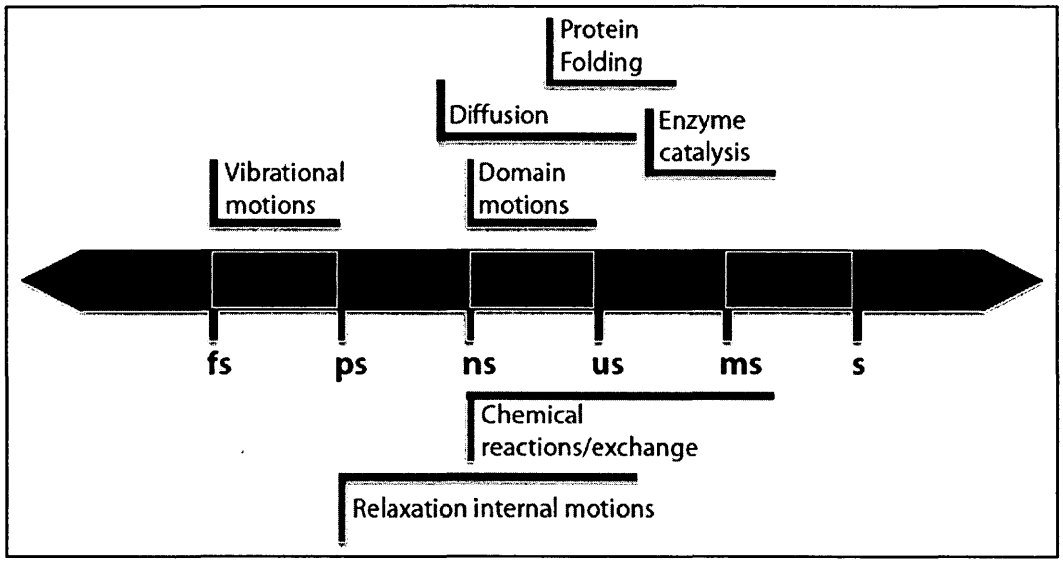


Figure 1.11: Timescales of protein motions observable by NMR spectroscopy. Adapted from [70].

Fast timescale dynamics measures the spin-lattice relaxation rate (R_1), the spin-spin relaxation rate (R_2), and the hetNOE. As the protein moves, the relaxation observed is a consequence of local molecular motions and global molecular tumbling that causes instability of the local magnetic field. These motions result in a loss of coherence (T_2 relaxation) and equilibrium restoration (T_1 relaxation). Therefore, relaxation depends on the probability of field fluctuations of each frequency arising from the fluctuations of the molecule, specifically motions of the N-H bond vector. This probability is known as the spectral density function, $J(\omega)$, Eq. (3) [72].

$$J(\omega) = \tau_m/(1 + \omega^2 \tau_m^2) \qquad (3)$$

where τ_m is the molecular global correlation time and ω is the nuclear Larmor frequency. Relaxation of the ^{15}N nucleus bonded to ^1H nuclei is based on dipolar interactions and the chemical shift anisotropy of the system, resulting in R_1, R_2, and hetNOE relaxation rates given by Eq. (4), Eq. (5), and Eq. (6) respectively [73]

$$1/T_1 = R_1 = d^2/4\,[J(\omega_H-\omega_N) + 3\,J(\omega_N) + 6\,J(\omega_H+\omega_N)] + c^2\,J(\omega_N) \qquad (4)$$

$$1/T_2 = R_2 = d^2/8\,[4J(0) + J(\omega_H-\omega_N) + 3\,J(\omega_N) + 6\,J(\omega_H) + 6\,J(\omega_H+\omega_N)] + (c^2/6)\,[4J(0) + 3J(\omega_N)] + R_{ex} \qquad (5)$$

$$\text{hetNOE} = 1 + (d^2/4)\,(\gamma_N/\gamma_H)\,[6J(\omega_H+\omega_N) - J(\omega_H-\omega_N)]\,T_1 \qquad (6)$$

where $d = (\mu_0 h\gamma_N\gamma_H/8\pi^2)\,(rNH^{-3})$, $c = \omega_N/\sqrt{3}\,(\sigma|| - \sigma\perp)$, h is Planck's constant, μ is the permeability of free space, ω_N and ω_H are the Larmor frequencies of ^{15}N and ^1H

respectively, γ_N and γ_H are the gyromagnetic ratios of ^{15}N and ^{1}H respectively, $\sigma||$ and $\sigma\perp$ are the parallel and perpendicular components of the axially symmetric ^{15}N chemical shift tensor (^{15}N CSA value of 160 ppm), r_{NH} is the bond length with typical values of 1.02 or 1.04 Å, and R_{ex} is the contribution to R_2 from motions slower than the overall molecular motion.

Once relaxation data is successfully acquired, methods for analyzing relaxation rates include spectral density mapping [74] and the model free approach [75, 76]. Spectral density mapping incorporates the direct relationship between the spectral density function and the R_1, R_2 and hetNOE relaxation rates [74]. The model-free approach provides information about internal motions in proteins without forcing the R_1, R_2, and hetNOE relaxation data into any particular analysis model and the value of S that describes the freedom of rotational motion about the N-H bond vector is calculated without a model as well [76]. However, the model-free approach requires assumptions about molecular motions and about the correlation function. These limitations are not present in spectral density mapping [77].

1.10 MACROMOLECULAR CRYSTALLOGRAPHY

Aside from NMR spectroscopy, the technique of X-ray crystallography is one of the main methods used for three-dimensional structure determination of macromolecules such as proteins. Although X-ray crystallography can only observe structure in a static crystal state, the benefit of this technique over NMR spectroscopy is simple. X-ray crystallography is not prohibited by size, which is very advantageous for studying large macromolecules that would otherwise exceed the size limit for NMR [78].

To solve a structure by X-ray crystallography, a crystal of the macromolecule of interest must first be produced. A crystal is composed of structural units of atoms, ions, or molecules that are a regular periodic arrangement. In this arrange of the crystal lattice, the smallest regular repeating unit is known as the unit cell. The unit cell is defined by its lattice dimensions (a, b, c) and lattice angles (α, β, γ) (Figure 1.12). Symmetry operations such as rotations, translations, and reflections, are performed on the crystal lattice unit cell that allow for a copy of the original unit cell to be moved in space. This copy is called the asymmetric unit [78, 79].

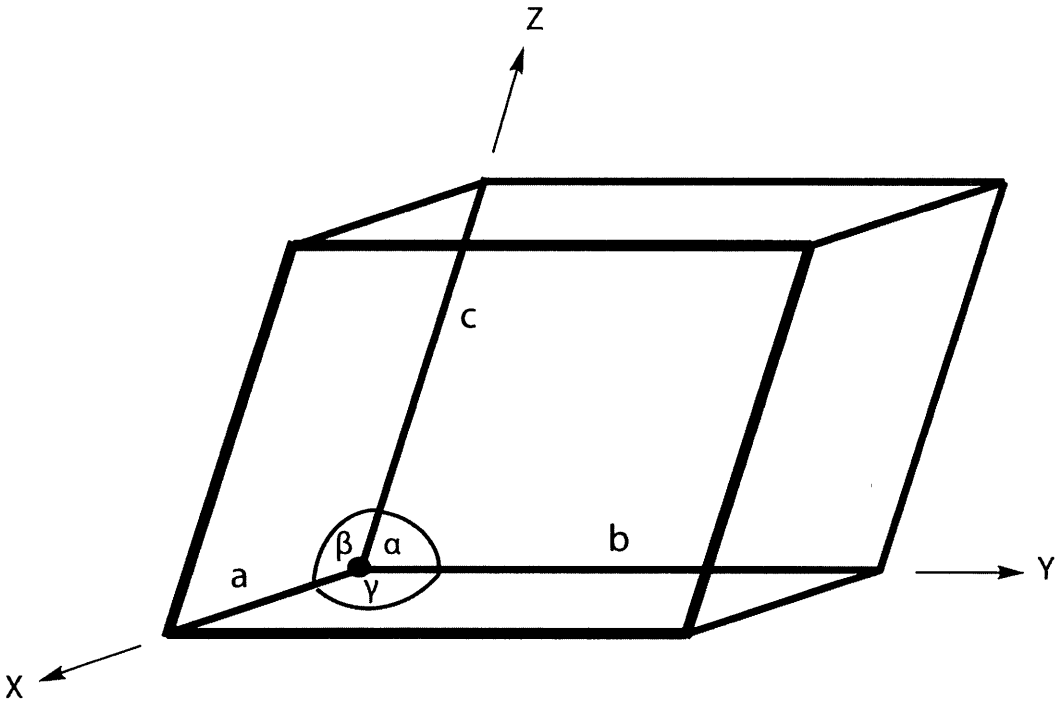


Figure 1.12: The unit cell in a lattice crystal defined by dimensions a, b, c and angles α, β, γ. The origin is denoted by a dot at the vertex. Adapted from [79].

For protein crystallography, the crystal must be pure. This means that the purified and isolated protein must be a homogeneous sample that only contains the protein of interest. In general, standard protein purification methods such as column chromatography are sufficient for removing impurities and obtaining milligram quantities (usually 2-50 mg/mL) of pure sample. The difficulty in protein crystallography usually occurs at the crystallization step when a series of crystal screens are set up and optimal protein concentrations, salt, pH, and temperature

conditions must be met for ordered crystal formation and growth. Crystallization kits with pre-made buffer solutions are available for initial crystallization screening, upon which further optimization is required [79].

Since proteins are composed of amino and carboxylic functional groups that are readily protonated or deprotonated when in an acidic or basic environment respectively, proteins are often quite sensitive to small pH changes, causing proteins to crystallize over a small pH range in a select amount of buffers [80].

The solubility of the macromolecule is important for crystallization to occur. The macromolecule must be pulled out of a soluble state into a nucleation state that promotes nuclei formation from which a crystal will grow (Figure 1.13). In terms of phases, crystal nucleation and growth occur in the region of supersaturation. To reduce the solubility of the protein and control precipitation, a precipitant is included in the buffer system [80].

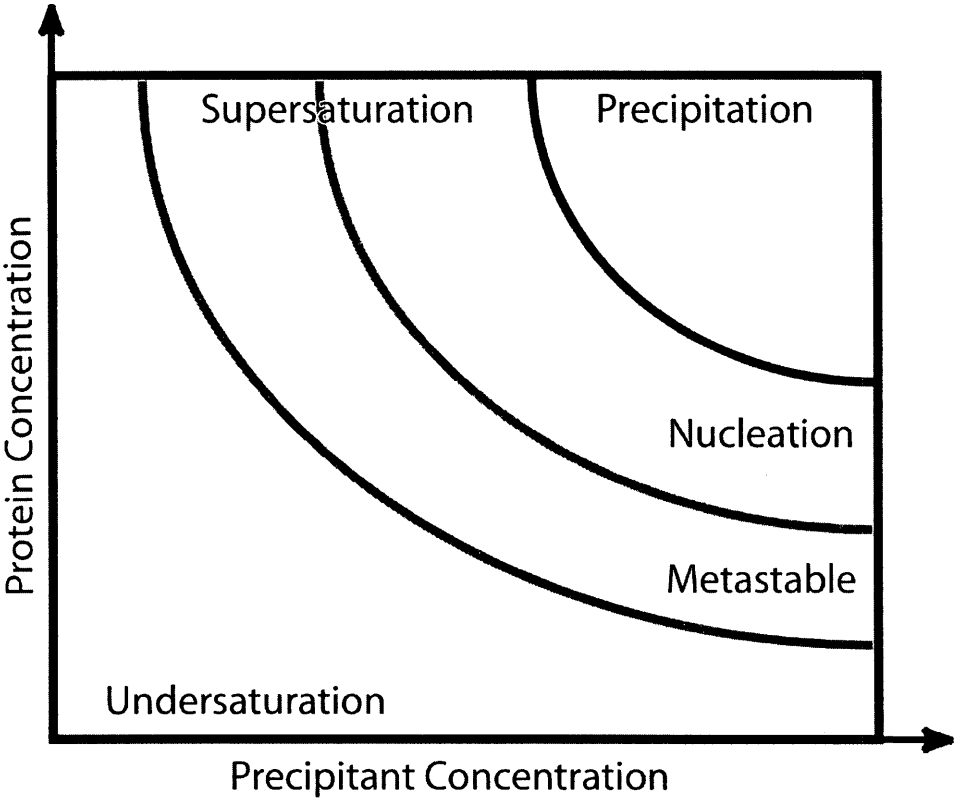


Figure 1.13: Protein crystallization phase diagram. In the area of undersaturation the protein and/or precipitation concentrations do not produce crystal growth. In the precipitation zone, the protein comes out of solution as an aggregate and is not useful. In the area of nucleation, crystal nucleation occurs and initial crystal growth is facilitated. As the crystal grows, protein concentration is depleted and protein crystallization moves to the metastable (growth) zone. Adapted from [79].

Two types of vapor diffusion methods are commonly used for achieving crystallization. The hanging drop technique, which is the most used method, involves mixing a few microlitres (usually 1-2 μL) of protein with an equal ratio of reservoir (buffer) solution containing the precipitant. This protein/precipitant mixture is placed on the surface of a cover slip that covers the reservoir. The buffer solution is also added to the reservoir in a larger volume. Since the protein/precipitant mixture in the drop is less concentrated than the reservoir solution, water evaporates from the drop into the reservoir. As a result, the concentration of the protein and precipitant in the drop increases, allowing for protein crystal formation. The sitting drop method follows the same principles as the hanging drop method with the exception that the droplet is positioned on a pedestal or ledge above the reservoir (Figure 1.14). It is important to note that both the hanging drop and sitting drop methods require a closed system that is achieved by sealing the cover slip to the top of the reservoir with a high vacuum grease [79]

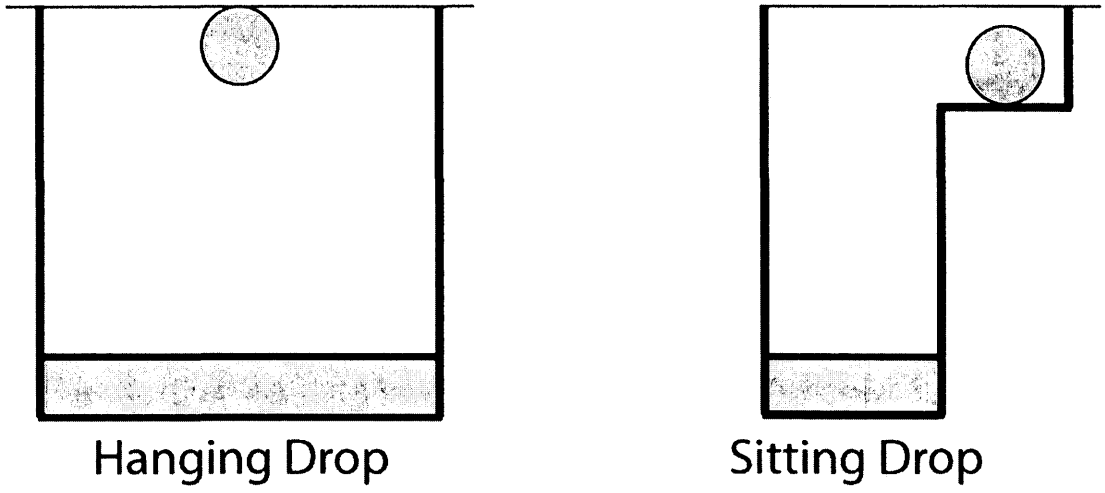


Figure 1.14: Two types of vapor diffusion methods commonly used for protein crystallization. Adapted from [79].

Once a protein crystal of suitable order and size is achieved, the crystal must be mounted in preparation for diffraction. To begin this process, a droplet of mother liquor with the protein crystal of interest is added to the surface of a plastic cover and placed under a microscope. A crystal mounting tool, such as a cryogenic loop, is used to select the protein crystal from the droplet. Once mounted, the crystal is immediately cryoprotected with liquid nitrogen and stored in a dewar until data is collected by a diffractometer or at a synchrotron [79].

At the synchrotron or diffractometer, an X-ray beam is projected at the crystal and thereby scattered to produce a diffraction pattern that provides

information about the lattice planes. This diffraction event is defined by Bragg's Law (Figure 1.15) [78].

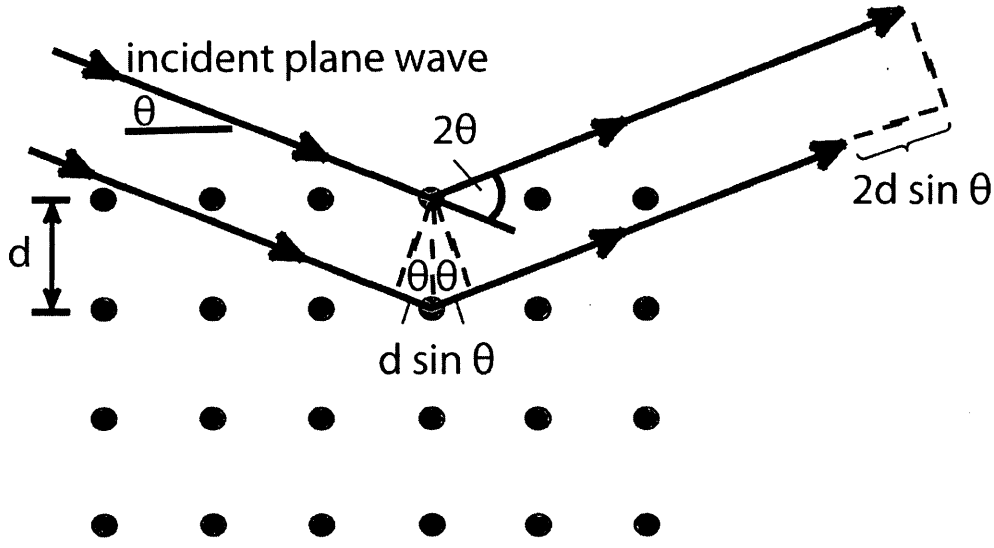


Figure 1.15: Bragg's Law scattering. Crystal scattering occurs in all directions, however it is only visible in a finite number of directions that obeys Bragg's Law. Adapted from [78].

Although scattering from the crystal occurs in all directions, scattering is only visible in a finite direction by wavelengths of plane waves that do not interfere with one another. Constructive interference occurs when $n\lambda = 2d\sin\theta$. Perpendicular to the diffraction planes is a diffraction vector (S) with length that is the reciprocal to the spacing between the diffraction planes (1/d). The diffraction vector provides a

directional component to the reciprocal space unit cell parameters a*, b*, c* as described in Eq. (7) [78]

$$S\,(hkl) = ha^* + kb^* + lc^* \qquad (7)$$

where h, k, l are Miller indices and each index describes a plane orthogonal to a direction. The diffraction process can be understood in visual terms by the Ewald's Sphere, which incorporates the scalar (Bragg's Law) and vector (Miller indices) components of diffraction (Figure 1.16) [78].

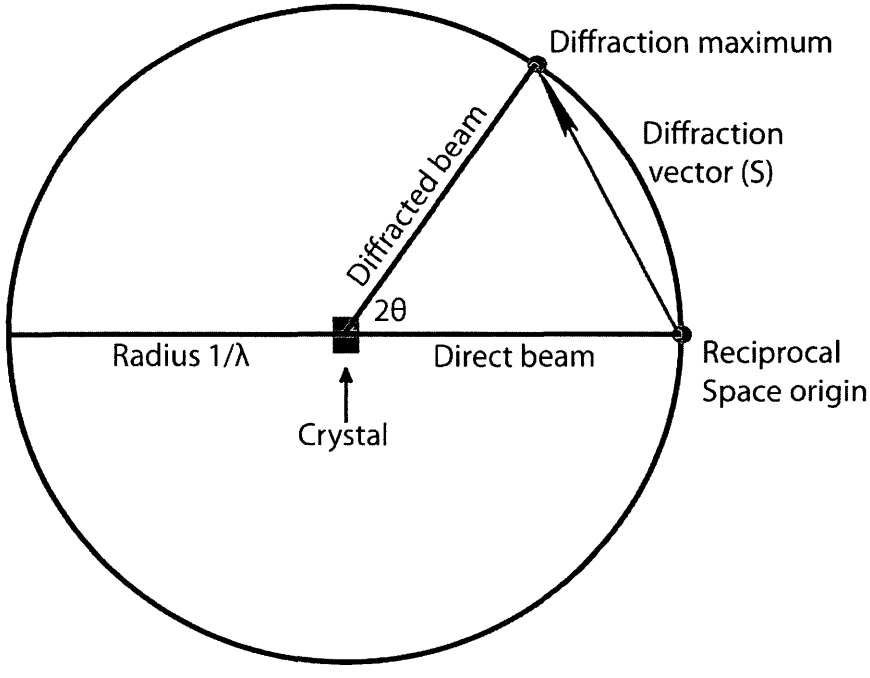


Figure 1.16: Ewald's Sphere describes the requirements for diffraction as a graphical representation. For a diffraction point to adhere to the diffraction conditions, the diffraction point must lie on the surface of Ewald's Sphere. Adapted from [78].

In the Ewald Sphere construction, the crystal sits in the center of the sphere then the X-ray beam enters from the left and is scattered by the crystal. The beam that passes directly through the crystal to the surface of the sphere is called the origin of reciprocal space. The diffracted beams have an angle of 2θ from the origin of reciprocal space. In order for a diffracted beam to appear in the crystal's diffraction pattern, the diffraction point must lie on the surface of the Ewald's Sphere, thus satisfying the diffraction condition. By using this diagram, it is possible to determine which points satisfy to the diffraction condition, making data collection easier to interpret [78].

1.11 SMALL ANGLE X-RAY SCATTERING (SAXS)

A wealth of information about a protein's three-dimensional structure can be gained from the high-resolution techniques of NMR spectroscopy and crystallography. In addition to both these methods, small angle X-ray scattering (SAXS) is an established low-resolution technique that is commonly used as a fundamental structural analysis tool for the study of biomolecules to obtain structural information about the shape and size of a protein [81-83]. Unlike crystallography, a crystalline sample is not needed for SAXS experiments and the sample is not destroyed in the process. Although this is an advantageous quality for protein samples that are difficult to crystallize, the loss of information due to the random orientation of dissolved or partially ordered molecules, compared to

crystallography, is a disadvantage [81]. Another benefit of SAXS is that unlike NMR methods, SAXS experiments are not limited by biomolecular size. Most often SAXS is employed to compliment NMR and crystallography methods and by doing so, useful structural information can be obtained [81, 84].

In a standard SAXS experiment, a monochromatic and collimated X-ray beam illuminates macromolecules to produce a scattering pattern that is recorded by an X-ray detector. Intensity (I) of the scattering is isotropic and is dependent on the scattering angle 2θ [81]. Measurements of the buffer serve as a blank and the scattering intensities of the buffer are subtracted from the scattering intensities of the sample. The result is a scattering profile that is expressed as a function of the momentum transfers $s = 4\pi\sin\theta/\lambda$, where λ is the wavelength. The value of "s" is proportional to the solute concentration and to the scattering from a single particle when all orientations are averaged. The scattering profile provides information about the main geometrical parameters of the particle, such as molecular mass (MM), hydrated particle volume (V), specific surface area (S), the maximum distance within the particle (D_{max}), and the folding state of the particle. Most significantly, the one-dimensional scattering profile provides useful information about the three-dimensional macromolecular structure. The low-resolution (1.5-2nm) capability of SAXS allows for *ab initio* particle shape predictions to be produced from the scattering pattern intensity [81]. It must be noted that multiple shapes may be compatible with the same scattering profile; however, this can be limited by

including restrictions to the parameters used when composing the macromolecule model.

The SAXS method is often used to identify biologically active conformations and oligmeric compositions, to determine flexible systems quantitatively, and to corroborate the structural similarity between crystal and solution states by quickly computing *ab initio* biomolecular shapes and sizes [81-83, 85]. Because of such applications, SAXS remains to be a useful technique for studying biomolecules that are well behaved in solution.

1.12 ISOTHERMAL TITRATION CALORIMETRY (ITC)

In addition to NMR spectroscopy, biomolecular interactions can be further characterized from an energetic perspective using the thermodynamic technique of isothermal titration calorimetry (ITC). ITC directly measures the heat released and absorbed during a biomolecular binding interaction to generate accurate thermodynamic parameters, making it possible to determine the binding constant (K_a), reaction stoichiometry (n), entropy (ΔS) and enthalpy (ΔH) values. The result is a complete thermodynamic profile of a biomolecular interaction [86].

The versatility of ITC allows common interactions to be studied easily and quickly, including protein-protein, protein-nucleic acid, protein-small molecule, and antibody-antigen interactions. In addition, a range of applications can be studied

including binding characterizations, enzyme kinetics, biological activity assessment, and studies of the effects of structural changes on binding mechanisms [87].

A major advantage of using this analytical technique is that K_d binding affinity values can be accurately measured in the millimolar to nanomolar range in a single experiment. Binding constants as large as 10^8-10^9 M^{-1} can be measured and heat changes as small as 0.4 μJ (0.1 μcal) can be detected with considerable accuracy. Furthermore, single and multiple binding sites can also be detected [88].

The ITC apparatus manufactured by Microcal (USA) is designed to titrate a solution of biomolecule (A) from a syringe into a cell containing a solution of the associated binding ligand (B) (Figure 1.17) [89]. Most often the small molecule is titrated into the macromolecule, but this is not a rule.

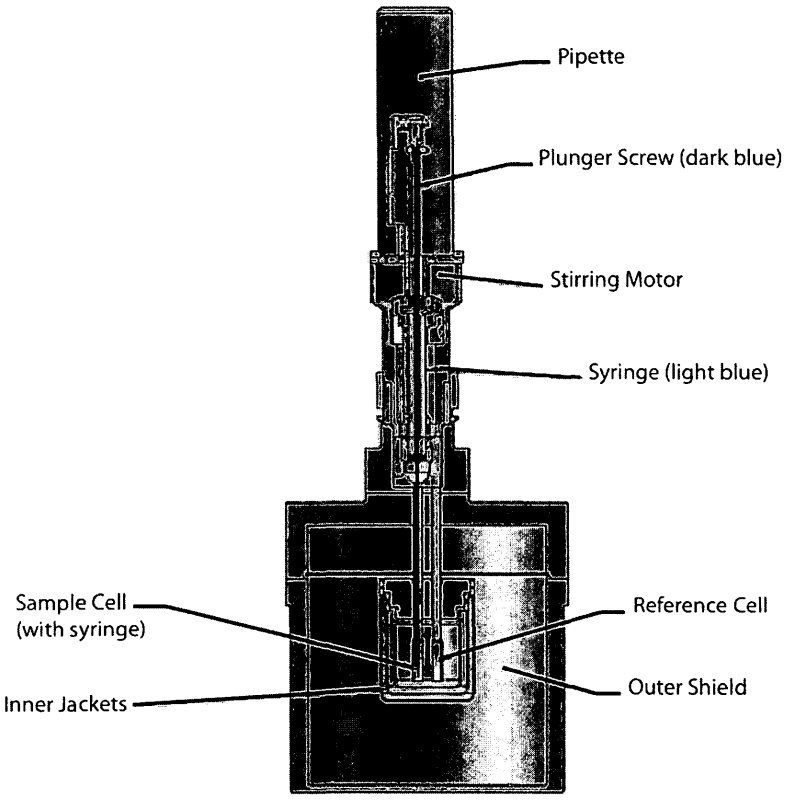


Figure 1.17: Isothermal titration calorimetry (ITC) apparatus. Adapted from [89].

The heat released or absorbed from the binding interaction upon titration is then monitored over a period of time with the raw data represented as peaks on a chromatogram (Figure 1.18, top panel). The largest ITC signal is generated when A is added to B because all of B is available for binding and a significant amount of A is bound. As less B becomes available, the ITC signal diminishes. The amount of heat absorbed or released during the titration is directly proportional to the amount of

binding such that when B is fully in complex to A, there is no B available for binding and no heats of reaction are observed. Instead, heats of dilutions can be subtracted from the data if needed. Integration of the heats produced from each titration generates a binding isotherm curve (Figure 1.18, bottom panel) that is plotted against the heats from each injection and the molar ratio of the binding partners. The curve is fit to either a one-site or multiple-site binding model.

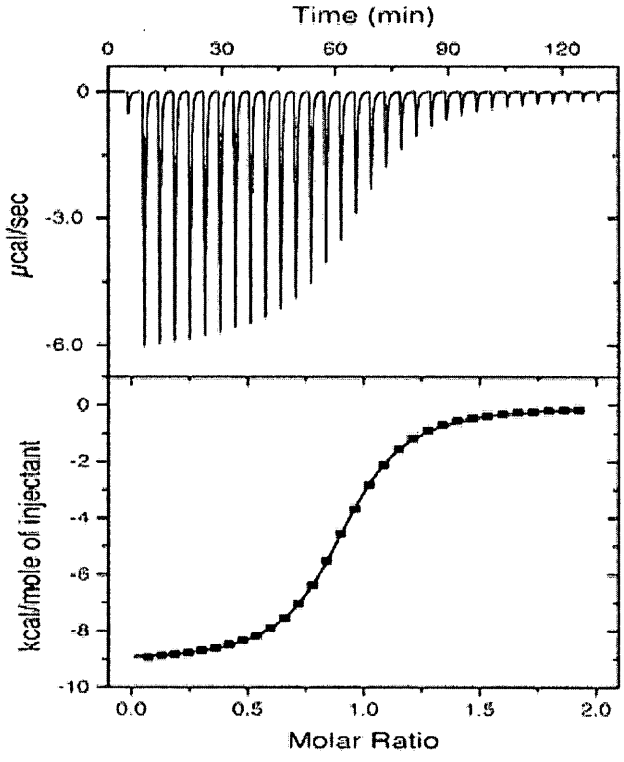


Figure 1.18: Output given by the ITC apparatus. The top panel is a graph of the power supplied per unit time recorded in real time. Integrating the data in the top panel provides the heat per mole per injection with respect to the molar ration, as shown by the bottom panel. Adapted from [89].

The binding model selected is representative of the binding process used to calculate K_a, ΔH, and n. From these thermodynamic parameters, the change in free energy (ΔG) can be calculated using Eq. (8)

$$\Delta G = -RT \ln K_a \qquad (8)$$

where R is the gas constant and T is the temperature of the system in Kelvin. From this, the change in entropy can be determined from Eq. (9)

$$\Delta G = \Delta H - T\Delta S \qquad (9)$$

Spontaneous ligand-macromolecule interactions are when ΔG is negative, which can occur during exothermic reactions when ΔH is negative and ΔS is positive.

Information about the binding affinity, K_d, can be determined using Eq. (10)

$$K_d = 1/K_a \qquad (10)$$

Determination of binding affinity is an important quantitative aspect when considering molecular interactions [89].

1.13 GEL AFFINITY ELECTROPHORESIS

Macromolecules are commonly separated and analyzed using electrophoresis. This technique separates molecules based on electric charge and

electrophoresis analysis has been shown to generate precise quantitative data. Investigating the interaction between protein drugs and their molecular targets is particularly useful when developing stable protein pharmaceuticals [93]. For example, binding constants have been estimated for the interaction between the immunosupressive drug cyclosporine A (CsA) to the enzyme cyclophilin (Cyp) [94].

CHAPTER 2

A STRUCTURAL STUDY OF *C. ELEGANS* RDE-4, A DUAL DOUBLE STRANDED RNA-BINDING PROTEIN INVOLVED IN RNA INTERFERENCE

2.0 RDE-4: THE DUAL DOUBLE STRANDED RNA-BINDING PROTEIN

The usage of short RNA molecules as gene regulators has gained significant interest since the discovery of siRNA [29]. Small interfering RNAs (siRNAs) are a class of double stranded RNA that is cleaved into short sequences of approximately 22 nucleotides in length, and have an integral role in RNA interference (RNAi). The process of RNAi uses siRNA to bind to a complimentary target mRNA, thus tagging the mRNA for degradation and ultimately controlling gene expression [29]. For this reason, RNAi is critical for our understanding of cell development, cell growth, and consequently for our understanding of the molecular basis of cancers.

Double stranded RNA binding proteins (dsRBPs) in the RNAi pathway act in concert with the endonuclease, Dicer-1 (DCR-1), to cleave exogenous or endogenous double stranded RNA into short interfering RNA (siRNA) [95]. The antisense strand of the siRNA is bound to the RNA induced silencing complex (RISC) and subsequently base paired to a strand of mRNA, causing strand degradation. In *C. elegans*, the dsRBP component in the DCR-1 complex is RDE-4, which contains two dsRNA binding domains (dsRBDs) that are each composed of a conserved $\alpha\beta\beta\beta\alpha$ fold. RDE-4 is necessary to process the trigger RNA into siRNA, most likely through binding the dsRNA and recruiting it to the DCR-1 complex (Figure 2.1) [96].

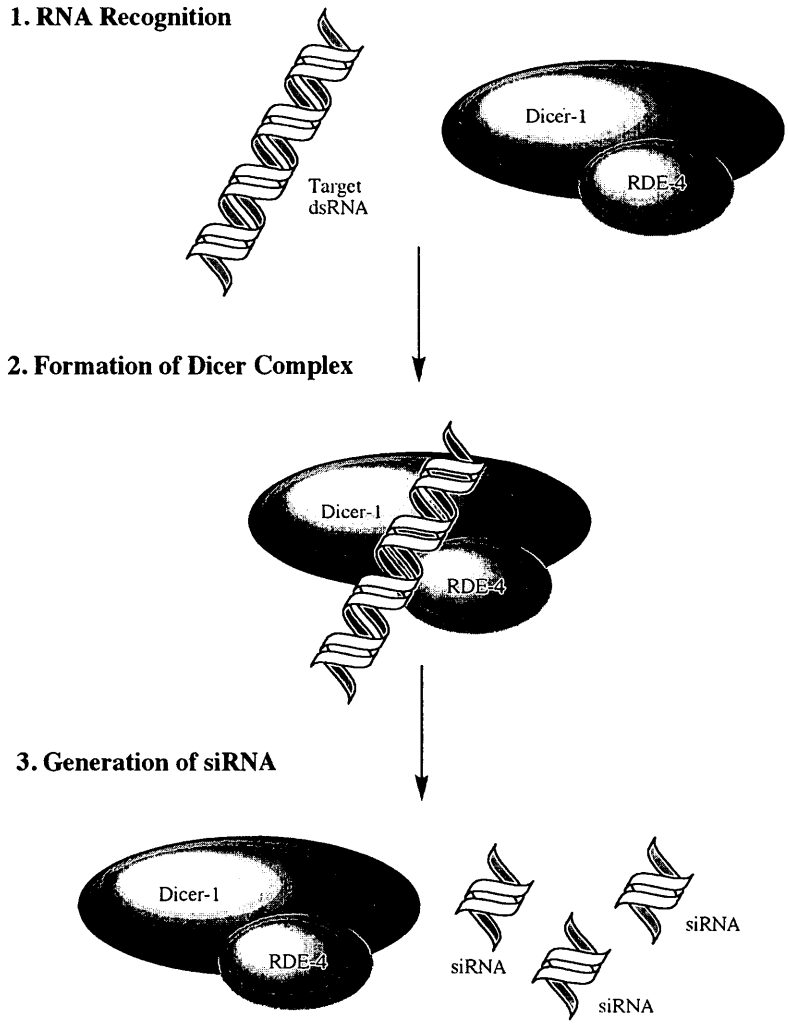


Figure 2.1: The early events of RNA interference. 1) The dual dsRNA binding protein RDE-4 (green) forms a complex with the Dicer endonuclease (grey) to bind the target dsRNA. 2) The Dicer complex cleaves the target dsRNA into siRNA molecules. 3) The siRNA molecules are available for incorporation into the RNA induced silencing complex (RISC) where it can perform gene regulation tasks.

RDE-4 contains a dual double-stranded RNA binding domain (dsRBD), each predicted to be composed of a αββββα motif (Figure 2.2). The full protein sequence of RDE-4 is 385 amino acids (aa) long. The first dsRBD is located from 43-105aa and the second dsRBD is located from 169-236aa. A linker region of approximately 63aa separates the two dsRBDs. In addition, the C-terminal (254aa-385aa) is necessary for the dimerization of the protein, resulting in the formation of a stable homodimer. Although the C-terminal is not required for dsRNA binding or binding with increased affinity, it is believed to facilitate the formation of an active Dicer complex [96]. In the case of Rnt1p, a dsRBD containing protein in the RNase III family, the helical region that flanks the main dsRBD has been found to control the binding specificity of the protein [97, 98].

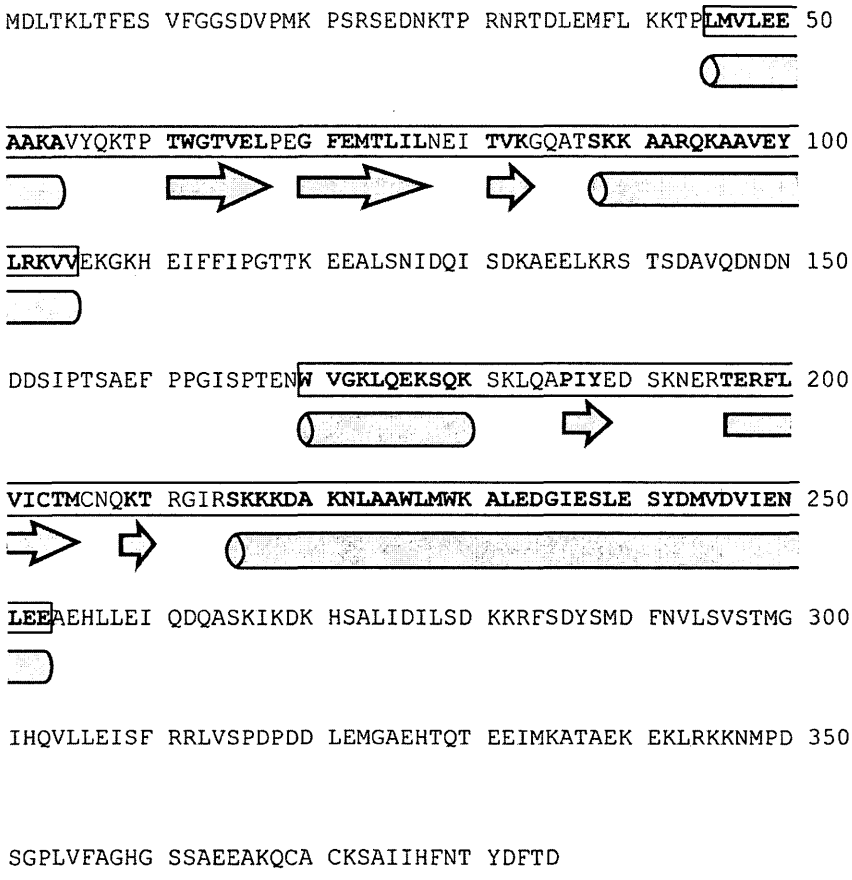


Figure 2.2: The αββββα motif present in RDE-4. The cylinders outlined in red represent α-helices, the arrows outlined in blue represent the β-strands, the boxed regions indicates the two dsRBDs (based on PSIPRED protein structure prediction program).

The characterized function of the αββββα motif in RDE-4 is to bind structured RNA strands, more specifically dsRNA molecules. Although RDE-4 can bind short dsRNA, the affinity for long dsRNA is much higher. For example, RDE-4 can bind a 650 base-pair (bp) dsRNA with a K_d of ~23 nM compared to a 20 bp dsRNA with a K_d of ~1.5 μM [96].

Structural studies performed by Chang *et al.* [99], reveal that the dsRBD of RNA binding proteins can bind across approximately 15-16 RNA base pairs by interacting with the 2'OH groups and phosphates groups of the successive minor, major, and minor grooves in an RNA duplex (Figure 2.3). This binding interaction is believed to apply to the dual dsRBD of RDE-4 [99]. Both dsRBDs of RDE-4 are required for RNA binding and RDE-4 can bind dsRNA independently of DCR-1 with high affinity for long dsRNA (≥40 base pair), but not short dsRNA molecules. This binding specificity is consistent with the function of the DCR-1 complex to bind the trigger dsRNA, but not the product siRNA molecules [96].

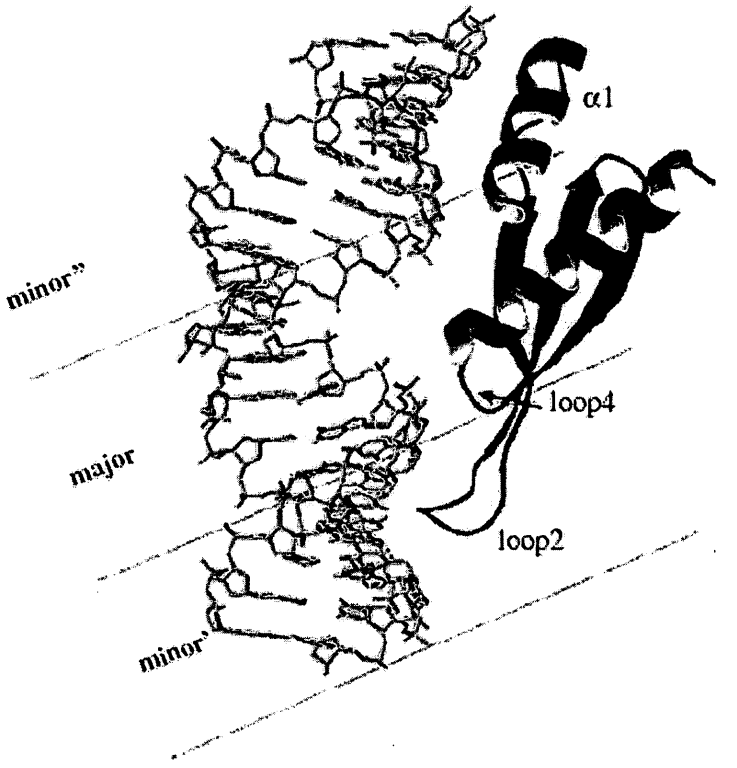


Figure 2.3: The binding interaction of dsRBD and dsRNA in *Xenopus laevis* RNA binding protein A. The first α-helix in the dsRBD, loop4, and loop2 are able to interact with the minor, major, minor grooves of the dsRNA [99].

It must be noted that many different dsRNA-binding proteins in different RNAi pathways in various organisms have been discovered, with most of these proteins identified to contain a dsRBD. In most cases, only a single copy of the dsRBD is present. However, proteins that have two dsRBD regions have been reported, including adenosine deaminase (ADAR) [100] and protein kinase (PKR)

[101]. The two dsRBD structures in both these proteins do not interact with each other, thereby operating as independent structured units. When the structure of the protein involved in human RNAi, DGCR8, was reported in 2007, DGCR8 became the first dual dsRBD protein in RNAi to have its structure determined as well as the first example of a dual dsRBD protein with two dsRBD regions that interacted with each other to create a single ordered structure [102]. The structure of DGCR8 that was reported identified the conserved αββββα motif in both dsRBD regions. In addition, a helix in the region between the two dsRBD structures and another helix at the C-terminus was found to interact with the second helix in the first dsRBD to create a structured single unit of the tandem dsRBDs (Figure 2.4).

Figure 2.4: Domain structures and sequence alignments of dual dsRBD family of proteins, *Drosophila* R2D2, *C. elegans* RDE-4, and Human DGCR8. In panel A, the domain location of the dsRBDs and C-terminal domain (CTD) within the protein family is shown. In panel B, the sequence alignment of the proteins is shown where "H" and "E" indicate α-helix and β-strand secondary structures contained in the dsRBD regions of DGCR8 respectively, and "h" and "e" indicate the α-helix and β-strand secondary structures outside of the dsRBD regions of DGCR8. Residues conserved in all three family members are indicated by "*", residues conserved in two of the three family members are indicated by ":", and residues with similar function are indicated by ".".

2.1 RESEARCH GOALS

The purpose of this project was to understand how the Dicer complex recognizes the trigger RNA by determining the structures of the dual dsRBD-containing protein and studying how they interact with RNA. More specifically, this project aims to understand how the structure of the dual dsRBD interacts with dsRNA by identify if the dual dsRBD of the *C. elegans* RDE-4 protein forms a single structured unit to allow for binding to long target dsRNA and not to short product dsRNA molecules. Answering this question will provide a structural explanation for how this binding specificity is achieved in the RNAi pathway. Proteins that contain dsRNA-binding domains are necessary during the initial steps of RNAi to process the trigger dsRNA into siRNA products [95]. However, the mechanism by which the dsRNA binding protein is capable of binding the dsRNA is not well understood. Furthermore, this research aimed to investigate which residues in the protein sequence interact with RNA and are important for RNA binding to occur, and to provide insight into the tertiary structure of these dsRBDs. To achieve these goals, the function and structure of the dual dsRBDs of RDE-4 was investigated through the techniques of NMR spectroscopy, small angle X-ray scattering (SAXS) data, isothermal titration calorimetry (ITC), and gel shift assay.

The ability of RNAi to silence the activity of specific genes has directed the use of RNAi to being a therapeutic modality [103]. Understanding the RNAi

mechanism will certainly prove advantageous towards the progression of RNAi-based therapies.

2.2 MATERIALS AND METHODS

2.2.1 Assembly of RDE-4 dual dsRBD Protein Truncations

(i) Truncations Made Via Assembly PCR (RDE-4^{38-268})

Oligomers of two selected regions (22aa-268aa and 38aa-268aa) of the RDE-4 DNA sequence were determined using Assembly PCR Oligo Maker [104] and were provided by Sigma-Aldrich. The oligomers were subjected to assembly polymerase chain reaction (PCR) in order to obtain the two template DNA regions of interest. The assembly PCR products were analyzed on a 1.5% agarose gel for confirmation of size. The assembly PCR product corresponding to the template DNA encoding for RDE-4^{22-268} was amplified by PCR using Vent Polymerase and the forward primer 5'-CCGGGAATTCTCCCGTTC-3' (Sigma-Aldrich) and the reverse primer 5'-CGGCCAAGCTTTTTAATCTTGG-3' (Sigma-Aldrich). Similarly, the template DNA encoding for RDE-4^{38-268} was amplified by PCR using the forward primer 5'-CCGGGAATTCATGTTTCTGAAGA-3' (Sigma-Aldrich) and the reverse primer 5'-CGGCCAAGCTTTTTAATCTTGG-3' (Sigma-Aldrich). All forward and reverse primers included 3' EcoRI and 5' HindIII restrictions sites, respectively. The PCR products were analyzed on a 1.9% low melt agarose gel. The DNA bands of interest were

excised from the gel and the DNA extracted using the Qiagen QIquick DNA Gel Extraction Kit. Double digestion reactions using EcoRI and HindIII restriction enzymes were performed on the inserts and on a pET23a vector. This involved a 3-hour incubation time followed by heat inactivation of the restriction enzymes at 65°C for 20 minutes. To prevent the digested vector from self-ligating, the 5' hydroxyl group of the DNA was removed by incubating with Antarctic Phosphatase for 1 hour at 37°C, then heat inactivating the enzyme at 65°C for 15 minutes. Each insert was ligated to pET23a vectors by incubating the reaction with T4 DNA Ligase. The ligated plasmids were transformed to DH5αF' *E. coli* competent cells by heat shocking and incubated overnight at 37°C on kanamycin agar plates. Colony PCR was used to determine if the cells contained the inserts of interest. The colony PCR reactions were analyzed by 1.9% agarose gel, which determined that only the insert of RDE-$4^{38\text{-}268}$ was successfully ligated to the vector. The Sequencing Lab in the Core Facility at York University sequenced the DNA of this colony. After receiving positive sequencing results, the plasmid was transformed via heat shocking to BL21 (DE3) *E. coli* expression cells.

(ii) Truncations Made Via GenScript (RDE-$4^{28\text{-}253}$, RDE-$4^{28\text{-}137+165\text{-}253}$, RDE-$4^{253\text{-}385}$)

The following three regions of the full RDE-4 DNA sequence were obtained as plasmids (in pET15b vectors) and as Top10 stab cultures by GenScript: RDE-$4^{28\text{-}253}$, RDE-$4^{28\text{-}137+165\text{-}253}$, and RDE-$4^{253\text{-}385}$. These truncations contained the dual dsRBDs, the dual dsRBDs without the region of residues that links the two dsRBDs together,

and the C-terminus respectively. The plasmids were transformed to BL21 (DE3), and BL21 (pLysS) *E. coli* competent expression cells by heat shocking and incubated overnight at 37°C on ampicillin LB agar plates.

2.2.2 Protein Expression of RDE-4 Protein Truncations

Expression tests were performed on all truncations mentioned in section 2.2.1 *(i)* and *(ii)* that were transformed to BL21 (DE3) expression cells. To determine expression, the cells were grown in small-scale culture growths of 5 mL TYP growth media with 5 μL of the appropriate antibiotic in the shaker at 37°C, and then induced with 1.0 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) when the $O.D_{600}$ reached 0.8. The uninduced and induced samples were analyzed on 15% SDS-PAGE gels. Once successfully expression was confirmed, large-scale growths of the each domain were made in two 1 L flasks of TYP media containing 1 mL per flask of the appropriate antibiotic. The cells were induced with 1.0 mM IPTG when $O.D_{600}$ reached 0.8, then incubated for 4 hours at 37°C in the shaker. After the incubation period, the media was centrifuged for 20 minutes at 5,000 rpm. The supernatant was discarded and the pellet resuspended in wash buffer (30 mM Tris pH 8, 250 mM NaCl, 5 mM BME, 5% Glycerol, 0.02% NaN_3).

2.2.3 Protein Purification of RDE-4 Protein Truncations

The cells were sonicated using a Fisher Scientific Model 500 Sonic Dismembrator (6 cycles: 30 seconds pulse on, 30 seconds pulse off) and centrifuged

at 15,000 rpm for 20 minutes. Both the supernatant and pellet were separated and retained. The pellet was resuspended in 15 mL of 6 M guanidine hydrochloride (GdnHCl) and allowed to gently shake overnight at 4°C. Following centrifugation at 15,000 rpm for 30 minutes, the supernatant was passed through a 5.00 μm filter then a 0.45 μm filter. The supernatant was purified using a nickel affinity column under denaturing conditions (denaturing buffer: 6 M GdnHCl, 30 mM Tris pH 8, 250 mM NaCl, 5 mM BME, 5% (v/v) Glycerol, 0.02% (w/v) NaN_3; wash buffer: 30 mM Tris pH 8, 250 mM NaCl, 5 mM BME, 5% (v/v) Glycerol, 0.02% (w/v) NaN_3; elution buffer: 400 mM imidazole, 30 mM TrisHCl pH 8, 250 mM NaCl, 5 mM BME, 5% (v/v) Glycerol, 0.02% (w/v) NaN_3). Fractions containing the protein of interest were collected and analyzed by 15% SDS-PAGE gel. The fractions containing the purified protein of interest were concentrated using a 10,000 molecular weight cut-off (MWCO) Amicon concentrator and the buffer exchanged to a Q anion exchange column wash buffer (20 mM Tris pH 7, 0.5 mM EDTA, 0.5 mM DTT, 0.02% (w/v) NaN_3) for further purification. The resulting sample (~1.5-2.0 mL) was loaded and purified on a Q anion exchange column using the same wash buffer as in the sample and an elution buffer containing 1 M NaCl, 20 mM Tris pH 7, 0.5 mM EDTA, 0.5 mM DTT, and 0.02% (w/v) NaN_3. The fractions containing the protein of interest were collected and analyzed on a 15% SDS-PAGE gel and concentrated to 500 μL using a 10,000 MWCO Amicon concentrator.

2.2.4 RDE-4 dsRBDs NMR Sample Preparation

The pET15b plasmid containing the RDE-4 DNA sequence encoding from amino acids 28-253 was obtained from GenScript (2.2.1 (ii)). The plasmid was transformed into *E. coli* BL21 (DE3) competent expression cells by heat shocking and incubated overnight at 37°C on ampicillin LB agar plates. Cells were isotopically labeled with ^{15}N, $^{15}N^{13}C$, or $^{2}H^{15}N^{13}C$ by growing 1L cultures of cells in minimal media containing the isotope or isotopes of choice along with 1 mL of ampicillin antibiotic (100mg/mL). Note that protein expression was enhanced by adding 10% of the total media volume of the appropriate isotopically labeled Spectra 9 growth media to the prepared minimal media. Protein expression was induced with 1.0 mM IPTG at $O.D_{600}$=0.6-0.8 followed by a 4 hour growth period. The cells were sonicated using a Fisher Scientific Model 500 Sonic Dismembrator (6 cycles: 30 seconds pulse on, 30 seconds pulse off) and centrifuged at 15,000 rpm for 20 minutes. Both the supernatant and pellet were separated and retained. The pellet was resuspended in 15 mL of 6 M guanidine hydrochloride (GdnHCl) and allowed to gently shake overnight at 4°C. Following centrifugation at 15,000 rpm for 30 minutes, the supernatant was passed through a 5.00 μm filter, a 0.45 μm filter, and finally a 0.22 μm filter. The supernatant was refolded and purified using a nickel affinity column (denaturing buffer: 6 M GdnHCl, 30 mM Tris pH 8, 250 mM NaCl, 5 mM BME, 5% (v/v) Glycerol, 0.02% (w/v) NaN_3; wash buffer: 30 mM Tris pH 8, 250 mM NaCl, 5 mM BME, 5% (v/v) Glycerol, 0.02% (w/v) NaN_3; elution buffer: 400 mM imidazole,

30 mM TrisHCl pH 8, 250 mM NaCl, 5 mM BME, 5% (v/v) Glycerol, 0.02% (w/v) NaN_3). Column fractions were collected and analyzed by 15% SDS-PAGE gel. The fractions containing the purified protein of interest were concentrated down to a volume of 500 μL using a 10,000 molecular weight cut-off (MWCO) Amicon concentrator. All NMR samples were prepared by exchanging the sample buffer (using a 10,000 Amicon concentrator) with a buffer that contained 50 mM KH_2PO_4, 50 mM K_2HPO_4, 250 mM NaCl, 0.5 mM DTT, 0.02% (w/v) pH 7.0. In addition, 10% (v/v) D_2O was added to all NMR samples. The final sample volumes for all NMR samples were no less than 500μL at concentrations of approximately 1 mM.

Binding studies (1:1 complex) of the RDE-4 dual dsRBD and 83mer RNA by NMR spectroscopy were completed at concentrations of 0.15 mM in a shegimi NMR tube. Buffer used for the complex was identical as used for all NMR experiments.

2.2.5 RDE-4 dsRBDs NMR Processing and Analysis

All NMR spectra were acquired on a Bruker Avance 600 MHz and Bruker 800 MHz NMR spectrometers. All NMR spectra were processed for NMRView software [105]. Analysis of the acquired 2D and 3D NMR data were performed using CcpNMR Analysis 2.0.6 and NMRViewJ 8.1 NMR spectra software.

2.2.6 RNA Synthesis of 83mer Oligo for RDE-4 dsRBD Binding Studies

RNA for binding studies were transcribed from either a 100mer DNA strand annealed to a Top17 promoter sequence or from two 100mer DNA complimentary

strands annealed together to make a double stranded template for RNA synthesis. All RNA 40 mL large-scale reactions were ethanol precipitated and purified on a 20% large scale (19:1, acrylamide:bisacrylamide) polyacrylamide 6 M urea gel that ran at 25 W for 24 hours. The band containing the RNA was excised and electroeluted using an Elutrap at 100 V overnight. Samples were collected and concentrated using a 3,000 MWCO Amicon concentrator. The samples were exchanged 4 times with 2 mL of 1 M NaCl and then washed 4 times with 2 mL of distilled water to a final volume of ~500 μL.

2.2.7 Gel Shift Binding Studies of the Dual dsRBD Region of RDE-4 to a Long dsRNA

Samples of the complex were prepared of RDE-4^{28-253} (1.2 μM) and an 83mer RNA oligo (1.4 μM) at a molar ratio of approximately 1:1 of protein to RNA, followed by increasing protein concentrations of 10X, 100X, and 1000X. Subsequently, the reaction mixtures were run on 8% native (19:1, acrylamide:bisacrylamide) at 300 V for 35 minutes polyacrylamide gels, allowing for visualization of the protein-rna complexes.

2.2.8 ITC Study of RDE-4 Dual dsRBD:83mer RNA Complex

Isothermal titration calorimetry (ITC) was performed using a MicroCal VP-ITC instrument. All samples were in the identical buffer as used for the NMR experiments. The sample buffers were degassed using the MicroCal Thermovac.

Based on the gel shift assays, it was decided that titrating 2.5x molar ratio of protein into the RNA would be sufficient for binding to occur. Titrations were performed by injecting 2 μL of protein (0.263 mM) followed by 34 consecutive injections of 8 μL aliquots of protein (0.263 mM) into the ITC cell containing 83mer RNA (20 μM). All titrations were performed at 25°C. Enthalpy, equilibrium association constant and binding stoichiometry was determined by fitting the data to one-site and two-site binding models using Origin 5.0 (MicroCal Software Inc).

2.2.9 Crystal Screen Trials

Unlabeled protein samples of the RDE-4 dsRBDs were grown and purified as mentioned above with the exception of TYP media being used instead of minimal media containing isotopes. The final protein concentration of the RDE-4 dsRBDs for all screens was 0.5 mM (12.8 mg/mL) when in a buffer of 20 mM Tris HCl (pH 7.0), 400 mM NaCl, 0.5 mM EDTA, 0.02% (w/v) NaN_3. Crystallization screens of the RDE-4 dsRBDs protein were set up using the Qiagen NeXtal Classics Crystallization Kit via hang drop method. Optimized crystal screens were prepared under buffer conditions of 0.2 M sodium Citrate, 0.1 M Tris HCl (6.0, 7.0, 8.0, and 9.0), and PEG (20%, 25%, 35%, and 40% (v/v)). Crystallization screens of the RDE-4 dsRBDs were also prepared using the Qiagen NeXtal PEGs I Suite, Qiagen NeXtal PEGs II Suite, Qiagen NeXtal JCSG+ Suite, and Qiagen NeXtal MPDs Suite.

Samples (^{15}N labeled) of the RDE-4 dsRBDs protein and the 83mer RNA were also prepared (see sections 2.1.2, 2.1.4, 2.1.6) for crystal screening at Dr. Matthew

Wilce's laboratory at the University of Melbourne in Austalia. The samples prepared were as follows: 1) free RDE-4 dsRBDs protein (34 mg/mL), 2) 83mer RNA (34.4 mg/mL). Crystal screens of these samples were performed by Simone Beckham in Dr. Matthew Wilce's laboratory.

2.2.10 SAXS Sample Preparation

Isotopically labeled samples are not required when performing SAXS, however the samples of RDE-4 were ^{15}N labeled. The protein expression and purification of the RDE-4 was performed as described in sections 2.1.2 and 2.1.4. The 83mer RNA was prepared as described in section 2.1.6. The samples prepared for SAXS experiments were as follows: 1) free RDE-4 protein, 26 KDa (20 mg/mL), 2) 83mer RNA, 27 KDa (20 mg/mL) and 3) RDE-4 protein:83 mer RNA complex, 53 KDa (20 mg/mL). All SAXS experiments were performed and analyzed by Simone Beckham in Dr. Matthew Wilce's laboratory at the University of Melbourne in Australia. The *ab initio* derived structural envelope for RDE-4 was calculated using GASBOR [106].

2.3 RESULTS AND DISCUSSION

2.3.1 Protein Expression, Purification, and NMR Analysis

At the start of this project, a truncation of the dual dsRBD from amino acids 38-268 (RDE-4^{38-268}) was desired for cloning and expression. After many attempts, the RDE-4^{38-268} truncation was successfully cloned into pET23a plasmids, transformed to DH5αF' *E. coli* cells and expressed in BL21 (DE3) *E. coli* competent cells (2.2.1 (i)). Purification of RDE-4^{38-268} by nickel affinity column yielded a sample that was pure enough for NMR analysis by ^{15}N HSQC. However, the HSQC of the ^{15}N RDE-4^{38-268} NMR sample showed fewer than the expected number of peaks that was characteristic of an unfolded protein. It was suspected that the α-helix secondary structure preceding the RDE-4^{38-268} truncation was necessary for proper protein folding.

To investigate the necessity of particular regions of the protein for dsRNA binding and to obtain a workable protein for further studies, the viability of three new truncations were explored. RDE-4^{28-253} was a truncation of RDE-4 from amino acids 28-253. RDE-$4^{28-137+165-253}$ was a truncation of the dual dsRBD without the linker region that exists between the two domains (2.2.1 (ii)). RDE-$4^{253-385}$ was a truncation of the C-terminus region. Unlike the previous truncation, RDE-4^{38-268}, these three truncations were designed to include an additional α-helix predicted at the N-terminus of the first dsRBD to encourage proper refolding of the protein.

All three truncations were expressed in BL21 (DE3) and BL21 (pLysS) *E. coli* competent expression cells to determine which construct expressed the most amount of protein. Analysis of all protein expression on a 15% SDS-PAGE gel showed expression of RDE-$4^{28\text{-}253}$ (dual dsRBD) (Figure 2.5, 1A-1B) and RDE-$4^{253\text{-}385}$ (C-terminal, 23 KDa) (Figure 2.5, 3A-3B) in both expression cells. However, expression in BL21 (DE3) cells proved to produce the most amount of protein of both truncations compared to BL21 (pLysS) cells. As a result, BL21 (DE3) transformants were used for all subsequent culture growths. The molecular weights of RDE-$4^{28\text{-}253}$ and RDE-$4^{253\text{-}385}$ were 28 KDa and 16 KDa respectively. There was no expression of RDE-$4^{28\text{-}137+165\text{-}253}$ (dual dsRBD without the linker region) (Figure 2.5, 2A-2B) in either of the expression cells, indicating that the two dsRBDs of RDE-4 in *C. elegans* cannot be expressed without the linker. The molecular weight of RDE-$4^{28\text{-}137+165\text{-}253}$ was 23 KDa.

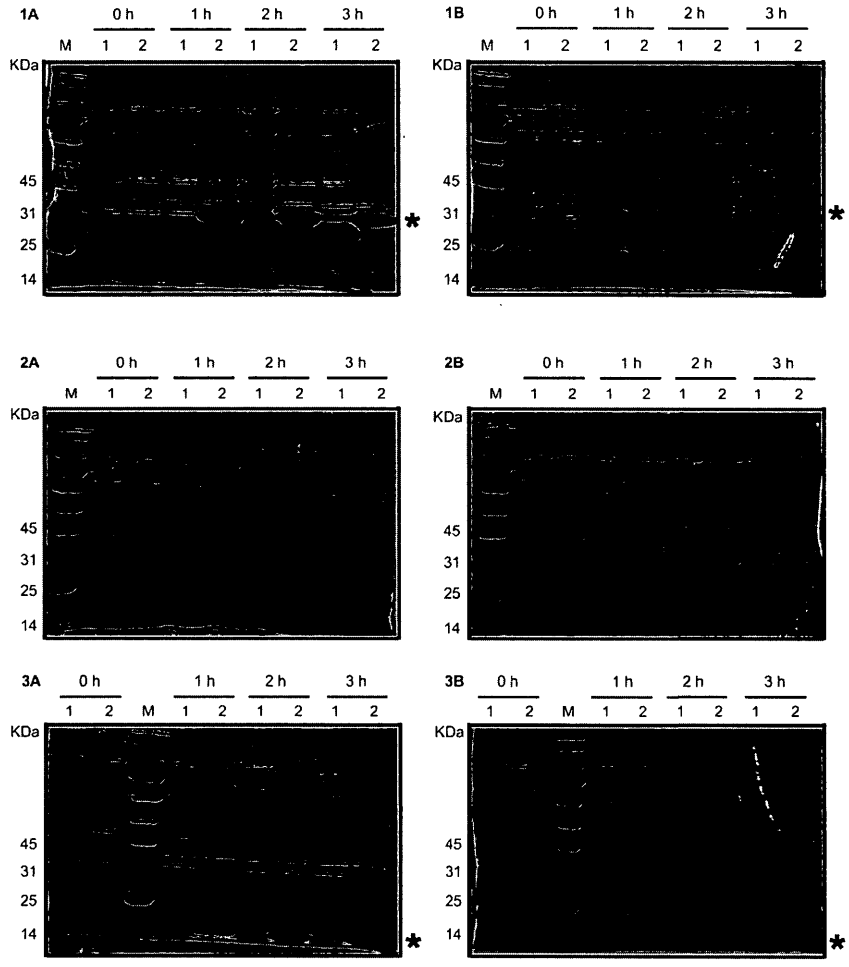


Figure 2.5: Protein expression of RDE-4 protein truncations on a 15% SDS-PAGE gel from 0-3 h. 1A is RDE-4^{28-253} in BL21 (DE3) cells and 1B is RDE-4^{28-253} in BL21 (pLysS) cells. 2A is RDE-$4^{28-137+165-253}$ in BL21 (DE3) cells and 2B is RDE-$4^{28-137+165-253}$ in BL21 (pLysS) cells. 3A is RDE-$4^{253-385}$ in BL21 (DE3) cells and 3B is RDE-$4^{253-385}$ in BL21 (pLysS) cells. Expression was seen only for RDE-4^{28-253} and RDE-$4^{253-385}$, with optimal expression occurring in BL21 (DE3) cells. "M" is the protein marker, 1 and 2 are uninduced and induced (1 mM IPTG) samples respectively, and "*" indicates expression of the protein of interest.

Truncation RDE-4$^{28-253}$ containing the dual dsRBD was selected for further studies. Note that from this point forward, the RDE-4$^{28-253}$ construct is referred interchangeably with RDE-4 or RDE-4 dsRBDs. Since protein expression was observed in unlabelled growth media, a large scale culture growth of ^{15}N minimal media was prepared in order to later observe the isotopically labeled protein using NMR spectroscopy. Analysis by a 15% SDS-PAGE gel confirmed protein expression in minimal media (Figure 2.6).

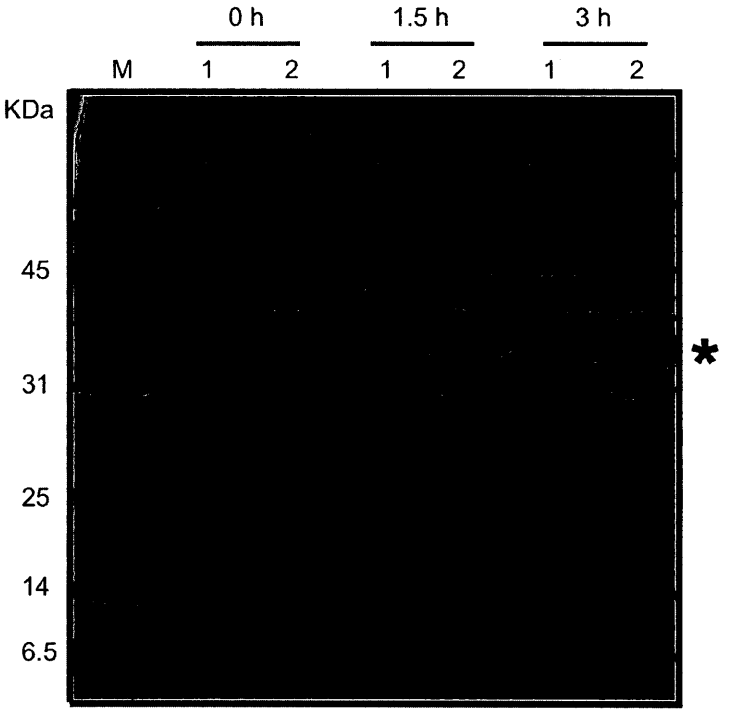


Figure 2.6: Protein expression of ^{15}N-labelled dual dsRBD of RDE-4 on a 15% SDS-PAGE gel after 0 h, 1.5 h, and 3 h. "M" is the protein marker, 1 and 2 are uninduced and induced (1 mM IPTG) samples respectively, and "*" indicates expression of the protein of interest, RDE-4.

After sonication of the cells, analysis of the lysate and pellet revealed that the protein was insoluble. In order to solubilize the protein, the pellet was resuspended with 6M guanidine hydrochloride (GdnHCl) and allowed to gently shake at 4°C overnight. GdnHCl was used because of its ability to denature proteins by disrupting non-covalent bonds, hence making the protein soluble. After centrifuging the mixture the following day, the protein was found to be in the supernatant.

A nickel column was used as a means to refold the denatured protein, purify, and then elute it. This process of on-column refolding [107] allowed the denatured protein to gradually refold over a linear GdnHCl gradient from 6M to 0M. As a result, the occurrence of the protein precipitating out of solution due to a sudden denaturant concentration change was avoided. Throughout the course of this project, the on-column refolding method had repeatedly proved to be an efficient way of refolding proteins compared to conventional dialysis methods. A nickel affinity column was used to refold RDE-4 because the protein was tagged with a hexahistidine tag, which allowed the protein to bind to the column for the refolding process, then elute during the purification step once it was refolded. This type of column has Ni^{2+} beads stationed inside it that would bind to the electron-rich tag of the protein. The protein of interest would then be eluted off the column by adding increasing amounts of imidazole, which would compete for the Ni^{2+} beads.

RDE-4 was found to elute at ~40% imidazole concentration (Figure 2.7). The resulting nickel affinity fractions were analyzed by 15% SDS-PAGE (Figure 2.8).

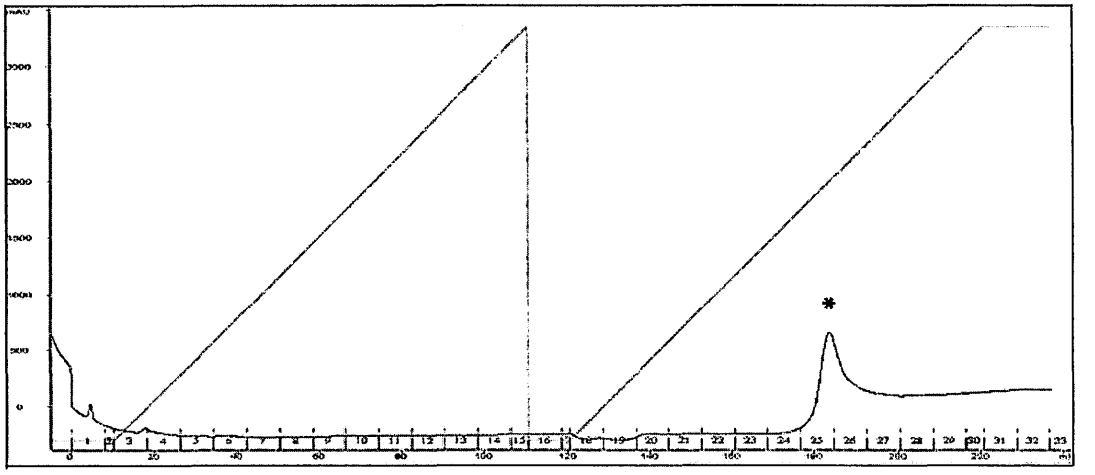


Figure 2.7: Chromatogram of ^{15}N RDE-4 dsRBDs nickel affinity column purification. The protein refolding step is shown by the first gradient in green. The elution step is shown by the second gradient in green. Absorbances (UV280nm) are in blue. Fractions are indicated in red. "*" indicates the protein of interest.

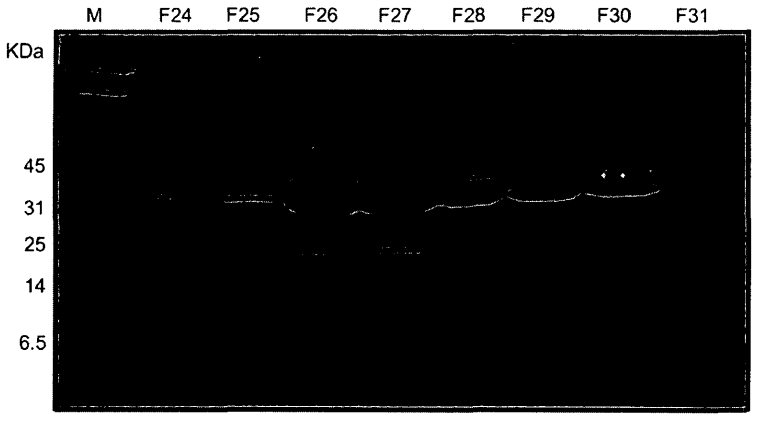


Figure 2.8: 15% SDS PAGE gel analysis of chromatogram peak fractions (F24-F31). ^{15}N Rde-4 dsRBDs were eluted in fractions F24-F30. "M" is the protein marker.

Concentration of fractions containing RDE-4 followed by buffer exchange to the NMR buffer consistently resulted in samples with a protein concentration of ~4 mM at 1 mL. Although the NMR sample was not completely pure, RDE-4 was in much higher concentration compared to all other proteins, which was sufficient for NMR studies (Figure 2.9).

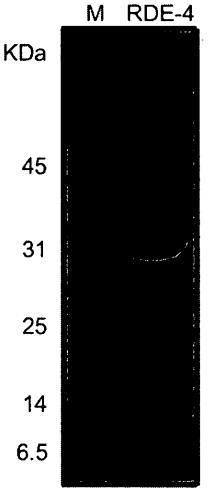


Figure 2.9: Purified ^{15}N RDE-4 dsRBDs on a 15% SDS PAGE gel.

2.3.2 NMR Analysis of RDE-4 dsRBDs

Many previous preliminary experiments conducted have determined that the 28 KDa RDE-4 dsRBDs protein expresses well and can be purified clean enough for NMR purposes with final protein sample concentrations in the range of 4 mM. However, to obtain better resolved NMR data, all NMR samples were diluted to approximately 1mM.

Excellent 2D and 3D data were acquired from the experiments performed on the Bruker 800MHz. The raw data for the ^{15}N-HSQC (Figure 2.10), HNCO, HNCA, HNCOCA, HNCACB, HCCH-TOCSY, HCCH-COSY, ^{15}N-NOESY HSQC, and ^{13}C-NOESY HSQC (aliphatic optimized and aromatic optimized) experiments were processed to obtain the corresponding spectra.

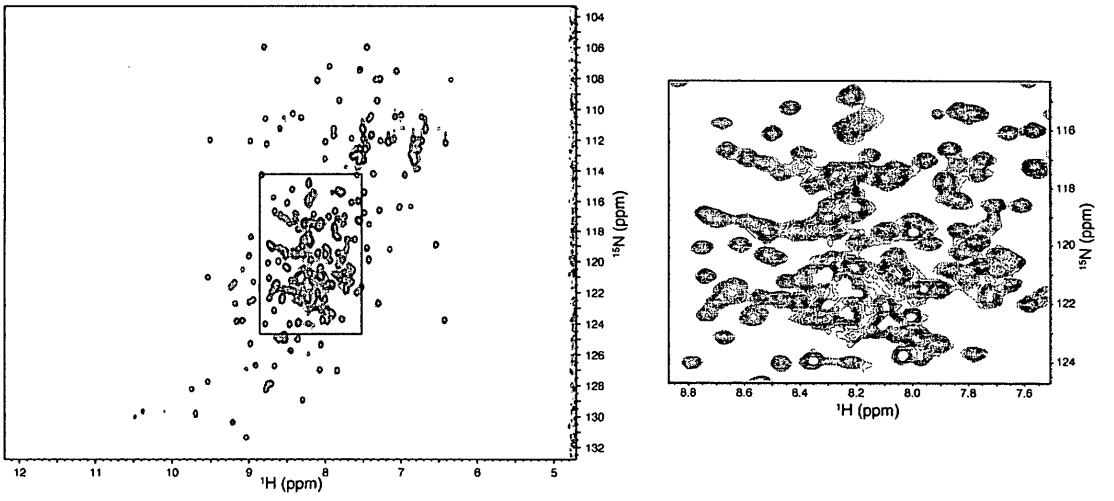


Figure 2.10: ^{15}N HSQC of a 1 mM sample of RDE-4 dsRBDs. The expansion of the boxed region shows well-dispersed peaks.

CcpNmr Analysis 2.0.6 software was used to peak pick all of the acquired 2D and 3D spectra in order to generate residue assignments. Submission of the peak lists to the PINE Server v.1.0 produced inconclusive and misleading residue

assignment output data. As a result, it was determined that in order to achieve any useful PINE Server v.1.0 assignment probabilities, additional 3D NMR experiments must be conducted to obtain a more complete set of spectral data for analysis. This was not surprising since the RDE-4 dsRBDs protein being studied was very large (226 amino acids).

2.3.3 Chemical Shift Assignments of the Dual dsRBD Region of RDE-4

Assignment of the backbone HN, N, C_α and C_β chemical shifts of the RDE-4 dsRBDs were completed by acquiring and analyzing data from CBCACONH, HNCACB, HNCA, HNCACO, and HNCO experiments. The peaks in the ^{1}H-^{15}N HSQC that were most intense with the least amount of peak overlap upon increasing the contour levels were assigned first using the HNCACB and CBCACONH as a guide (Figure 2.11).

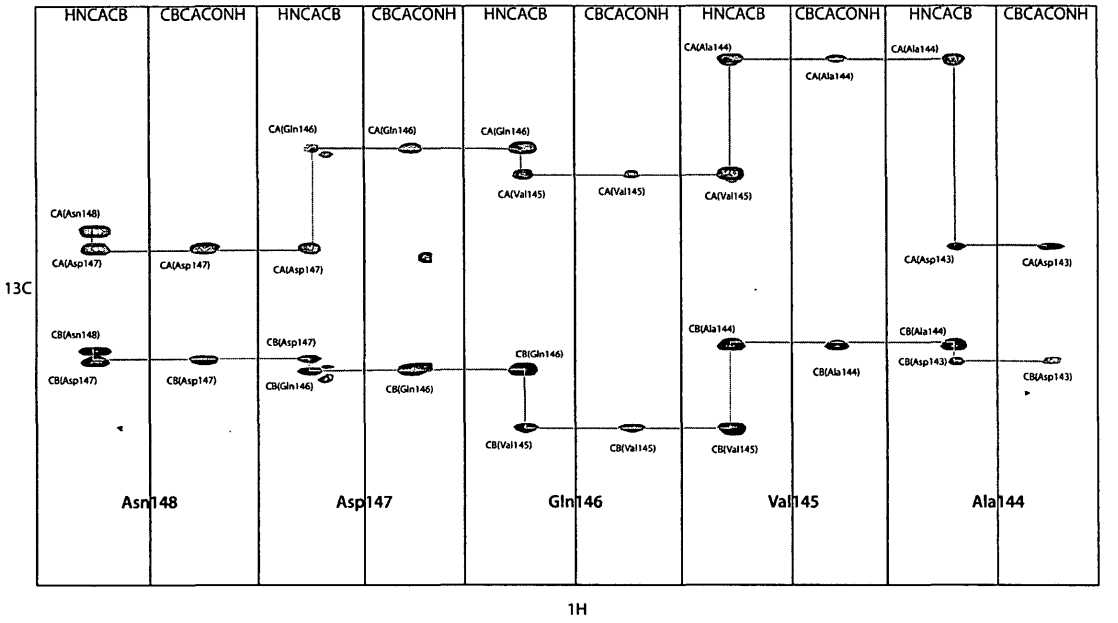
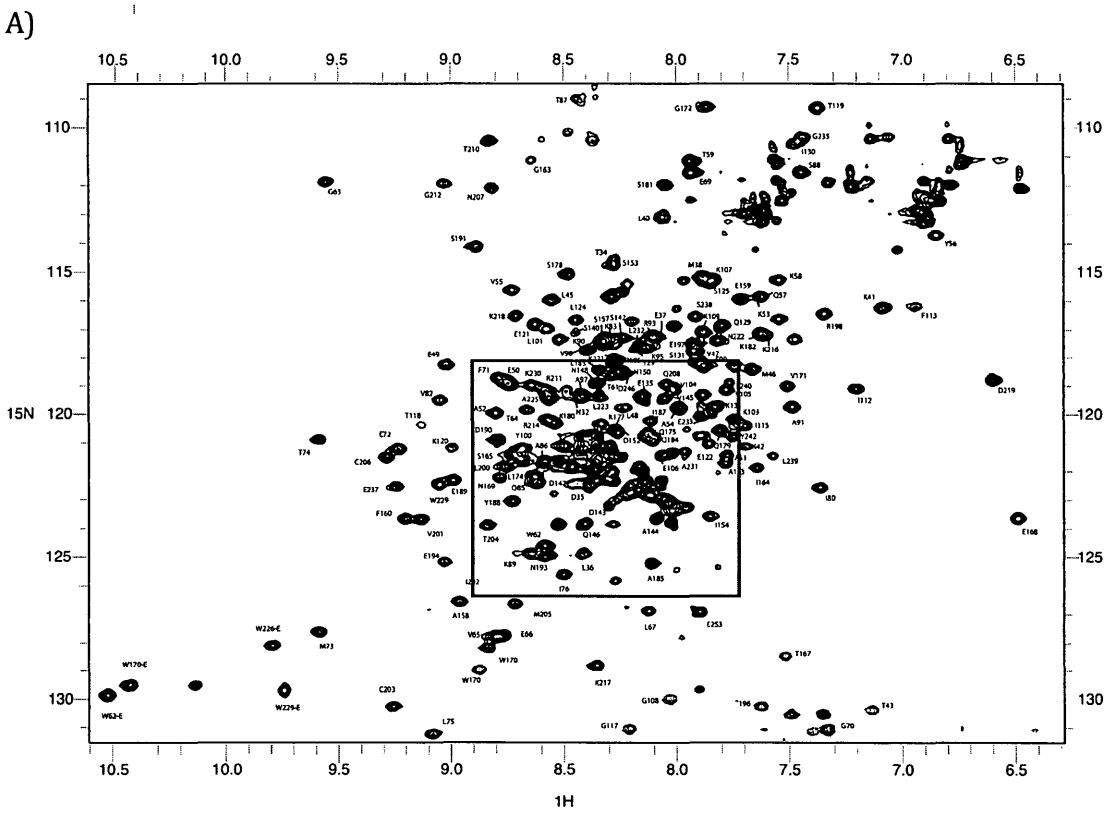


Figure 2.11: HNCACB and CBCACONH strips of RDE-$4_{28\text{-}253}$ used for assigning the backbone amides. Red peaks in the HNCACB strips are the C_α resonances and black peaks are the C_β resonances of residues i and i-1. Peaks in the CBCACONH strips are the C_α and C_β of residue i-1.

Chemical shift assignments were checked against assignments published by Chiliveri *et al.* (BMRB databank accession No. 17703) [108]. Of the 216 non-proline residues in RDE-4, backbone amide resonances were assigned for 197 resonances (Figure 2.12). The residues N78, E79, V98, E111, F114, T141, S215, L227, M228, I236, S241, M244, V245, V247, I248, E249, N250, L251, and E252 could not be assigned either because the of extensive peak overlapping in the triple resonance

experiments or because there were no correlating peaks in the triple resonance data. It is possible that the presence of μs-ms timescale dynamics resulted in peak broadening. The proline residues were not assigned since prolines lack an amide and as a result do not provide signal in the ^{1}H-^{15}N HSQC spectrum. The 10 proline residues include 30, 44, 60, 68, 116, 155, 161, 162, 166, and 186. The assigned backbone ^{1}HN and ^{15}N as well as C_{α} and C_{β} resonance assignments are listed in Table 2.1.

A)



B)

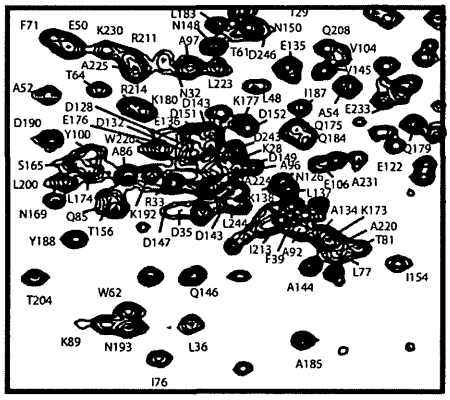


Figure 2.12: NMR peak assignments of the dsRBDs of RDE-4 shown on A) an ^{15}N-HSQC and B) expansion of the boxed central region of the spectrum.

Table 2.1: Chemical shift assignments of backbone ^{1}HN and ^{15}N as well as C_{α} and C_{β} resonances (ppm) of RDE-4 dsRBDs in 30 mM Tris HCl (pH 8), 250 mM NaCl, 5 mM BME, 5% glycerol, 10% D_2O at 25°C ('na' denotes assignments that could not be made due to an absence of triple resonance data, '?' means resonance data is inconclusive). Residues with inconclusive C_{α} and C_{β} triple resonance data were assigned based on comparisons with chemical shift assignments made by Chiliveri *et al.* [108].

Residue	^{15}N (ppm)	^{1}HN (ppm)	C_{α} (ppm)	C_{β} (ppm)
K28	121.353	8.303	56.200	32.780
T29	117.568	8.163	59.890	69.480
P30	----	----	63.360	32.390
R31	121.221	8.409	55.700	30.900
N32	119.387	8.418	53.260	38.970
R33	121.515	8.358	56.530	29.830
T34	114.604	8.273	61.940	69.880
D35	122.456	8.383	55.770	41.020
L36	124.889	8.413	?	?
E37	117.277	8.055	?	?
M38	115.169	7.884	?	?
F39	122.352	8.143	?	?
L40	113.086	8.062	56.090	?
K41	116.239	7.081	56.290	32.980
K42	121.160	7.693	?	?
T43	130.442	7.124	58.770	68.430
P44	----	----	66.370	32.400
L45	115.958	8.552	58.190	42.930

M46	118.465	7.665	59.350	34.480
V47	118.163	7.915	60.690	?
L48	119.821	8.244	58.420	42.830
E49	118.277	9.022	60.770	29.810
E50	118.812	8.759	?	?
A51	121.449	7.777	55.280	19.230
A52	120.005	8.799	56.280	18.740
K53	115.860	7.627	59.780	32.920
A54	119.742	7.990	55.210	19.560
V55	115.641	8.729	64.520	?
Y56	113.721	6.850	?	39.52?
Q57	115.851	7.617	59.730	?
K58	115.270	7.545	54.750	36.980
T59	111.180	7.941	58.810	69.480
P60	----	----	62.940	30.380
T61	118.630	8.290	61.870	70.350
W62	124.592	8.587	55.280	32.930
G63	111.903	9.559	44.760	----
T64	119.835	8.663	62.880	71.030
V65	127.821	8.804	61.950	35.420
E66	127.751	8.767	56.310	29.880
L67	126.898	8.123	52.300	41.560
P68	----	----	65.860	31.330
E69	111.586	7.936	55.730	29.890
G70	130.949	7.335	45.820	----
F71	118.737	8.787	57.450	44.110
E72	121.223	9.235	55.210	33.320

M73	127.653	9.584	52.160	32.910
T74	120.895	9.587	62.860	69.990
L75	131.240	9.081	53.220	44.160
I76	125.587	8.499	60.310	41.550
L77	123.784	8.030	55.230	38.720
N78	na	na	----	----
E79	?	?	?	?
I80	122.527	7.360	61.320	39.180
T81	123.255	7.960	60.960	70.970
V82	119.509	9.050	59.250	36.510
K83	117.402	8.323	54.750	37.03?
G84	128.889	8.853	44.620	----
Q85	122.387	8.653	54.230	31.900
A86	121.718	8.587	51.780	22.240
T87	109.014	8.443	?	?
S88	111.549	7.445	55.480	66.450
K89	124.873	8.646	60.900	31.850
K90	117.750	8.396	?	?
A91	119.742	7.487	54.330	18.850
A92	122.863	8.118	55.160	18.630
R93	117.271	8.098	59.840	29.370
Q94	116.678	7.547	58.810	27.820
K95	117.623	8.126	?	?
A96	121.619	8.336	?	?
A97	119.285	8.408	?	?
V98	na	na	----	----
E99	118.297	7.880	58.850	28.800

Y100	121.229	8.683	62.370	38.960
L101	116.994	8.576	?	?
R102	117.354	8.521	60.890	?
K103	120.191	7.741	58.230	31.320
V104	119.147	8.014	?	?
V105	119.156	7.780	66.320	?
E106	121.427	8.049	?	?
K107	115.289	7.841	56.210	32.380
G108	130.162	7.999	46.550	----
K109	117.091	7.882	55.240	33.450
H110	119.005	8.643	61.360	30.840
E111	na	na	----	----
I112	119.103	7.205	62.43?	37.49?
F113	116.199	6.941	?	?
F114	?	?	57.990?	35.420?
I115	120.451	7.696	57.810	38.990
P116	---	----	62.410	31.610
G117	131.057	8.161	45.150	----
T118	120.381	9.132	64.36?	69.00?
T119	109.324	7.371	58.750	72.810
K120	121.179	8.997	60.230	32.380
E121	116.838	8.628	60.830	28.780
E122	121.076	7.856	59.080	30.900
A123	121.729	7.782	55.220	17.680
L124	116.681	8.445	58.240	40.940
S125	115.352	7.842	?	?
N126	121.952	8.190	?	?

I127	121.520	8.251	?	?
D128	120.763	8.404	?	?
Q129	116.865	7.859	59.290	29.770
I130	110.557	7.478	63.300	38.910
S131	117.750	7.925	60.520	63.520
D132	120.901	8.381	55.24?	40.95?
K133	119.842	7.839	56.800	32.710
A134	122.967	8.060	52.830	18.930
E135	119.361	8.151	56.960	30.040
E136	121.007	8.396	?	?
L137	122.001	8.151	55.24?	42.11?
K138	121.895	8.165	56.650	32.890
R139	122.058	8.305	56.270	32.860
S140	117.115	8.448	58.320	63.910
T141	na	na	----	----
S142	117.305	8.248	?	?
D143	121.944	8.359	54.120	41.100
A144	123.663	8.089	52.680	19.300
V145	119.445	8.053	62.570	32.770
Q146	123.837	8.409	55.780	29.840
D147	121.942	8.362	54.340	41.250
N148	118.911	8.354	53.350	39.250
D149	121.515	8.291	54.230	41.100
N150	118.500	8.258	53.090	39.250
D151	121.070	8.362	54.790	41.130
D152	120.638	8.262	56.692	32.224
S153	114.751	8.276	58.210	63.680

I154	123.589	7.851	59.470	?
P155	----	----	?	?
T156	122.202	8.628	?	?
S157	117.616	8.292	58.590	63.920
A158	125.839	8.274	19.270	52.720
E159	115.949	7.714	?	?
F160	123.691	9.203	57.150	38.540
P161	----	----	?	?
P162	----	----	63.830	31.510
G163	111.144	8.643	45.080	----
I164	121.888	7.642	57.190	34.920
S165	121.440	8.721	55.290	63.950
P166	----	----	62.350	32.930
T167	128.900	7.505	60.950	68.420
E168	123.645	6.490	55.290	29.840
N169	122.246	8.783	50.630	35.950
W170	128.182	8.834	60.340	28.880
V171	119.004	7.508	65.910	32.480
G172	109.292	7.866	47.610	----
K173	123.071	8.044	59.400	31.420
L174	121.535	8.688	57.770	42.040
Q175	120.698	8.127	59.260	27.840
E176	120.926	8.417	59.860	29.420
K177	120.565	8.285	59.340	32.310
S178	115.058	8.480	62.380	?
Q179	120.600	7.808	58.250	28.430
K180	120.277	8.544	59.270	32.890

S181	111.992	8.054	58.790	63.870
K182	117.127	7.626	57.280	28.810
L183	118.452	8.345	?	?
Q184	120.946	8.106	?	?
A185	125.226	8.113	50.710	17.650
P186	----	----	?	?
I187	120.226	8.116	?	?
Y188	123.058	8.727	56.730	41.220
E189	122.341	8.991	55.200	33.930
D190	120.912	8.792	52.710	45.610
S191	114.137	8.886	57.860	65.600
K192	121.713	8.521	54.760	35.350
N193	124.954	8.582	52.670	39.550
E194	125.205	9.028	58.310	29.890
R195	118.180	8.284	58.280	30.280
T196	130.356	7.604	61.320	69.870
E197	117.488	7.929	57.270	27.280
R198	116.438	7.345	54.730	33.360
F199	118.923	8.745	58.250	41.950
L200	121.824	8.755	54.720	41.570
V201	123.691	9.135	60.830	33.950
I202	126.570	8.962	59.300	40.460
C203	130.269	9.256	57.160	28.820
T204	123.892	8.838	61.350	69.940
M205	126.651	8.715	56.260	34.860
C206	121.529	9.288	60.390	25.790
N207	112.088	8.819	54.340	37.970

Q208	118.941	8.047	54.250	31.510
K209	119.359	7.883	55.170	?
T210	110.457	8.833	60.260	?
R211	119.272	8.575	57.570	?
G212	111.951	9.030	?	----
I213	122.648	8.171	?	?
R214	120.189	8.575	52.220	35.950
S215	na	na	----	----
K216	117.197	7.613	54.390	34.320
K217	128.829	8.353	?	?
K218	116.526	8.711	59.320	32.260
D219	118.814	6.600	56.300	40.900
A220	123.159	8.020	55.780	18.920
K221	118.050	8.268	59.330	33.040
N222	117.392	7.817	56.150	39.460
L223	119.387	8.344	57.210	?
A224	121.773	8.325	56.040	18.060
A225	119.435	8.566	54.710	18.750
W226	128.109	9.793	?	?
L227	na	na	----	----
M228	na	na	----	----
W229	122.508	9.052	61.450	28.790
K230	119.168	8.584	57.380	30.850
A231	121.322	7.960	54.240	17.690
L232	117.549	8.163	?	?
E233	120.092	7.894	58.670	30.190
D234	116.718	8.196	?	?

G235	110.369	7.438	45.090	----
I236	na	na	----	----
E237	122.522	9.244	?	?
S238	116.555	7.916	60.380	62.920
L239	121.439	7.574	55.840	41.570
E240	118.897	7.765	58.870	34.36?
S241	?	?	?	?
Y242	120.800	7.749	56.35?	35.27?
D243	121.169	8.297	54.650	41.080
M244	?	?	?	?
V245	?	?	?	?
D246	118.565	8.231	?	?
V247	na	na	----	----
I248	na	na	----	----
E249	na	na	----	----
N250	na	na	----	----
L251	na	na	----	----
E252	na	na	----	----
E253	126.920	7.897	?	?

Based on the CO, C_α, and C_β assigned resonances, the secondary structural elements of RDE-4 were analyzed to provide a consensus chemical shift index (CSI) of the protein (Figure 2.13). The resulting CSI shows a complete αββββα motif for dsRBD1 (residues 33-104) and partial αββββα motif for dsRBD2 (residues 170-232) that is lacking the last helix of the motif. A partial motif is seen for the dsRBD2

because the majority of residues that could not be assigned belonged to the amino acids located in that expected helix region of dsRBD2. Nonetheless, the CSI data for the assigned resonances of RDE-4 is consistent with the consensus that is characteristic for protein domains that bind to dsRNA and reinforces that RDE-4 assignments made are reasonable. Furthermore, the CSI data shows a structured linker region between the dsRBD domains that contains α-helices. The indication of this structured region reinforces that the linker sequence is necessary for RDE-4 expression and possibly for the binding function of the protein.

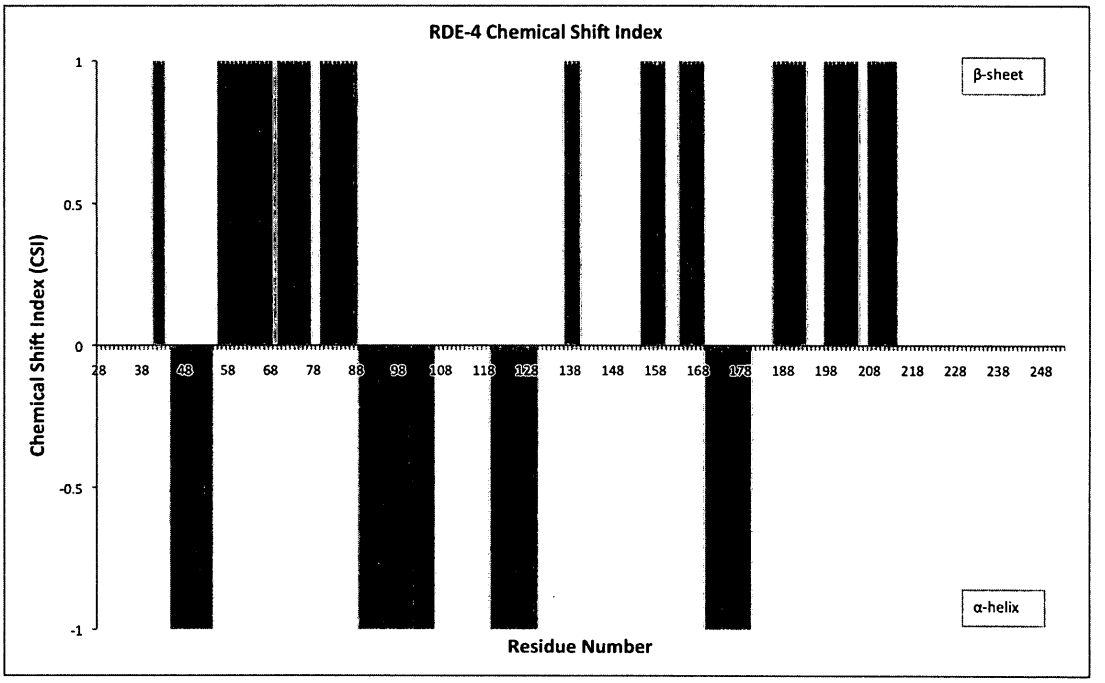


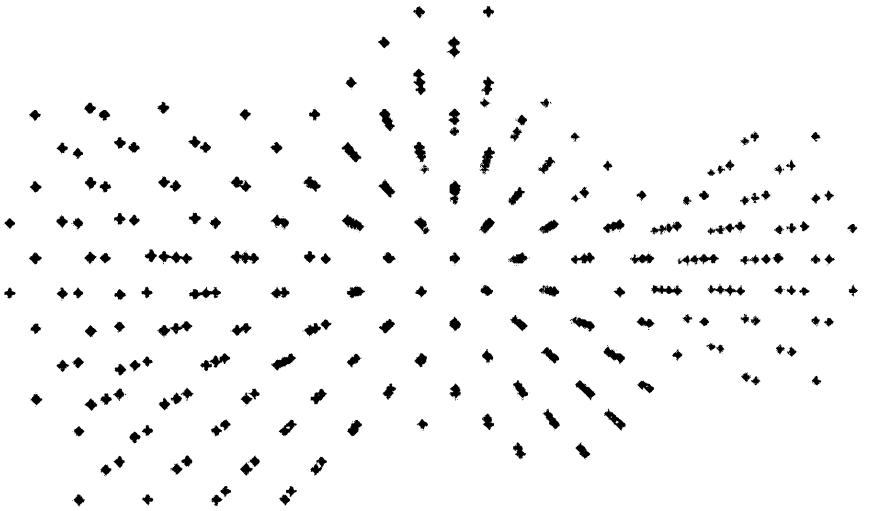
Figure 2.13: The consensus chemical shift index of the dual dsRBD region of RDE-4 derived from the assigned CO, C_α, and C_β resonances with respect to residue number. Helices = -1, loops = 0 and sheets = 1.

2.3.4 SAXS Analysis of RDE-4 dsRBDs

Small angle X-ray scattering (SAXS) was performed on RDE-4 dsRBDs to obtain structural information about the shape and size of this protein. Unlike crystallography, a crystalline sample is not needed for SAXS experiments, which is a desirable feature since the RDE-4 protein proved difficult to crystallize. The SAXS data obtained of the RDE-4 protein effectively complements the NMR-based structural studies presented.

SAXS data on RDE-4 dsRBDs reveals that the protein, containing two domains separated by a linker region, has a football-like shape opposed to being dumbbell-like (Figure 2.14). This implies that a flexible linker does not separate the two domains and the two domains do not interact with one another, thus forming a continuous interface. This single structured unit would enable RDE-4 to bind long RNA molecules that fit the entire binding site, but not RNA molecules that are shorter than the length of the binding surface. This also implies that important RNA-interacting residues exist at both ends of the binding site.

a)



b)

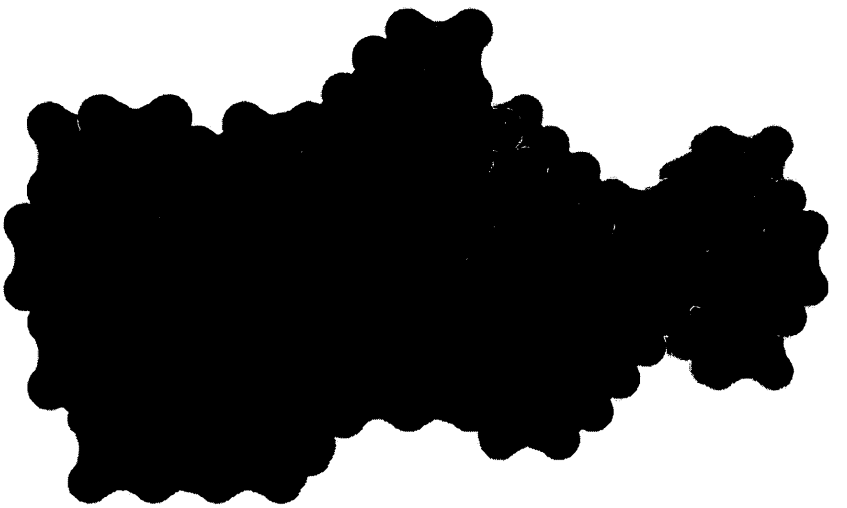


Figure 2.14: SAXS 3D structural map of RDE-4 dsRBDs protein represented as a) points in space and b) a continuous surface. Both maps suggest that the dual dsRBD of RDE-4 has a football-like structure. The data implies that a continuous surface exists where the two domains do not interact with one another.

An *ab initio* calculation of the structural envelope of the RDE-4 protein overlays very well with a derived model of the RDE-4 protein that is based on the structure of the DGCR8 protein (Figure 2.15) [102]. The largest deviation between the model and SAXS-derived shape occurs at residues 81-104. Here, RDE-4 is expected to form a long loop ending with a short four-residue helix. This difference is insignificant and occurs because there are no restraints on this unstructured loop. Future models of RDE-4 would be refined against the SAXS-derived molecular shape and should accommodate this loop within the structural envelope.

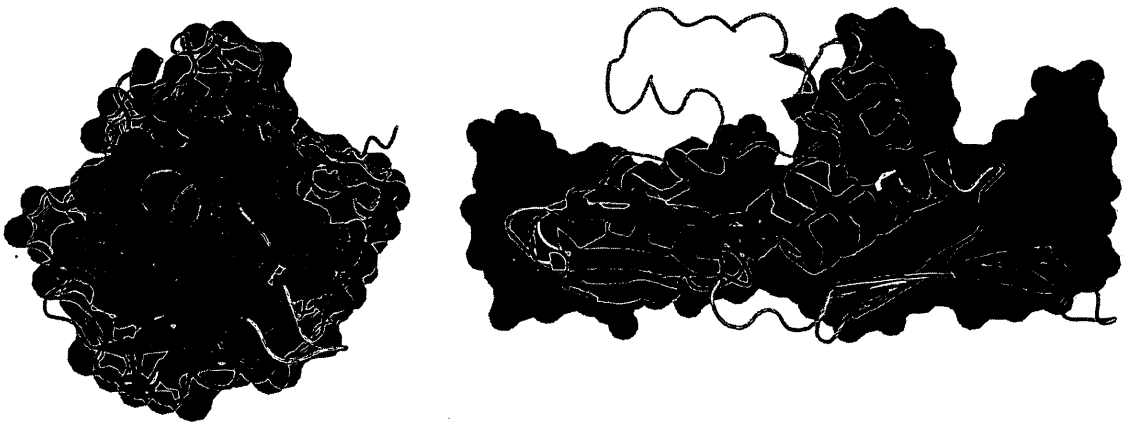


Figure 2.15: Overlay of a structural model and the SAXS-derived structural envelope of the dual dsRBD-containing RDE-4 protein. Shown in the green cartoon is the structure of RDE-4 modelled on the DGCR8 structure [PDB ID: 2YT4] [102]. In red is the *ab initio* determined structural envelope of RDE-4. Overall, there is a fair agreement between the model and the experimentally derived shape. This structural overlay demonstrates that RDE-4 likely forms a single structural entity as opposed to being two separate nodes connected by a linker region.

2.3.5 Crystallization of RDE-4 dsRBDs

Crystals of the RDE-4 dsRBDs proved to yield promising results under optimized crystal screen conditions of 0.2 M sodium citrate, 0.1 M Tris HCl (pH 6.2), 20% PEG 400 and 0.2 M sodium citrate, 0.1 M Tris HCl (pH 6.8), 25% PEG 400. The crystal grown at pH 6.2 with 20% PEG 400 had measured dimensions of 2.2 mm (length) x 2.2 mm (width) (Figure 2.16 A). The crystal grown at pH 6.8 with 25% PEG 400 obtained had measured dimensions of 1.2 mm (length) x 1.4 mm (width) (Figure 2.16 B). Unfortunately, diffraction of these crystals showed no protein diffraction patterns, but instead evidence of salt peaks.

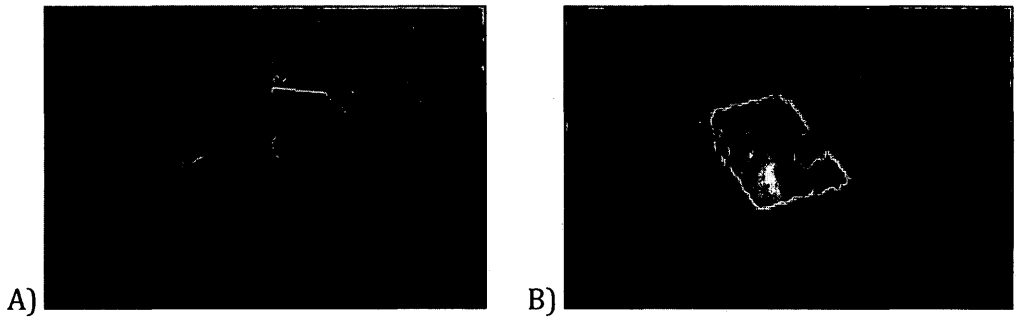


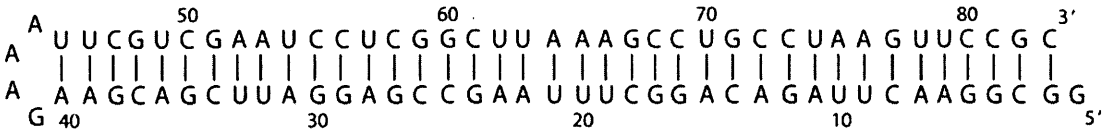
Figure 2.16: Crystallization of the RDE-4 dsRBDs protein grown by hanging drop method. The protein is 0.5 mM in a buffer of 20mM Tris HCl (pH 7.0), 400mM NaCl, 0.5mM EDTA. Screen conditions of A) 0.2M sodium citrate, 0.1M Tris HCl (pH 6.2), 20% PEG400 and B) 0.2M sodium citrate, 0.1M Tris HCl (pH 6.8), 25% PEG 400.

It must be noted that it was unknown when both crystals were formed because the crystal screen from which both crystals originated were not checked previously for more than a year. Since crystal formation of the RDE-4 dsRBDs occurred in the presence of polyethylene glycol (PEG), new screens using fresh protein (11mg/mL) were set up using Qiagen NeXtal PEGs I Suite and Qiagen NeXtal PEGs II Suite that contained several molecular weights of PEG in various combinations. In addition, screens were assembled using the Qiagen NeXtal JCSG+ Suite to check the crystallization possibilities with other commonly tested precipitants. And finally, crystal screens were set up using the Qiagen NeXtal MPDs Suite to determine if the organic precipitant 2-methyl-2,4-pentandiol (MPD) would encourage protein crystallization. After almost a year, no crystals were evident in any of these screens. On average, an optimized crystal diffracts at 5.0 Å. For publications, diffraction at 4.0 Å is most often required, but at this range only information about protein folds with resolution of only some side-chains is possible. At 3.5 Å side-chain positions can be resolved with confidence. For the purposes of the RDE-4 project, resolution of at least 3.0 Å would be necessary to make conclusions about specific interactions and biological function.

2.3.6 Binding Studies of the Dual dsRBD Region of RDE-4 to Long dsRNA by Gel Shift Assay

To investigate the nucleic acid binding properties of the dual dsRBD region of RDE-4, gel mobility shift assays were performed with purified RDE-4 and an in-house synthesized 83 nucleotide RNA substrate (83mer) composed of 39 Watson-Crick base pairs and a GAAA tetraloop (Figure 2.17). Tetraloops assist in initiating the RNA folding process as well as by providing sites for RNA tertiary contacts and for protein folding. The addition of a GAAA tetraloop has been found to significantly increase RNA folding stability, therefore an RNA hairpin without a GAAA tetraloop may affect protein binding due to the possibility of decreased stability of the RNA structure. For these reasons, binding studies were performed using a 83mer RNA containing a GAAA tetraloop that was synthesized from a 100mer DNA template (Figure 2.18).

(a)



(b)

Figure 2.17: (a) Sequence and putative secondary structure of the 83 nucleotide RNA molecule (83mer) used in these studies. The molecule is composed of 39 Watson-Crick base pairs and a GAAA tetraloop. (b) Model of the structure of the 83mer generated by MC-Sym. This structure was generated with the assumption that the molecule would form a standard A-form helix and have the canonical tetraloop structure.

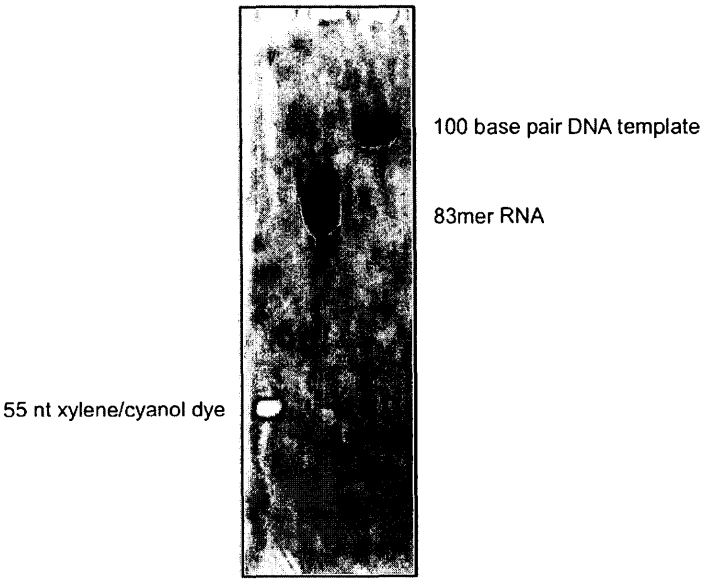


Figure 2.18: 20% denaturing polyacrylamide gel of 83mer RNA (lane 2) product transcribed from a 100 base pair DNA template (lane 3).

Shifts were observed on an 8% native polyacrylamide gel (19:1, acrylamide:bisacrylamide) for complexes with 0.75X, 1X, 2X and 3X protein excess, indicating that the dual dsRBD region of RDE-4 has an affinity for this long dsRNA and that RNA binding is significant with higher amounts of protein (Figure 2.19). Conversely, increasing the concentration of RNA rather than protein does not show binding. These data suggests that multiple RDE-4 molecules bind to an RNA substrate with positive cooperativity, thereby increasing the binding density and the

overall size of the complex, resulting in decreased mobility. The indication of binding events occurring between the RDE-4 dsRBDs and the RNA oligo makes it possible to identify the length and composition of RNA necessary for conducting structural studies of the RDE-4:RNA interaction by NMR methods.

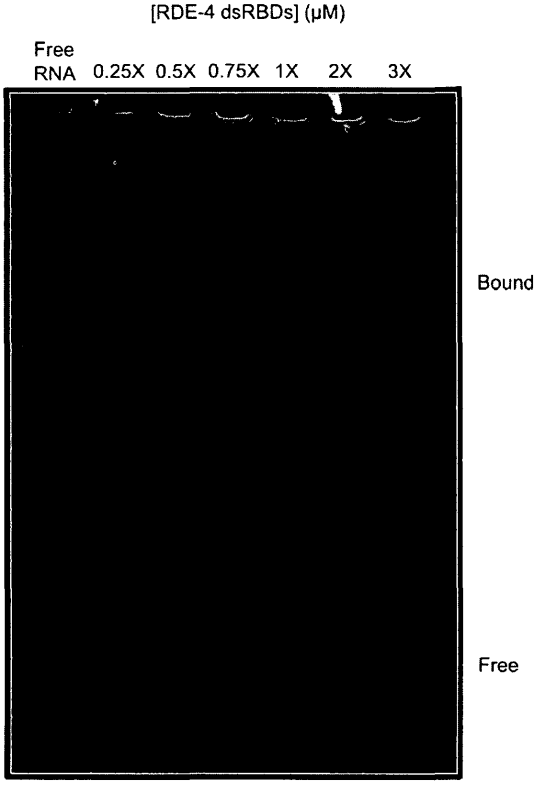


Figure 2.19: Gel mobility shift analyses of RDE-4 and 83mer RNA on an 8% native polyacrylamide gel. Bands corresponding to bound and free RNA are labelled. Increasing concentrations of RDE-4 added to 83mer RNA shows significant band shifting, indicating that protein-RNA complex formation occurs at high protein concentrations.

2.3.7 Binding Studies of the Dual dsRBD Region of RDE-4 to Long dsRNA by NMR and ITC Techniques

To identify residues from the dual dsRBD of RDE-4 that are important for RNA binding, chemical shift changes between the RDE-4 protein and RDE-4 protein:83mer RNA complex was completed using an ^{15}N-HSQC NMR experiment. This type of 2D experiment is a standard NMR experiment that shows all H-N correlations. These correlations are mainly backbone amide groups, however asparagine [109] and glutamine (Gln) side-chain Nδ-Hδ2/Nε-Hε2 groups, as well as tryptophan (Trp) residue side-chain Nε-Hε groups are also seen. The Nε-Hε peaks from arginine (Arg) residues are also visible, but are most often seen as negative peaks that are aliased/folded because they are at a chemical shift outside the recorded regions.

^{15}N-HSQC spectra of the RDE-4 protein and RDE-4 protein:83mer complex showed well dispersed peaks indicative of a well-folded protein and protein complex. The bound spectrum is clearly different, indicating that binding is occurring. From a superposition of the two spectra, a large number of peaks disappear, some peaks shift in chemical shift, and some remain unchanged (Figure 2.20).

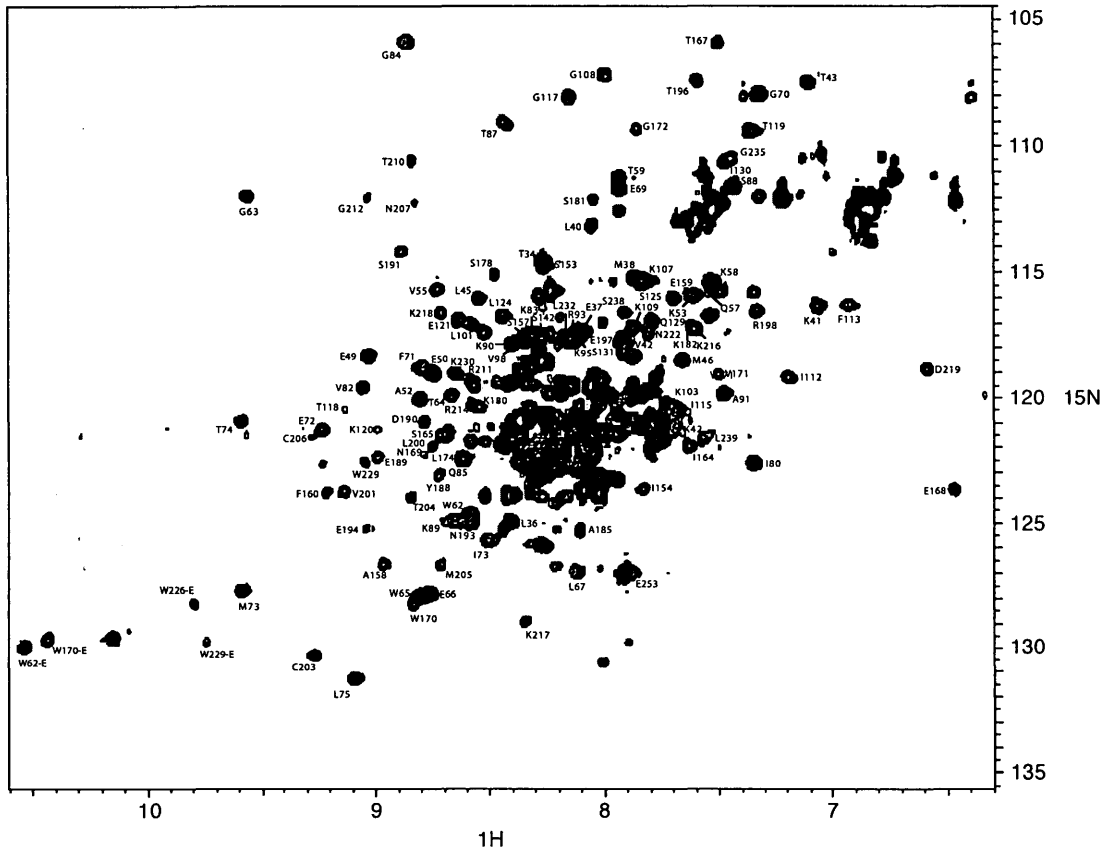


Figure 2.20: Overlay of RDE-4:83mer RNA (red) and RDE-4 protein (black) spectra.

Most likely, the residues that disappear lie directly at the binding site, those that shift are close to the binding site and those that remain unchanged are far removed from the binding site. Dynamics at the binding site may result in the observed signal loss because of movement of the RNA in the protein-binding site, or due to movement of the protein along the RNA helix. Furthermore, intermediate exchange as a result of RDE-4 binding to long RNA with an intermediate affinity may

cause peaks to disappear. Since the peaks disappear in the dual dsRBD, which is essentially the entire region of the protein studied, these results suggest that there is a significant global change between the free and bound structures and not just the residues on the binding face are affected. Furthermore, the binding does not stabilize the floppy terminus ends of the protein.

An ITC study of RDE-4 titrations into the 83mer RNA show binding is determined to occur at micromolar affinity of K_{d1} = 2.6 μM. This result is consistent with binding being in the intermediate exchange rate on the NMR timescale. Intermediate exchange leads to the line broadening that is seen in the NMR experiments. Furthermore, this result is in agreement with the micromolar binding affinity reported by Bass *et al.* [110]. A two-site binding model is used to fit the data to show that there is a second binding event occurring upon complex formation where a second protein molecule binds the RNA, however this binding is much weaker at approximately K_{d2} = 0.3 mM (Figure 2.21). This research presents the first report of an affinity value for the second binding site known to date.

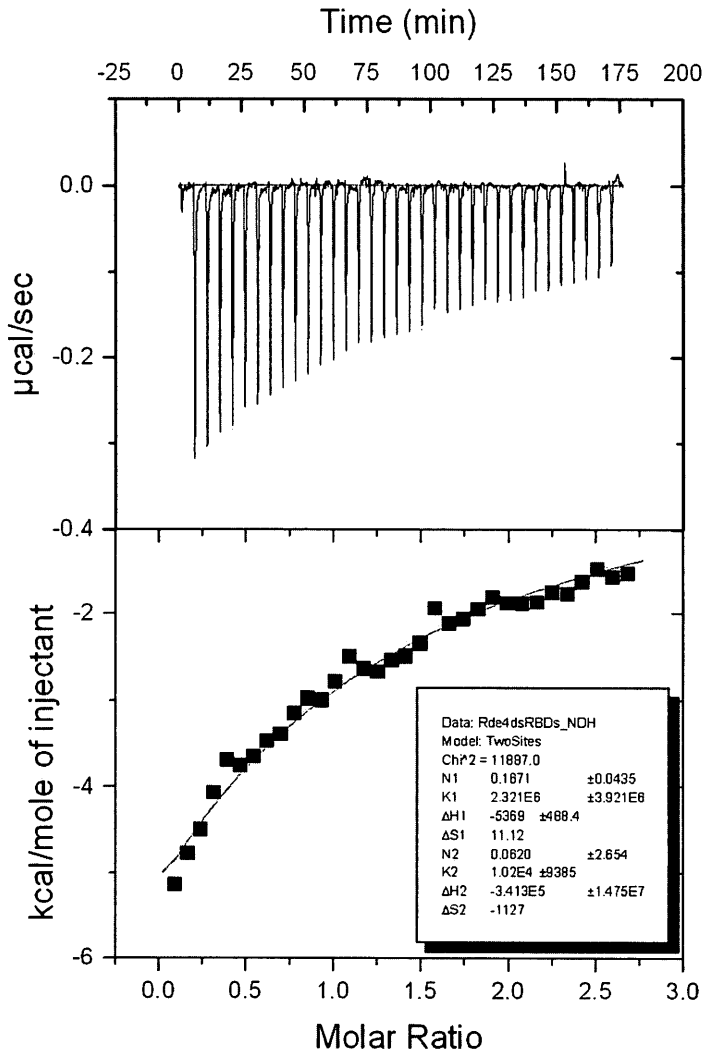


Figure 2.21: ITC profile of RDE-4 binding to the 83mer RNA. A two-site binding model shows that binding of RDE-4 dsRBDs to long dsRNA occurs via more than one protein molecule and that binding affinity is in the micromolar range.

2.4 CONCLUSIONS

For years scientists have studied proteins, such as transcription factors, that regulate gene activity. Now, the use of RNA through the RNAi process is recognized as a key approach for gene regulation. RNAi has expanded dramatically since its discovery, however, the intricacies of this natural process still remains largely unknown. By understanding the mechanistic details and the biomolecules involved in this process, the ability of RNAi to knock down genes that are responsible for the activity of diseases can essentially be accomplished easily and quickly. As a result, continued research into RNAi will allow this process to become a powerful tool for medical therapies.

RDE-4 in *C. elegans* served as a useful model for studying how double stranded RNA binding proteins that bind to dsRNA molecules are critical for RNAi cellular processes. The near complete backbone chemical shift assignments of the RDE-4 construct confirmed characteristic protein domains that bind to dsRNA. RDE-4 preferentially binds to long dsRNA rather than short dsRNA molecules [96]. For this binding event to occur, the two binding domains in the RDE-4 protein do not interact with one another, but instead forms a continuous interface onto which the long dsRNA can interact. The results presented here suggest that important residues exist at both ends of the binding site. The location of these important residues facilitate binding to long RNA molecules and not to short RNA molecules (Figure 2.22), which affords the RDE-4 protein the ability to bind to the long target

dsRNA and not to the short siRNA product. During binding, multiple RDE-4 molecules cooperatively bind to a non-sequence specific RNA substrate with micromolar affinity. This protein-RNA interaction causes a significant global change of the protein structure and not just the residues on the binding face are affected.

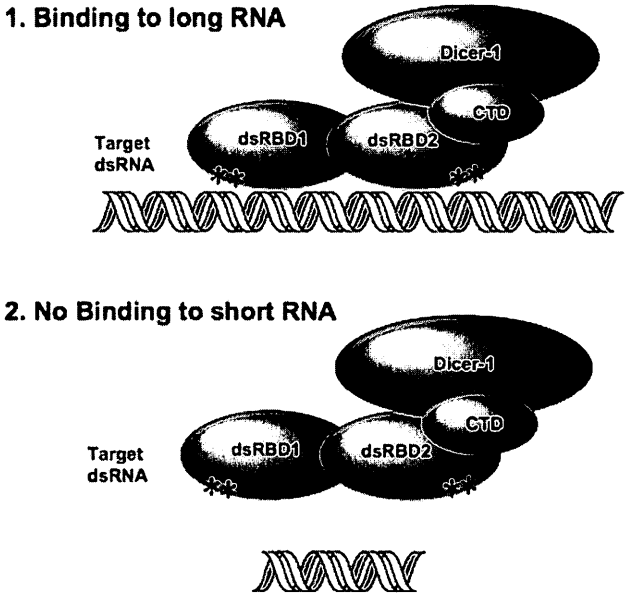


Figure 2.22: Proposed mechanism for the selective binding of RDE-4 to long, but not short RNA. This hypothesis proposes that the two dsRBDs in RDE-4 (brown) form one long binding surface with critical residues located at either end (shown by asterisks; *). In this model, long dsRNA can interact with the necessary regions of both dsRBDs (1). In contrast, short dsRNA molecules cannot interact with all the important residues on the binding surface and therefore does not bind (2).

2.5 FUTURE WORK

The future work of this research project aims to expand upon these studies by continuing to investigate the critical residues at the end of the binding site that are needed for a long RNA molecule to interact and for productive binding to occur. This information will allow for a plausible binding mechanism to be established that will explain why RDE-4 binds to long RNA rather than short RNA molecules. To accomplish this, mutagenesis experiments of the terminal residues of the dual dsRBD region of RDE-4 will allow for a better understanding of the structural basis for this binding specificity. To investigate this binding specificity further, additional NMR studies and binding studies of RDE-4 with a range of RNA molecules with varying lengths and GC content will be useful for studying binding affinity. Furthermore, it may be possible that sequence plays a role in RNA recognition by RDE-4, potentially through RNA duplex stability. The stability of these RNA molecules is expected to be different and can be measured by UV-monitored thermal melting methods. Additional aspects to be addressed include explaining how RDE-4 interacts with DCR-1 and determining the structure of the RDE-4 C-terminal domain in order to understand its function.

CHAPTER 3

AN INVESTIGATION OF THE DYNAMICS OF DIHYDROFOLATE REDUCTASE UPON INHIBITOR BINDING

3.0 DIHYDROFOLATE REDUCTASE

The enzyme DHFR is an important cancer and rheumatoid arthritis therapeutic target. DHFR regulates 5,6,7,8-tetrahydrofolate (THF) production in cells by using the cofactor nicotinamide adenosine dinucleotide phosphate (NADPH) to catalyze the reduction of 7,8-dihydrofolate (DHF) to (THF) via a cycle of conformational changes [111]. The THF product is essential for the synthesis of nucleic acid precursors and amino acids. Therefore, the inhibition of DHFR can limit the growth of rapidly dividing cells in cancers as well as limit this enzyme from preventing amino acid metabolism that leads to rheumatoid arthritis. Research of potential inhibitors of DHFR is of particular interest for the development of anti-cancer and anti-arthritis drugs such as methotrexate [112].

The three-dimensional structure of DHFR (159 amino acid) consists of eight-strand beta sheets that are connected by four alpha helices along with several flexible loops [113]. The enzyme can be divided into two subdomains: (1) the cofactor-binding domain (residues 38-88) that binds NADPH through the adenosine-binding loop (residues 67 to 69) and, (2) the major catalytic subdomain. The major catalytic subdomain is located in the active site and binds the substrate via an active-site loop called the Met20 loop (residues 9 to 24), FG loop (residues 116 to 132), and GH loop (residues 142 to 150), all of which participate in the catalysis to some extent (Figure 3.1) [114].

Figure 3.1: Three-dimensional structure of DHFR (PDB ID: 2ANQ) containing 8 strand beta sheets and 4 alpha helices connected by flexible loops including the Met20 loop (red), FG loop (blue), GH loop (purple), and adenosine-binding loop (orange). All other resides are colored green.

During the catalytic cycle, the enzyme (E) adopts either a "closed" conformation or an "occluded" conformation, depending on the intermediate formed (Figure 3.2) [114].

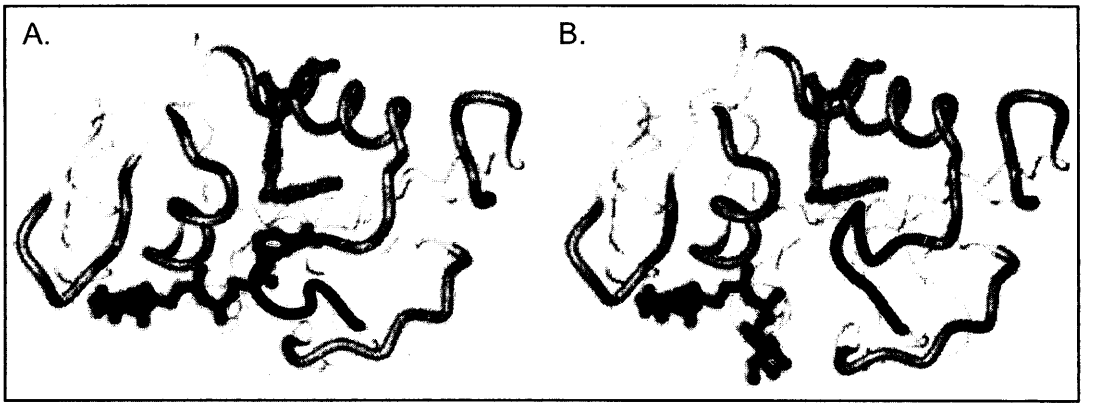


Figure 3.2: Conformations of DHFR during the catalytic cycle: A. closed, and B. occluded. The polypeptide backbone that surrounds the active site and adenosine-binding loop (blue and green), substrate (yellow) and cofactor (pink) are indicated [115].

In the binary haloenzyme, E:NADPH, the Met20 loop assumes a closed conformation that allows for close packing of the nicotinamide ring of NADPH (Figure 3.3) [111]. Water molecules near the active site donate a proton to N5 of DHF (Figure 3.3).

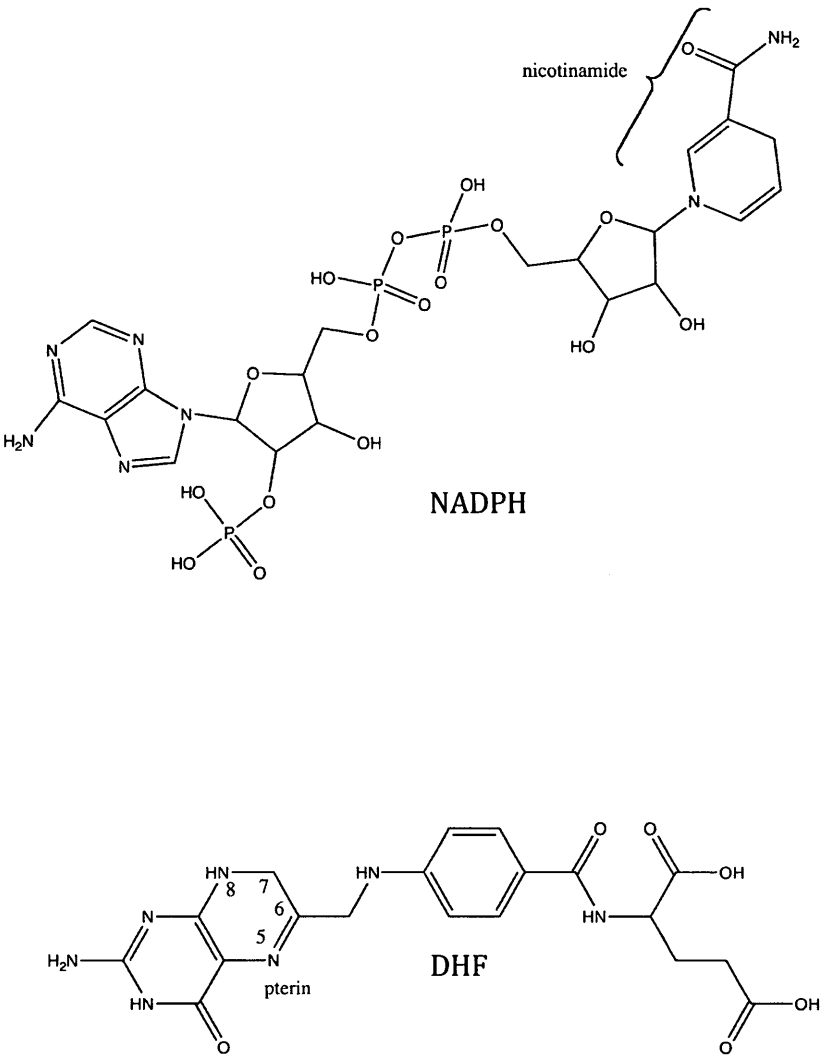


Figure 3.3: Chemical structures of NADPH and DHF.

Subsequent binding of the DHF substrate to the active site permits a hydride transfer from NADPH cofactor to the C6 atom of the pterin ring in the DHF molecule to yield the THF product and oxidation of NADPH to $NADP^+$ (Figure 3.4) [116]. As a result of this hydride transfer, the Met20 loop of the ternary complex, E:$NADP^+$:THF, adopts an occluded conformation that sterically prevents the nicotinamide ring from binding to the active site, causing the nicotinamide ring to be positioned outside of the substrate binding pocket [114]. The occluded conformation is made possible by the protruding sidechain of residue M16 of the Met20 loop into the active, as well as the disruption of the hydrogen-bonding network between the Met20 loop and the FG loop, and the formation of new hydrogen-bonding between the Met20 loop and the GH loop [114]. Replenishment of NADPH to the active site releases the THF product (rate-limiting step) and allows for the catalytic cycle to repeat. Studies of the kinetic characterization of intermediates show that DHFR proceeds through a conformational catalytic cycle involving five intermediates: E:NADPH, E:NADPH:DHF (Michaelis complex), E:$NADP^+$:THF, E:THF, E:NADPH:THF [111].

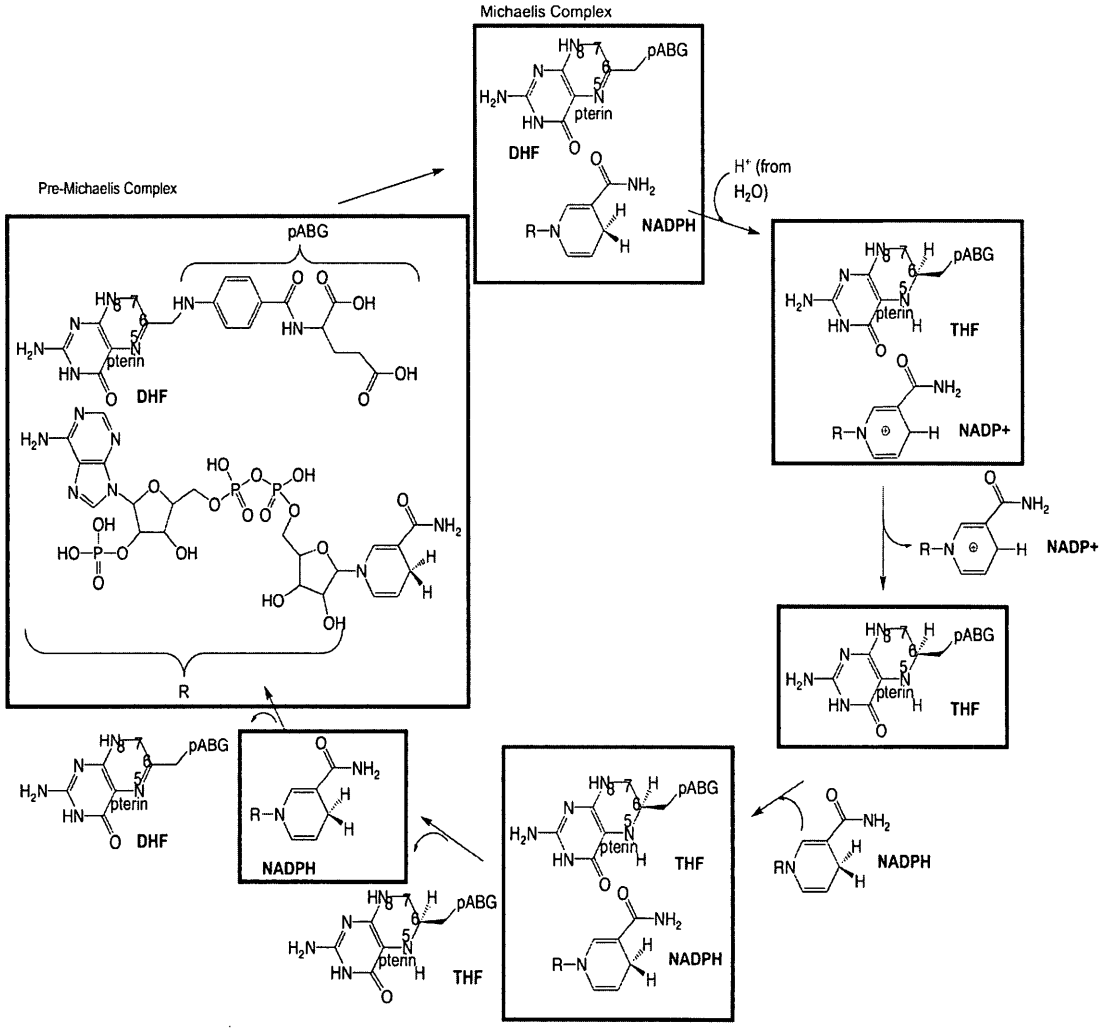


Figure 3.4: Catalytic cycle of DHFR. The boxed molecules indicate the various complexes of DHFR.

The active site structure of DHFR is maintained by the Met20 loops through hydrogen bonding with the F-G and G-H loops (Figure 3.5) [117]. It is this hydrogen-bonding network that determines the conformation of the Met20 loop and intervenes ligand binding. Various X-ray crystal structures of DHFR show that the Met20 loop exists in an occluded, closed, open conformations, and disordered conformations. However, the occluded and closed conformations occur most often [117]. Furthermore, NMR spectroscopy data demonstrate that only the closed and occluded conformations exist in solution with no evidence of the open conformation being present [118]. The occluded conformation is seen in the DHFR:$NADP^+$:THF, DHFR:THF, and DHFR:NADPH:THF complexes. The closed conformation is apparent in the DHFR:NADPH:DHF and DHFR:NADPH complexes [119].

Many studies of *E. coli* DHFR have shown that dynamics in the protein structure are important for function and catalysis [120]. A primary method for studying these dynamic motions and relating them to protein function and catalysis is by NMR relaxation techniques [111, 115, 121-123]. Previous studies on DHFR have shown that conformational changes in the active site loop are critical in mediating cofactor and substrate binding and product release. Further evidence of conformational differences in DHFR complexes arise from backbone dynamics obtained from ^{15}N NMR spin relaxation data on the closed DHFR:$NADP^+$:folate complex and occluded complexes. The Met20 loop and residues 119-123 in the F-G loop experience considerable motion and rotational freedom in the occluded complexes but become significantly diminished in the closed DHFR:$NADP^+$:folate

complex [122]. Moreover, it has been reported that replacing residues 16-19 of the Met20 loop with a single glycine resulted in a 500-fold reduction in hydride transfer without affecting the Michealis constant for DHF and NADPH [124]. This result emphasizes the importance of the Met20 loop in the catalytic cycle of DHFR. For these reasons, understanding how the dynamics of DHFR is affected when bound to an inhibitor and how the altered dynamics affects the protein structure and function is important for therapeutic development.

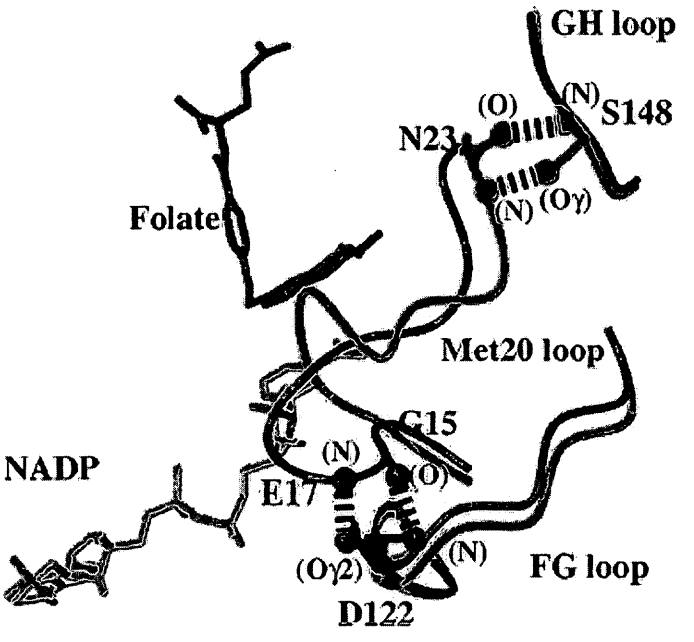
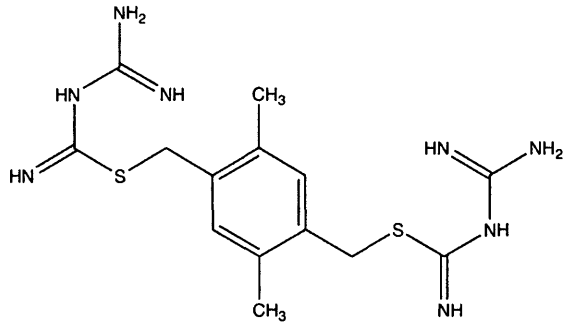


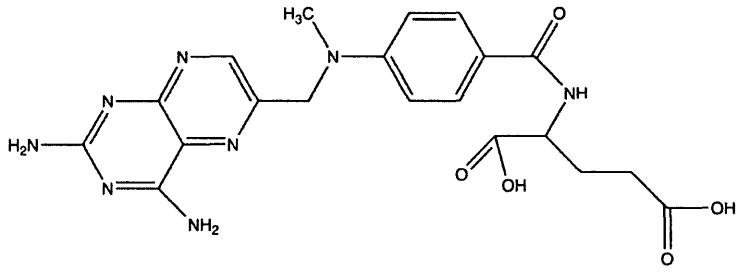
Figure 3.5: Hydrogen bonding network between the Met20, F-G, and G-H loops. The closed conformation of the Met20 loop is indicated in red and the occluded conformation of the Met20 loop is indicated in green [118].

3.1 RESEARCH GOALS

DHFR is one of the most studied enzymes to date, with its structure, function, and dynamics being extensively investigated. However, less has been reported on how inhibitors alter DHFR dynamics [114]. Establishing small molecules that act as antifolates against DHFR and understanding how these inhibitors affect the dynamics of DHFR is highly beneficial for drug discovery [112]. Most compounds known to be competitive inhibitors of *E. coli* DHFR are either 2,4-diaminoquinazoline or 2,4-diaminopyrimidine derivatives [114]. However, a recent inhibitor, 1,4-Bis-{[*N*-(1-imino-1-guanidino-methyl)]sulfanylmethyl}-3,6-dimethyl-benzene (from here on referred to as compound **1**) (Figure 3.6), has been identified by us in collaboration with Dr. Michael Organ, that does not contain the usual diaminoheterocycle known to be required for DHFR inhibition but instead contains two isothiourea groups that allows for new therapeutic possibilities [116]. This novel inhibitor binds DHFR:NADPH with a K_d of 13 ± 5 nM, has positive binding cooperativity with the NADPH cofactor, and binds to DHFR 285-fold tighter in the presence of the NADPH cofactor compared to in its absence [116]. The affinity of DHFR:NADPH for **1** is comparable with known inhibitors of DHFR such as methotrexate (MTX) (Figure 3.6), whose K_d is from 10^{-8} M to less than 10^{-11} M. Inhibitor **1** is particularly unusual because it interacts with the Met20 loop of DHFR, a region other DHFR inhibitors have not been seen to interact with.



1



MTX

Figure 3.6: Chemical structure of inhibitors **1** and MTX respectively.

The aim of this study was to better understand the enzyme function of *E. coli* DHFR by comparing the dynamics of the free enzyme and when bound to the recently discovered inhibitor **1**. ^{1}H-^{15}N NMR relaxation methods was used to investigate the dynamics of *E. coli* DHFR by comparing the change in molecular motions that occur between free DHFR:NADPH and when bound to inhibitor **1**.

In addition, ITC was used to investigate the thermodynamics of DHFR upon binding to inhibitor **1** and the established inhibitor MTX. To achieve this goal, enthalpies of binding were determined by ITC over a range of temperatures, allowing for estimates of heat capacity changes. The purpose of this data was to provide an understanding of the energetic contributions to the stability of the protein-inhibitor complexes of DHFR that is currently not known. More specifically, this part of the project aims to determine if the known X-ray structure is the same structure that binds inhibitor **1** and MTX. This study was important as it can lead to the development of new and better inhibitors of DHFR with different modes of action.

3.2 MATERIALS AND METHODS

3.2.1 ^{15}N Minimal Media Cell Preparation

BL21 (pLysS) *E. coli* cells were transformed with a pFW117.1 plasmid containing the *folA* gene coding for *E. coli* DHFR. Cells were grown at 37°C in M9 minimal media containing 1 g/L of 99% $(^{15}NH_4)_2SO_4$ (Cambridge Isotope Laboratories, Inc.). Over expression was induced with 1 mM IPTG when the cells reached OD_{600} = 0.6-0.7 and subsequently allowed to grow for 3 hours. The cells were harvested by centrifugation and suspended in buffer (20 mM Tris (pH 6.2), 0.5 mM EDTA, 0.5 mM DTT, 0.02% (w/v) NaN_3).

3.2.2 Protein Purification

Cells were lysed by sonication, centrifuged, and the lysate filtered using a 0.45 μm filter followed by a 0.2 μm filter. The lysate was loaded onto a quaternary (Q) anion-exchange column (GE Healthcare) pre-equilibrated with buffer (20 mM Tris (pH 6.2), 0.5 mM EDTA, 0.5 mM DTT, 0.02% (w/v) NaN_3) and eluted at 0.4 M NaCl over a NaCl gradient of 0.1-1.0 M. Fractions containing ^{1}H-^{15}N DHFR were pooled and concentrated using 5,000 MWCO concentrators (Amicon) and exchanged into buffer (50 mM sodium phosphate (pH 7.0), 1 M $(NH_4)_2SO_4$, 0.02% (w/v) NaN_3) that was also used to pre-equilibrate a hydrophobic interaction column (GE Healthcare). The protein was loaded onto the hydrophobic interaction column and eluted at 0.5 mM $(NH_4)_2SO_4$ over a 1.0-0.0 M $(NH_4)_2SO_4$ gradient. The fractions

containing ^{1}H-^{15}N DHFR were pooled and concentrated to 2 mL, and loaded onto a Sephacryl S-100 size exclusion column (GE Healthcare). DHFR was isolated with 100 mM sodium phosphate (pH 7.0), 150 mM NaCl, 5mM $MgCl_2$, 0.5 mM EDTA, and 0.02% (w/v) NaN_3. The fractions containing ^{1}H-^{15}N DHFR were pooled and concentrated. The final volume of purified ^{1}H-^{15}N DHFR was 500 μL.

3.2.3 NMR Sample Preparation

The final NMR sample contained 3 mM ^{1}H-^{15}N DHFR in NMR buffer (50 mM sodium phosphate, 100 mM KCl, 1 mM EDTA, 5 mM DTT, and 0.02% (w/v) NaN_3) that was deoxygenated by freeze-pump-thaw cycle under vacuum and then purged with argon. Subsequently, 10% (v/v) D_20, 60 mM NADPH and, when applicable, 4.5 mM of inhibitor **1** (Maybridge, Trevilett, UK) was added. The samples were placed in NMR tubes, purged with argon using a long needle inserted into the sample, and sealed with paraffin around the capped tube.

3.2.4 NMR Spectroscopy - Relaxation Dispersion

All ^{1}H-^{15}N NMR spectra were acquired at 25°C on a Varian 600 MHz spectrometer. Relaxation rates were measured using T_1, T_2, $\{^{1}H\}$-^{15}N NOE, and CPMG pulse sequences for both the DHFR:NADPH (free) and DHFR:NADPH:**1** bound complexes .

T_1 values were obtained at time delays τ = 10, 50, 90, 150, 210, 310, 410, and 510 ms. T_2 values were obtained at time delays τ = 10, 30, 50, 70, 90, 110, 130, and

150 ms. To test the reliability of the data, a repeat spectrum was recorded at 10 ms for both T_1 and T_2 experiments. To prevent sample heating, recycle delays of 1.5 s were used for both T_1 and T_2 during data acquisition. $\{^1H\}$-^{15}N NOE values were recorded from spectra in the presence and absence of proton presaturation during relaxation period of τ = 3.0 s.

The constant-time relaxation compensated CPMG dispersion experiments had a total relaxation period of 40 ms in all experiments. CPMG spectra were recorded with delays τ = 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34 ms. The effective field strength was modulated by changing the delay time, τ, allowing for the CPMG spectra to be collected at a series of 19 2D data sets with CPMG field strengths, ν_{CP}, of 25, 51, 76, 102, 129, 155, 182, 209, 237, 265, 293, 321, 350, 379, 409, 439, and 469 Hz, with a repeat experiment performed at ν_{CP} = 25 Hz and reference spectra obtained by omitting the CPMG period in the pulse sequence such that ν_{CP} = 0 Hz. 4 scans per FID were recorded. ^1H-^{15}N HSQC experiments were included before and after each of the four types of relaxation experiments to monitor the stability of the NMR sample through the duration of the experiments. All spectra were recorded at ^1H and ^{15}N spectral widths of 9058.0 Hz and 1700.0 Hz respectively. Data were acquired at matrices of 1536 by 160 total points.

3.2.5 Relaxation Analysis

NMR data were processed with NMRPipe [125] and visualized with NMRDraw [125]. Peak intensities were extracted using the NMRDraw module nlinLS. For the standard relaxation experiments T_1, T_2, and $\{^1H\}$-^{15}N NOE, the peak intensities were converted to R_1, R_2, and $\{^1H\}$-^{15}N NOE relaxation rates respectively. For CPMG experiments, the extracted peak intensities were converted into decay rates, R_2^{eff}, for a given CPMG field strength, ν_{CP}. The program 'pdbinertia' (developed by Dr. A. G. Palmer, Columbia University, 1996) was used to calculate the principle moments of inertia for the atoms in the pdb file of *E. coli* DHFR (Protein Data Bank, PDB ID: 2anq) and translate the protein's center of mass so that it is located at the origin and rotated, allowing the moments of inertia to be aligned with the Cartesian axes. The program 'r2r1_tm' (developed by Dr. A. G. Palmer, Columbia University, 1996) was used to calculate global and local effective correlation times from the ^{15}N relaxation data. From the local diffusions coefficients, the rotational diffusion tensor of the protein was calculated using the 'quadratic_diffusion' program. The newly generated pdb file corresponding to the proper tensor diffusion model was then used an input file for ModelFree 4.15 program using the Lipari and Szabo 'FAST-Modelfree' graphical user interface program software (Facile Analysis and Statistical Testing for Modelfree).

3.2.6 Determining ΔC_p of DHFR:NADPH:Inhibitor by ITC

ITC experiments were performed using a MicroCal VP-ITC instrument. All samples were prepared in 50 mM potassium phosphate (pH 6.8), 100 mM KCl, 1 mM EDTA, 5 mM DTT. The buffer was purged with Argon and degassed using the MicroCal ThermoVac. Titrations were performed by injecting with either 3 consecutive 8 μL aliquots of ligand solution (inhibitor **1** or MTX) into the ITC cell containing DHFR with NADPH over consecutive temperatures of 10°C, 15°C, 20°C, 25°C, and 30°C. The enthalpy of reaction at each temperature was determined from the average of the three enthalpy values obtained that were extrapolated from the y-intercept of the binding isotherm curve.

3.3 RESULTS AND DISCUSSION

3.3.1 Protein Purification

In order to obtain pure ^{1}H-^{15}N DHFR, three chromatographic columns were employed. Use of an anion-exchange column, followed by a hydrophobic column and lastly a size-exclusion column resulted in pure ^{1}H-^{15}N DHFR (Figure 3.7).

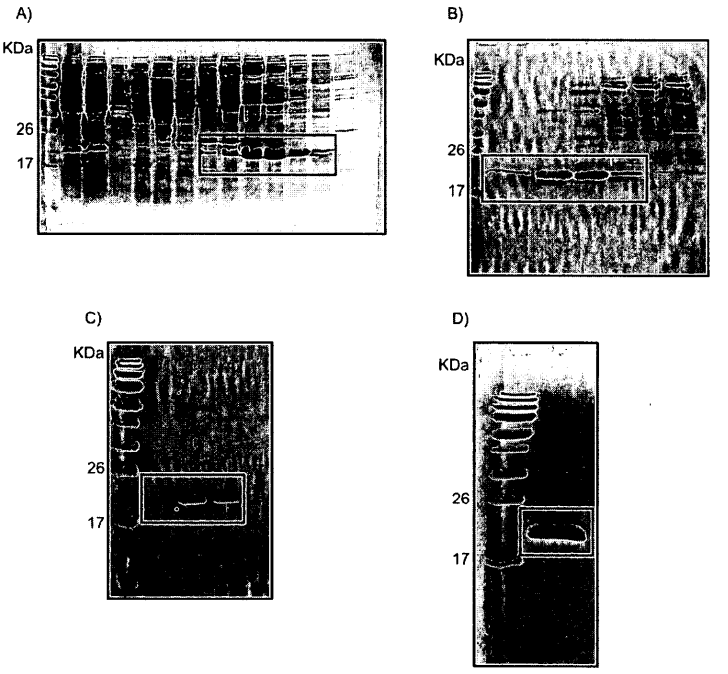


Figure 3.7: Purification scheme for ^{1}H-^{15}N DHFR involved three columns: A) anion exchange, B) hydrophobic interaction, and C) size exclusion to yield D) the pure protein sample. ^{1}H-^{15}N DHFR (21 KDa) is boxed in red.

3.3.2 Chemical Changes of DHFR:NADPH Upon Binding of Inhibitor 1

The majority of studies interested in investigating the dynamic properties of proteins and their alteration upon molecular interaction have been achieved using the amide ^{15}N nucleus that is attached to a ^{1}H nucleus in an isolated two-spin system, allowing for easier probing of backbone motions [71]. Prior to the setup of any relaxation experiments, a 2D ^{1}H-^{15}N HSQC (heteronuclear single quantum coherence) spectrum of ^{1}H-^{15}N DHFR was acquired to ensure that protein was detectable in the NMR. Confirmation of detectable protein allowed for addition of NADPH cofactor to the sample. The acquired spectrum of DHFR:NADPH showed better defined peaks compared to DHFR in the absence of NADPH, suggests that the cofactor binding to DHFR provides important conformational stability (Figure 3.8). The ^{1}H-^{15}N HSQC experiment is useful for detecting each unique proton attached to the nitrogen that is being considered, resulting in a peak in the HSQC spectrum for each amino acid along the protein's amide backbone. The HSQC spectrum of the binary complex indicated that the expected number of peaks were present for the 159 residue protein.

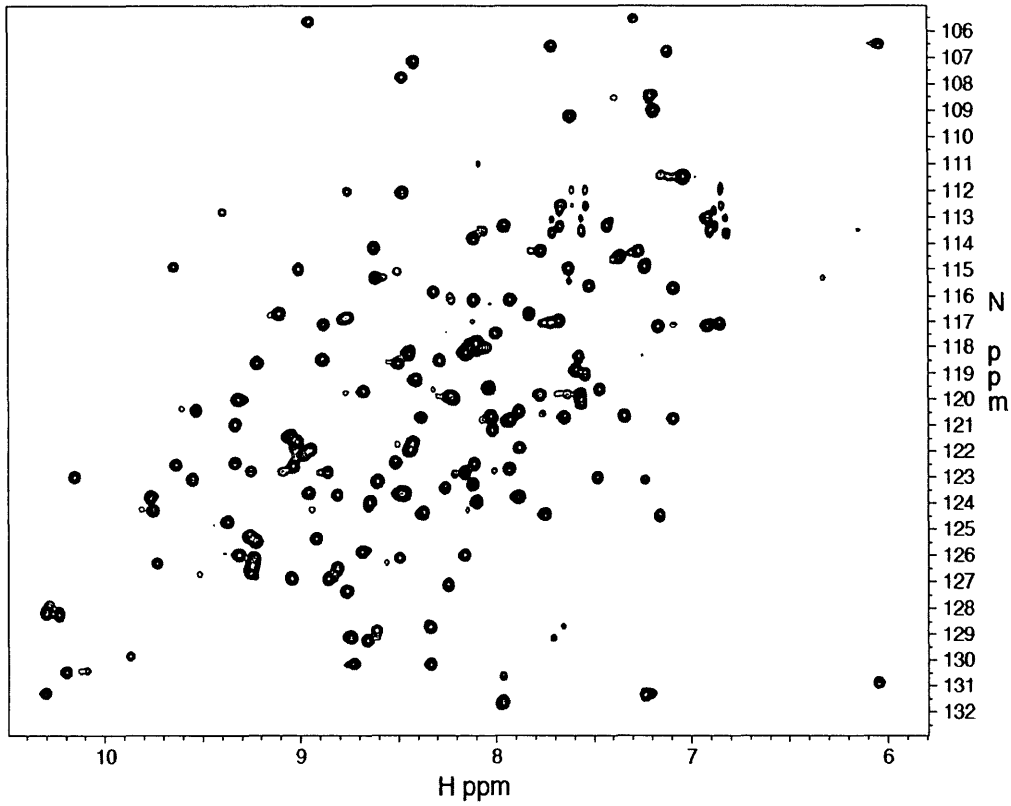


Figure 3.8: ^{1}H-^{15}N HSQC spectrum of the DHFR:NADPH complex.

Addition of inhibitor **1** to DHFR:NADPH resulted in a ^{1}H-^{15}N HSQC spectrum of the ternary complex (Figure 3.9) that showed chemical shift changes of specific residues when compared to the spectrum of the binary complex (Figure 3.10).

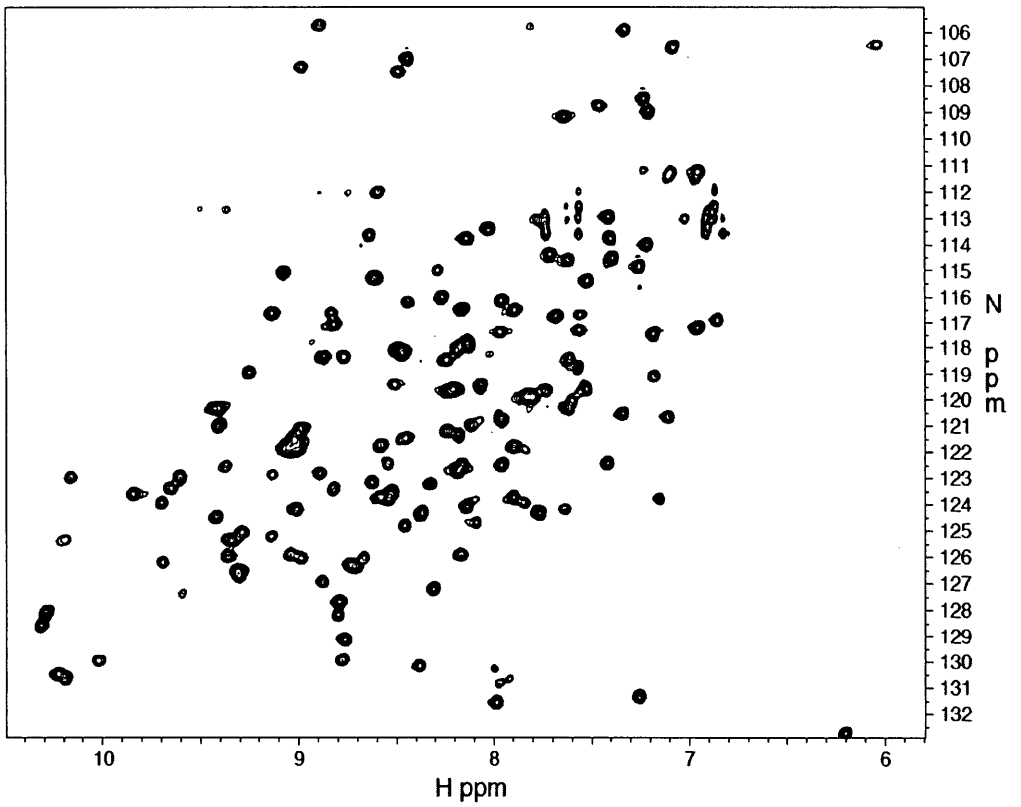


Figure 3.9: ^{1}H-^{15}N HSQC spectrum of the DHFR:NADPH:**1** complex.

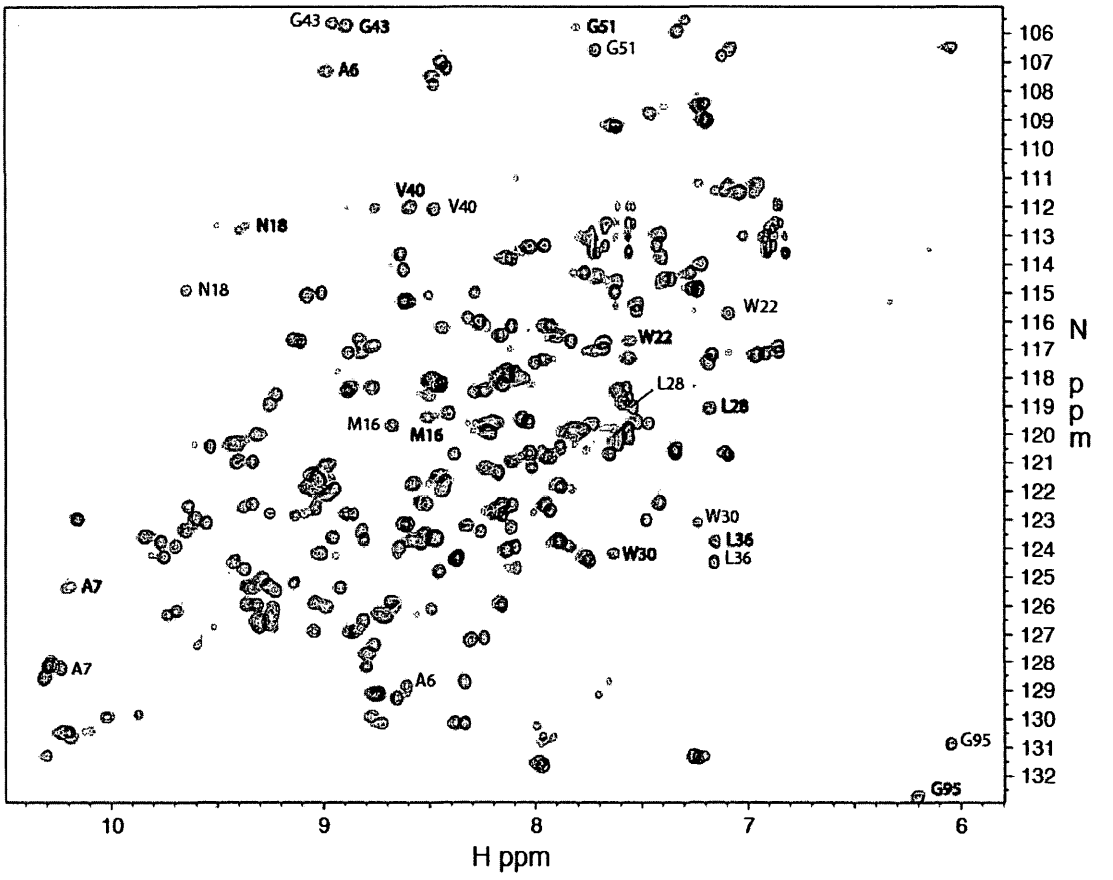


Figure 3.10: Overlay of DHFR:NADPH:**1** complex ^{1}H-^{15}N HSQC (red) and DHFR:NADPH complex ^{1}H-^{15}N HSQC (black).

The magnitudes of the chemical shift changes for all residues in the ^{1}HN and ^{15}N dimension for DHFR are listed in Table 3.1 and illustrated in Figures 3.11-3.12.

Table 3.1: List of backbone ^{1}HN and ^{15}N chemical shift changes in DHFR complexes (DHFR:NADPH and DHFR:NADPH:**1**).

Residue	DHFR:NADPH δ ^{1}HN (ppm)	DHFR:NADPH δ ^{15}N (ppm)	DHFR:NADPH:1 δ ^{1}HN (ppm)	DHFR:NADPH:1 δ ^{15}N (ppm)	Δδ ^{1}HN (ppm)	Δδ ^{1}N (ppm)
Met 1	----	----	----	----	----	----
Ile 2	9.368	124.774	9.419	124.461	-0.051	0.313
Ser 3	9.318	126.053	9.359	125.939	-0.041	0.114
Leu 4	8.435	121.987	8.586	121.76	-0.151	0.227
Ile 5	----	----	----	----	----	----
Ala 6	8.608	128.925	8.894	105.695	-0.286	23.23
Ala 7	10.295	128.186	10.195	125.371	0.1	2.815
Leu 8	9.039	122.613	8.988	121.703	0.051	0.91
Ala 9	8.492	126.167	8.665	126.053	-0.173	0.114
Val 10	7.574	118.376	7.621	118.461	-0.047	-0.085
Asp 11	9.089	122.84	9.133	122.897	-0.044	-0.057
Arg 12	8.479	107.77	8.492	107.486	-0.013	0.284
Val 13	7.097	120.793	7.112	120.679	-0.015	0.114
Ile 14	----	----	----	----	----	----
Gly 15	7.288	105.552	7.335	105.922	-0.047	-0.37
Met 16	8.677	119.684	8.479	119.542	0.198	0.142
Glu 17	10.091	130.46	10.195	130.659	-0.104	-0.199
Asn 18	9.645	114.879	9.398	113.221	0.247	1.658
Ala 19	----	----	----	----	----	----
Met 20	----	----	----	----	----	----
Pro 21	----	----	----	----	----	----
Trp 22	7.09	115.732	7.559	116.67	-0.469	-0.938
Asn 23	8.79	116.926	8.935	117.807	-0.145	-0.881
Leu 24	----	----	----	----	----	----
Pro25	----	----	----	----	----	----
Ala 26	9.337	121.02	9.403	121.02	-0.066	0

Asp 27	7.725	117.04	7.568	117.324	0.157	-0.284
Leu 28	7.543	119.058	7.185	119.058	0.358	0
Ala 29	----	----	----	----	----	----
Trp 30	7.232	123.153	7.637	124.177	-0.405	-1.024
Phe 31	----	----	----	----	----	----
Lys 32	----	----	----	----	----	----
Arg 33	7.832	116.727	7.892	116.528	-0.06	0.199
Asn 34	7.037	111.524	6.955	111.268	0.082	0.256
Thr 35	7.119	106.775	7.084	106.519	0.035	0.256
Leu 36	7.159	124.489	7.15	123.778	0.009	0.711
Asn 37	8.319	115.902	8.288	115.021	0.031	0.881
Lys 38	7.634	119.826	7.738	119.684	-0.104	0.142
Pro 39	----	----	----	----	----	----
Val 40	8.476	112.092	8.595	112.007	-0.119	0.085
Ile 41	8.423	121.703	8.46	121.447	-0.037	0.256
Met 42	8.957	123.693	9.02	124.177	-0.063	-0.484
Gly 43	8.96	105.666	8.991	107.344	-0.031	-1.678
Arg 44	7.59	118.916	7.568	118.803	0.022	0.113
His 45	----	----	----	----	----	----
Thr 46	8.02	121.219	8.237	121.219	-0.217	0
Trp 47	----	----	----	----	----	----
Glu 48	----	----	----	----	----	----
Ser 49	7.769	114.367	7.716	114.452	0.053	-0.085
Ile 50	----	----	----	----	----	----
Gly 51	7.713	106.519	7.713	105.589	0	0.93
Arg 52	6.855	117.097	6.855	116.954	0	0.143
Pro 53	----	----	----	----	----	----
Leu 54	9.73	126.366	9.692	126.252	0.038	0.114
Pro 55	----	----	----	----	----	----
Gly 56	----	----	----	----	----	----
Arg 57	7.166	117.239	7.185	117.466	-0.019	-0.227

Lys 58	----	----	----	----	----	----
Asn 59	8.922	125.428	9.045	125.911	-0.123	-0.483
Ile 60	8.809	126.565	8.731	126.366	0.078	0.199
Ile 61	8.762	127.418	8.793	127.702	-0.031	-0.284
Leu 62	8.159	126.025	8.174	125.939	-0.015	0.086
Ser 63	----	----	----	----	----	----
Ser 64	10.154	123.039	10.16	122.954	-0.006	0.085
Gln 65	8.602	123.21	8.627	123.153	-0.025	0.057
Pro 66	----	----	----	----	-----	----
Gly 67	7.625	109.249	7.64	109.192	-0.015	0.057
Thr 68	6.044	106.519	6.038	106.462	0.006	0.057
Asp 69	7.345	120.679	7.339	120.537	0.006	0.142
Asp 70	8.856	126.963	8.878	126.935	-0.022	0.028
Arg 71	8.885	118.461	8.878	118.376	0.007	0.085
Val 72	7.203	108.481	7.238	108.51	-0.035	-0.029
Thr 73	7.998	117.523	7.967	117.438	0.031	0.085
Trp 74	8.746	129.181	8.765	129.181	-0.019	0
Val 75	9.108	116.727	9.139	116.67	-0.031	0.057
Lys 76	8.228	116.158	8.266	116.016	-0.038	0.142
Ser 77	7.191	109.078	7.21	109.021	-0.019	0.057
Val 78	8.859	122.869	8.897	122.812	-0.038	0.057
Asp 79	----	----	----	----	----	----
Glu 80	7.936	120.85	7.958	120.793	-0.022	0.057
Ala 81	----	----	----	----	----	----
Ile 82	7.926	116.187	7.964	116.215	-0.038	-0.028
Ala 83	8.096	124.006	8.14	124.091	-0.044	-0.085
Ala 84	8.036	119.57	8.074	119.485	-0.038	0.085
Cys 85	7.37	114.594	7.395	114.594	-0.025	0
Gly 86	7.229	131.399	7.263	131.342	-0.034	0.057
Asp 87	8.511	122.471	8.542	122.442	-0.031	0.029
Val 88	7.235	114.936	7.263	114.879	-0.028	0.057

Pro 89	----	----	----	----	----	----
Glu 90	7.678	117.011	7.681	116.755	-0.003	0.256
Ile 91	8.727	130.176	8.775	129.977	-0.048	0.199
Met 92	7.882	121.873	7.898	121.76	-0.016	0.113
Val 93	9.224	125.485	9.343	125.371	-0.119	0.114
Ile 94	8.498	118.717	8.771	118.348	-0.273	0.369
Gly 95	6.044	130.944	6.204	132.764	-0.16	-1.82
Gly 96	----	----	----	----	----	----
Gly 97	7.389	108.566	7.458	108.794	-0.069	-0.228
Arg 98	9.761	123.807	9.849	123.608	-0.088	0.199
Val 99	7.48	123.068	7.42	122.442	0.06	0.626
Tyr 100	9.532	120.452	9.434	120.31	0.098	0.142
Glu 101	8.112	116.187	8.162	116.556	-0.05	-0.369
Gln 102	7.266	114.338	7.216	113.997	0.05	0.341
Phe103	8.115	113.855	8.143	113.798	-0.028	0.057
Leu104	----	----	----	----	----	----
Pro 105	----	----	----	----	----	----
Lys 106	----	----	----	----	----	----
Ala 107	7.936	122.698	7.958	122.499	-0.022	0.199
Gln 108	9.221	118.632	9.255	118.973	-0.034	-0.341
Lys 109	7.766	120.594	7.788	119.997	-0.022	0.597
Leu110	----	----	----	----	----	----
Tyr 111	9.334	122.499	9.372	122.527	-0.038	-0.028
Leu112	9.632	122.556	9.654	123.38	-0.022	-0.824
Thr113	----	----	----	----	----	----
His 114	----	----	----	----	----	----
Ile 115	9.048	126.906	8.988	126.053	0.06	0.853
Asp116	8.335	130.205	8.388	130.148	-0.053	0.057
Ala 117	7.747	124.489	7.766	124.376	-0.019	0.113
Glu 118	8.498	123.665	8.507	119.428	-0.009	4.237
Val 119	8.247	119.911	8.269	119.684	-0.022	0.227

Glu 120	8.372	124.404	8.372	124.376	0	0.028
Gly 121	8.413	107.173	8.438	106.974	-0.025	0.199
Asp122	8.753	116.812	8.834	116.642	-0.081	0.17
Thr123	7.521	115.675	7.521	115.419	0	0.256
His 124	9.513	126.792	9.592	127.418	-0.079	-0.626
Phe125	9.262	125.314	9.29	125.086	-0.028	0.228
Pro 126	----	----	----	----	----	----
Asp127	----	----	----	----	----	----
Tyr 128	7.471	119.684	7.533	119.599	-0.062	0.085
Glu 129	8.812	123.75	8.822	123.466	-0.01	0.284
Pro 130	----	----	----	----	----	----
Asp131	9.007	115.021	9.079	115.078	-0.072	-0.057
Asp132	----	----	----	----	----	----
Trp133	7.772	119.855	7.832	119.855	-0.06	----
Glu 134	9.551	123.153	9.604	122.982	-0.053	0.171
Ser 135	----	----	----	----	----	----
Val 136	----	----	----	----	----	----
Phe137	7.879	123.807	7.898	123.778	-0.019	0.029
Ser 138	----	----	----	----	----	----
Glu 139	8.683	125.911	8.454	124.802	0.229	1.109
Phe140	----	----	----	----	----	----
His 141	8.118	123.295	8.184	122.755	-0.066	0.54
Asp142	8.024	120.708	8.178	121.362	-0.154	-0.654
Ala 143	----	----	----	----	----	----
Asp144	9.051	121.418	8.995	121.134	0.056	0.284
Ala 145	----	----	----	----	----	----
Gln 146	7.958	113.372	8.024	113.457	-0.066	-0.085
Asn147	8.209	119.968	8.2	119.57	0.009	0.398
Ser 148	8.881	117.21	8.815	117.097	0.066	0.113
His 149	6.918	117.153	6.955	117.21	-0.037	-0.057
Ser 150	8.621	114.253	8.636	113.685	-0.015	0.568

Tyr 151	7.621	114.992	7.618	114.651	0.003	0.341
Cys 152	----	----	----	----	----	----
Phe153	----	----	----	----	----	----
Glu 154	9.749	124.347	9.698	123.978	0.051	0.369
Ile 155	8.646	124.034	8.589	123.722	0.057	0.312
Leu156	----	----	----	----	----	----
Glu 157	9.315	120.025	9.387	120.253	-0.072	-0.228
Arg 158	8.247	127.162	8.313	127.219	-0.066	-0.057
Arg 159	7.967	131.683	7.989	131.569	-0.022	0.114

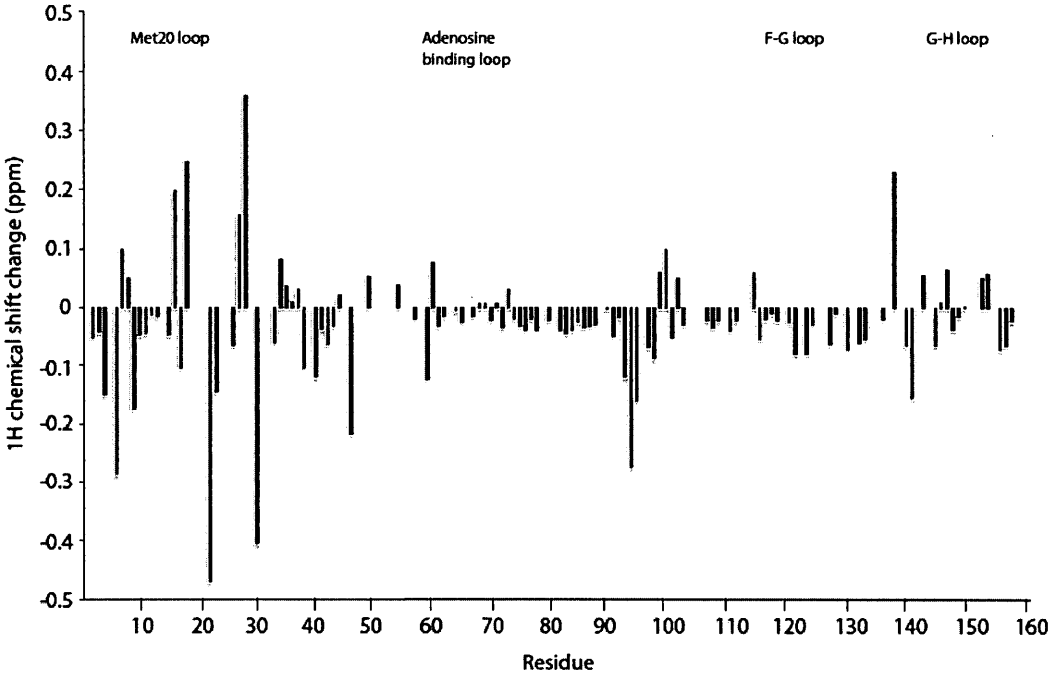


Figure 3.11: Histogram of the magnitude of ^{1}H chemical shift perturbation for each residue upon binding to inhibitor **1** to the DHFR:NADPH complex.

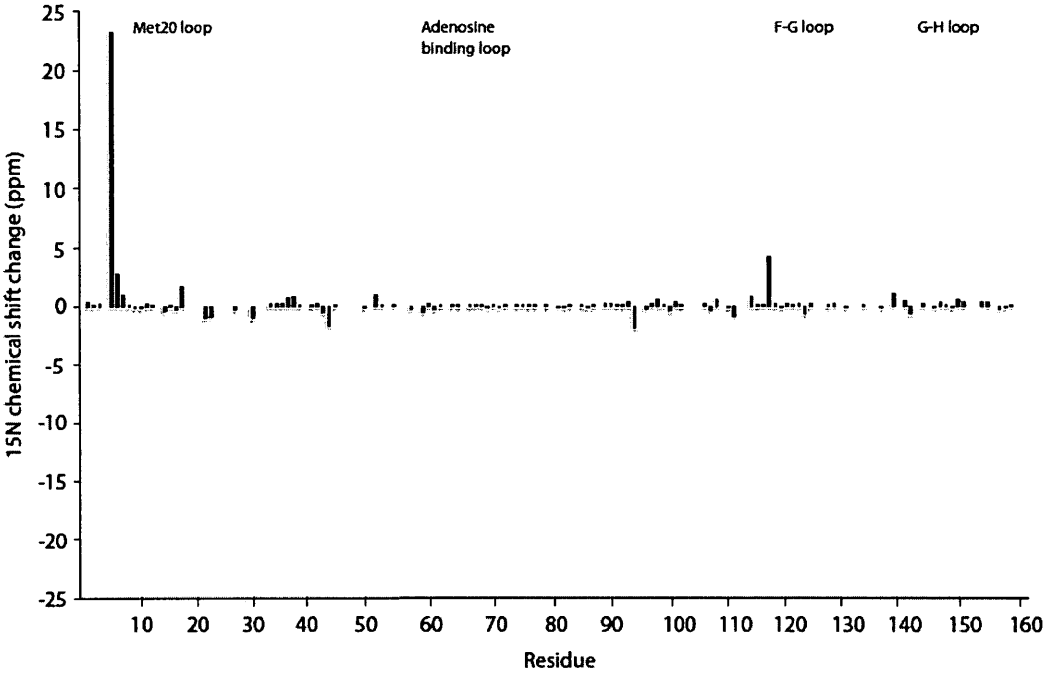


Figure 3.12: Histogram of the magnitude of ^{15}N chemical shift perturbation for each residue upon binding to inhibitor **1** to the DHFR:NADPH complex.

Significant chemical shift changes were observed for residues A6, A7, M16, N18, W22, L28, W30, L36, V40, G43, G51, and G95. All of these residues are in close proximity to the substrate (inhibitor). Analysis of the X-ray structure shows that one isothiourea group of the inhibitor interacts with a region of the protein that would form the binding pocket for the pteridine ring of the substrate and the other isothiourea group interacts with the Met20 loop (Figure 3.13) [126]. Residues M16 and N18 are part of the Met20 loop active site loop and are affected by the bound substrate. The occluded conformation in DHFR that occurs upon substrate binding is

attributed to movement of residue M16 such that it protrudes into the active site, explaining the detected chemical shift changes.

Figure 3.13: Residues (yellow) of DHFR (PDB ID: 2ANQ) [126] that are in close proximity to inhibitor 1 (blue) show significant chemical shift changes upon binding. Residue M16 in the Met20 loop is colored pink. The NADPH cofactor (red) does not appear to be considerably near to the residues showing chemical shift changes.

NMR studies by Venkitakrishnan *et al.* of alanine resonances in binary and ternary DHFR complexes with folate, showed that A6 and A7 are responsive to

substrate or to both substrate and cofactor binding [119]. The ^{15}N resonance of A6 is sensitive to the location of the pterin ring of the bound substrate or product [119]. The ^{1}HN and ^{15}N resonances of A7 are affected by hydrogen-bonding interactions with the carboxyamide group of the nicotinamide moiety of bound cofactor and to the bound substrate [119]. Residues W22, L28, and G51 are in close proximity to one of the isothiourea groups of the inhibitor, allowing for direct interactions. Residues W30, L36, V40, G43 and G95 are located closely to the second isothiourea group of the inhibitor so that direct interactions are possible.

3.3.3 Relaxation Analysis of DHFR:NADPH

Relaxation dispersion NMR spectroscopy was used to understand the dynamics of DHFR. In addition to being a powerful tool for characterizing the mechanisms of protein folding, this approach is often used for acquiring a detailed description of intermediate states in a protein's folding pathway [127]. Information about protein dynamics was obtained by performing T_1, T_2, $\{^{1}H\}$-^{15}N NOE, and CPMG relaxation experiments of both the binary and ternary complexes and then comparing the changes in dynamics parameters. Relaxation data was obtained for 121 residues out of 159 residues in the DHFR:NADPH complex (excluding 10 prolines and N-terminal methionine) (Figures 3.14-3.17). The T_1 values ranged from 428-876 ms with an associated error range of 2-15%. The average T_1 value was determined to be 668 ms with an associated error of 6%. The T_2 values ranged

from 46-97 ms with an associated error range from 2-19% and the average T_2 value was determined to be 16 ms with an associated error of 6%. Residues that had hetNOE values above 1.0 were excluded from the hetNOE relaxation data analysis. HetNOE values ranged from 0.47-0.99, with an average hetNOE of 0.86. To all hetNOE values, 5% error was assigned. Visual examination of the data points show residues 6, 34, 67-69, 88, 135, and 148 have both distinctively low hetNOE values and T_1/T_2 ratios, suggesting that these residues most likely undergo motions on the pico-nanosecond timescale. Conversely, high hetNOE values and low T_2 times result in higher than average T_1/T_2 ratios, which is evident for residues 9, 11, 30, 94, 97, 121, 128, 133. These values show a possible indication for motions on the micro-millisecond timescale.

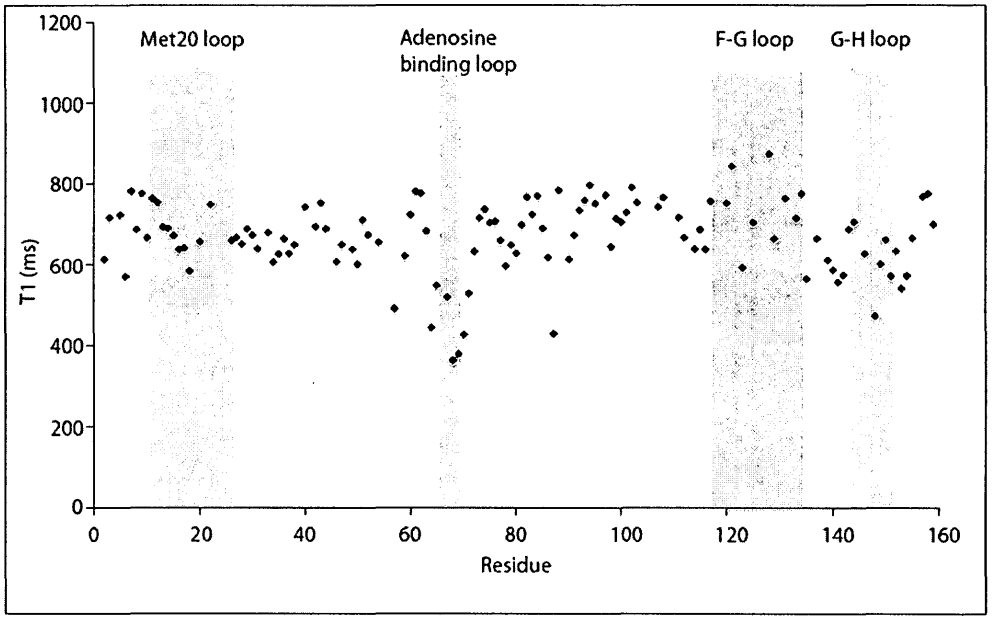


Figure 3.14: T_1 relaxation times for residues in the DHFR:NADPH complex.

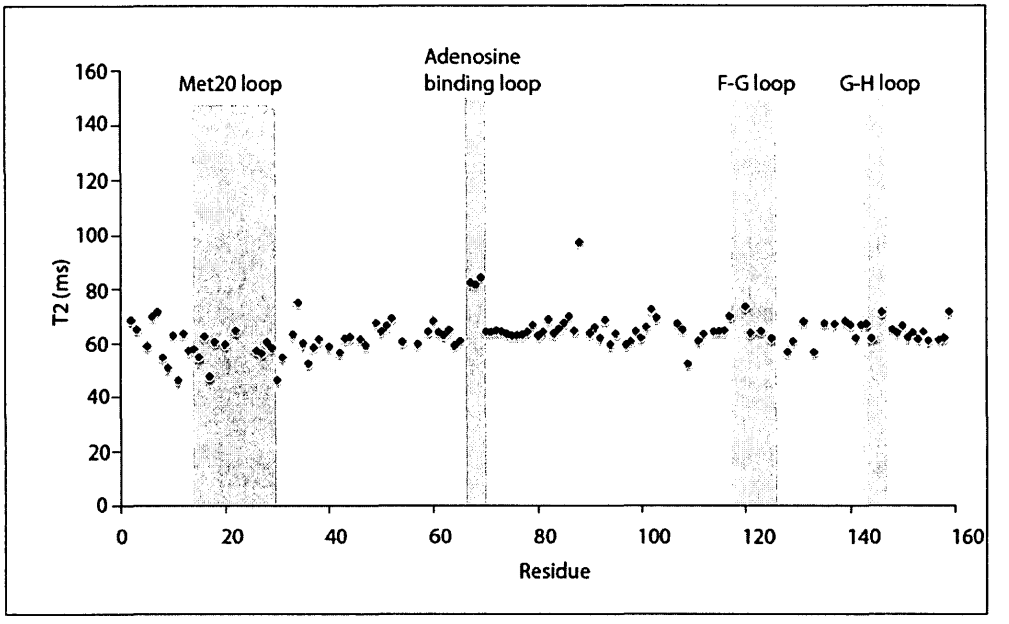


Figure 3.15: T_2 relaxation times for residues in the DHFR:NADPH complex.

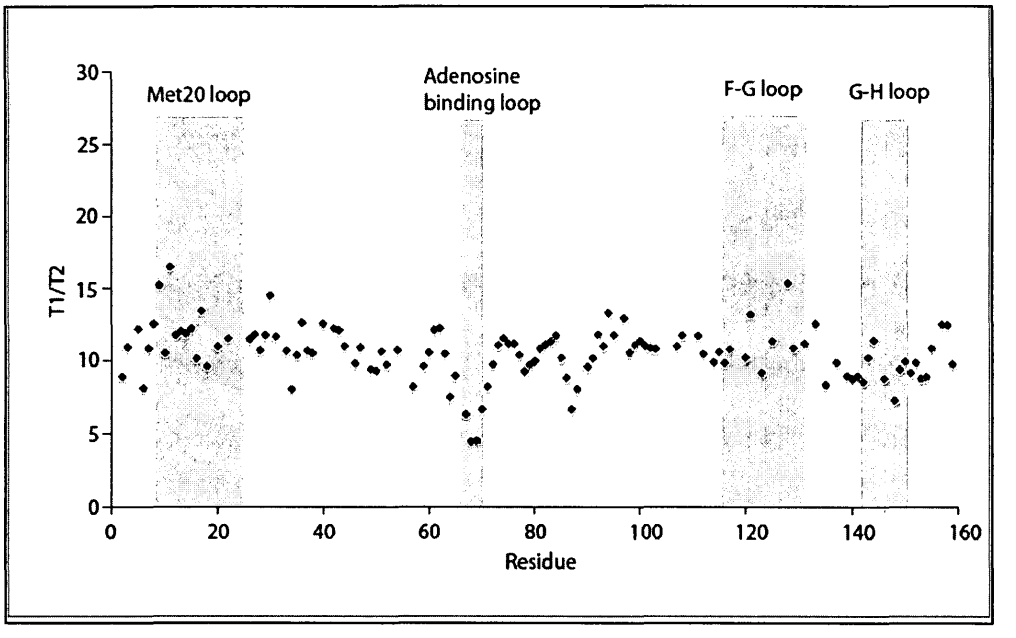


Figure 3.16: T_1/T_2 relaxation time ratios for residues in the DHFR:NADPH complex.

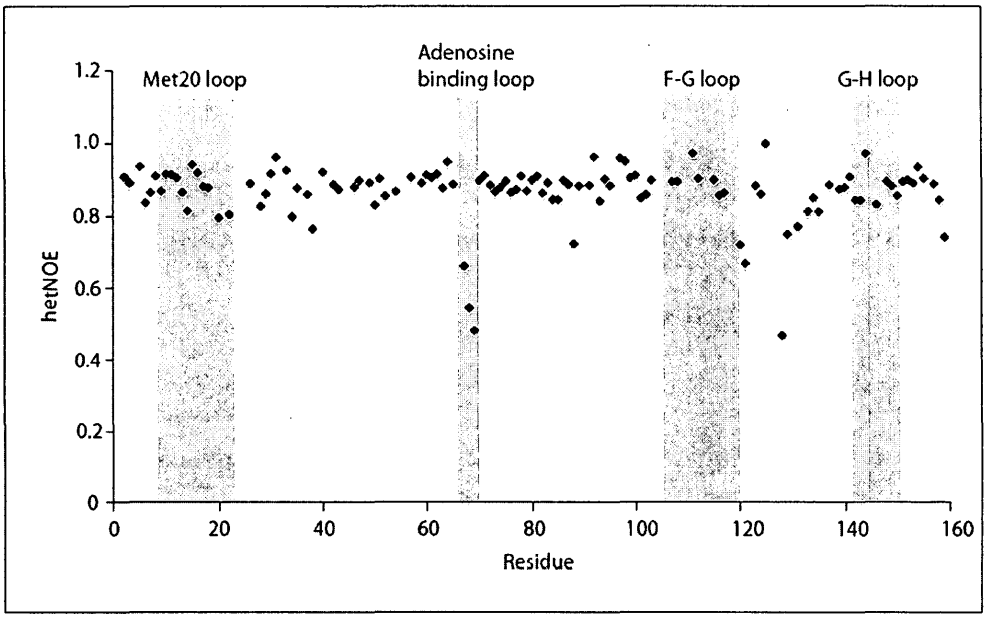


Figure 3.17: HetNOE for residues in the DHFR:NADPH complex.

3.3.4 Relaxation Analysis of DHFR:NADPH:1

Relaxation data was obtained for 106 residues out of a total of 159 residues in the DHFR:NADPH:**1** complex (excluding 10 prolines and N-terminal methionine) (Figures 3.18-3.21). The T_1 values ranged from 562-945 ms with an associated error range of 2-10%. The average T_1 value was determined to be 803 ms with an associated error of 7%. The T_2 values ranged from 39-102 ms with an associated error range from 1-7% and the average T_2 value was determined to be 67 ms with an associated error of 2%. Residues that had hetNOE values above 1.0 were excluded from the hetNOE relaxation data analysis. HetNOE values ranged from 0.48-0.99, with an average hetNOE of 0.88. To all hetNOE values, 5% error was

assigned. Visual examination of the data points show residues 34, 38, 58, 67-69, 80, 88, 120, 121, 128, 129, and 131 have both distinctively low hetNOE values and T_1/T_2 ratios, suggesting that these residues most likely undergo motions on the pico-nanosecond timescale. Residues 16, 40, and 59 have high hetNOE values and residues 11, 23, 28, 36, 42, 128, 155, and 158 display high T_1/T_2 ratio, indicative of motions on the micro-millisecond timescale. T_1 and T_2 relaxation data suggests that the DHFR:NADPH complex shows overall motions on the micro-millisecond timescale, which are referred to as 'slow' timescale motions that is often observed due to conformational exchange or ligand binding events [128].

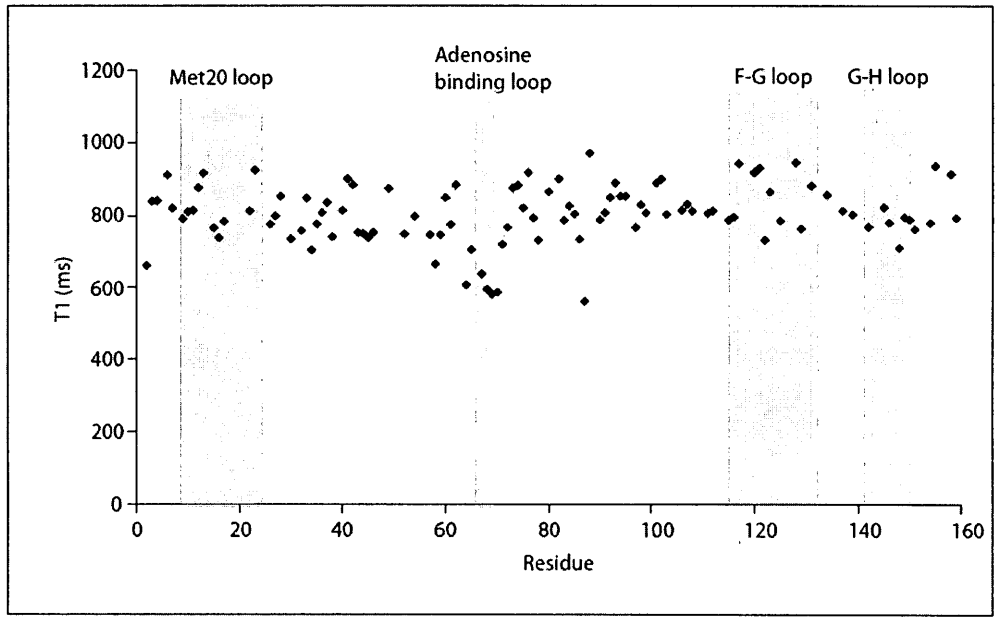


Figure 3.18. T_1 relaxation times for residues in the DHFR:NADPH:**1** complex.

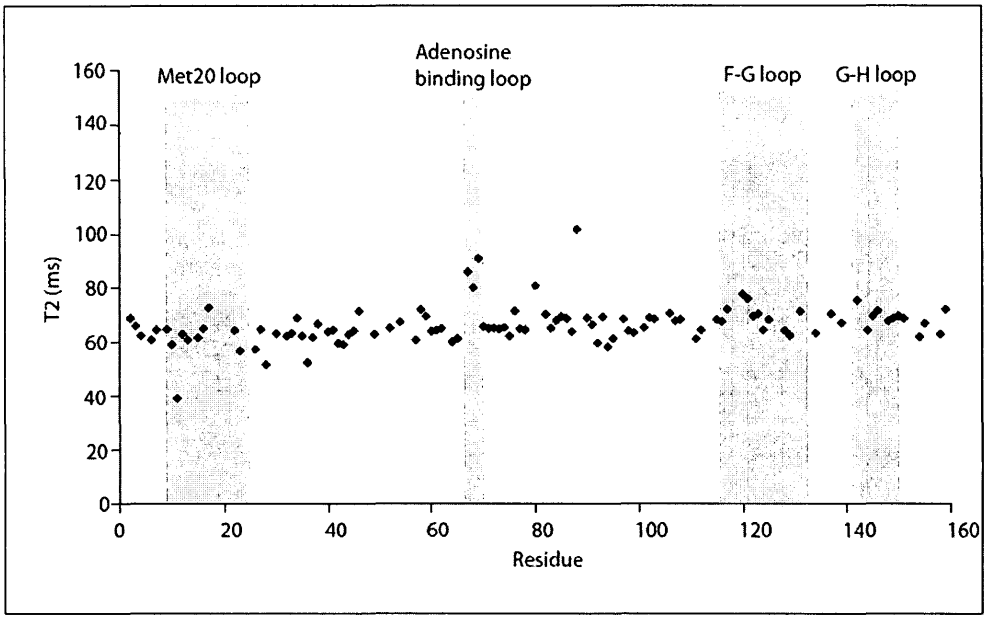


Figure 3.19: T_2 relaxation times for residues in the DHFR:NADPH:**1** complex.

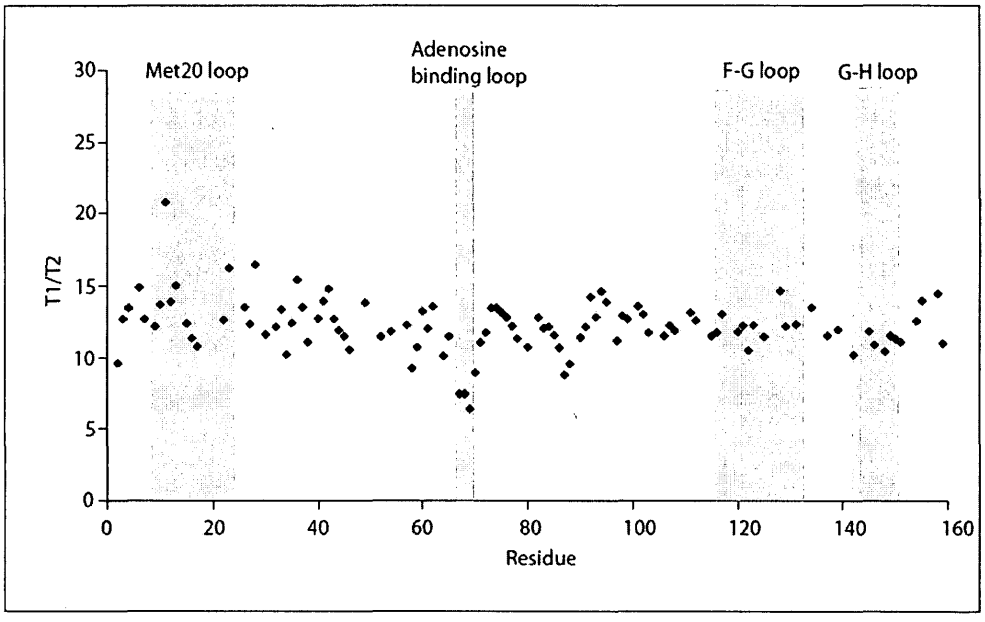


Figure 3.20: T_1/T_2 relaxation time ratios for residues in the DHFR:NADPH:**1** complex.

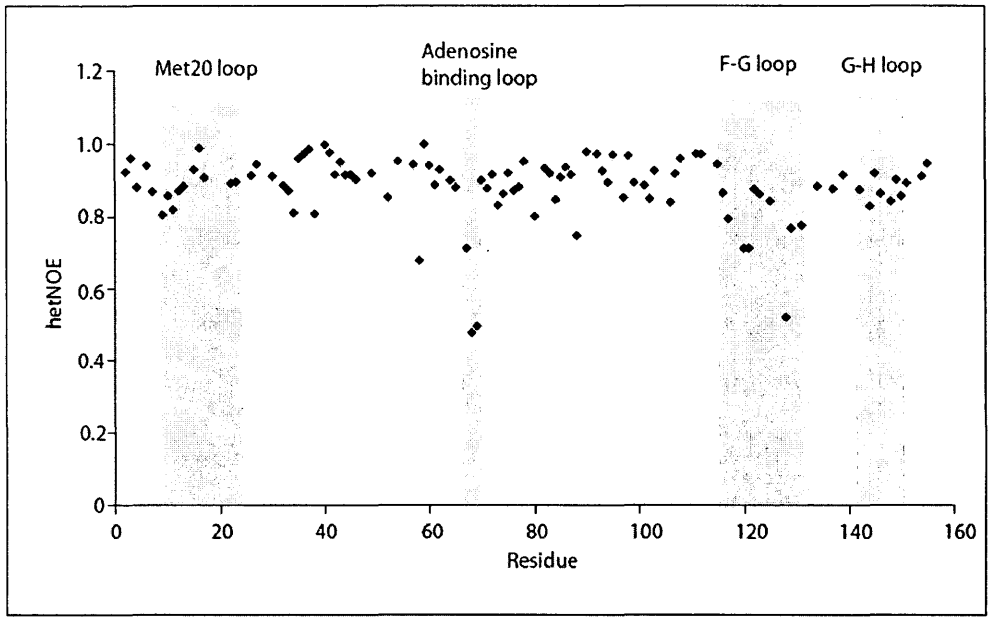


Figure 3.21: HetNOE for residues in the DHFR:NADPH:**1** complex.

3.3.5 CPMG NMR Spectroscopy for Protein Dynamics

The CPMG experiment monitors the signal intensity $I(\nu_{CPMG})$ of spins from the ground state, where the signal intensity is a function of the rate of applied refocusing pulses that are described as $\nu_{CPMG} = 1/(2\tau)$ with τ being the time between pulses. The rate of decay of the signal can be quantified by converting $I(\nu_{CPMG})$ into the effective transverse relaxation rate Eq. (11),

$$R_{eff}(\nu_{CPMG}) = -\ln(I(\nu_{CPMG})I_o)/T \quad (11)$$

where I_o is defined as the reference signal intensity. In doing so, relaxation dispersion (RD) profiles of $R_{eff}(\nu_{CPMG})$ are obtained. ^{15}N RD profiles were generated

for each residue of DHFR when the protein is free (binary complex) and when bound to inhibitor **1** (ternary complex). Since the RD profiles depend on the nonlinear rates of exchange and on the differences in chemical shift, each residue should have a unique RD profile. Comparing the RD profiles of each protein residue in the binary and ternary complexes should show a difference between the shape of the profiles and hence provide insight into which residues are involved in protein folding upon binding of the inhibitor. RD profiles of the binary complex of DHFR showed generally straight lines for all residues whereas RD profiles of the ternary complex of DHFR showed slight curves for all residues (Figure 3.22 a-b). Comparison of the binary and ternary data is inconclusive because a visible change in the profiles should be apparent for at least some protein residues only and not for all residues. Therefore no conclusions could be made regarding relaxation dispersion or protein folding of DHFR. One possible reason for this lack of difference may be that the exchange is not in the intermediate exchange range. Furthermore, acquiring CPMG data at a different frequency in addition to the frequency under which this data was obtained would largely improve this experiment. No conclusions about the protein folding of DHFR could be made from the CPMG relaxation dispersion data acquired.

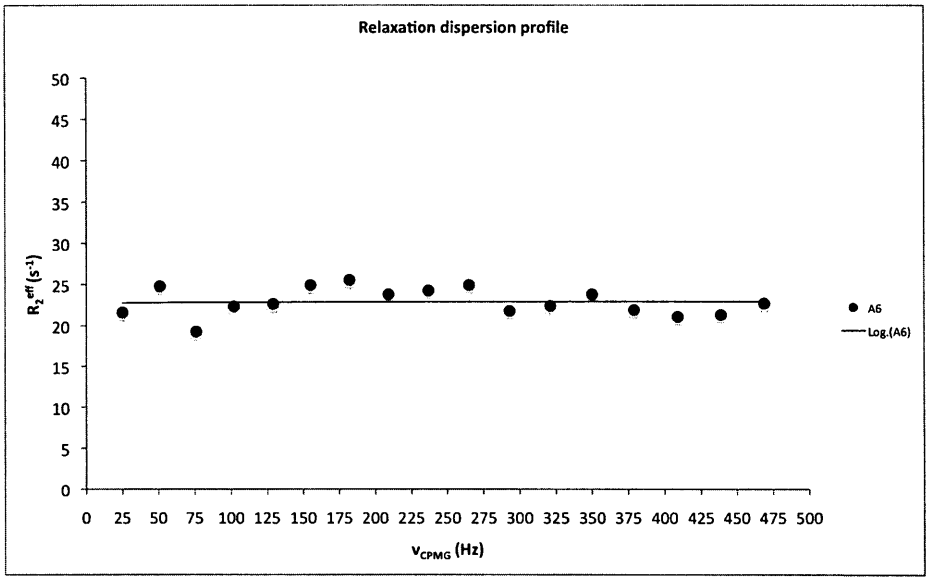


Figure 3.22 a. Relaxation dispersion profile of residue A6 of free DHFR (binary).

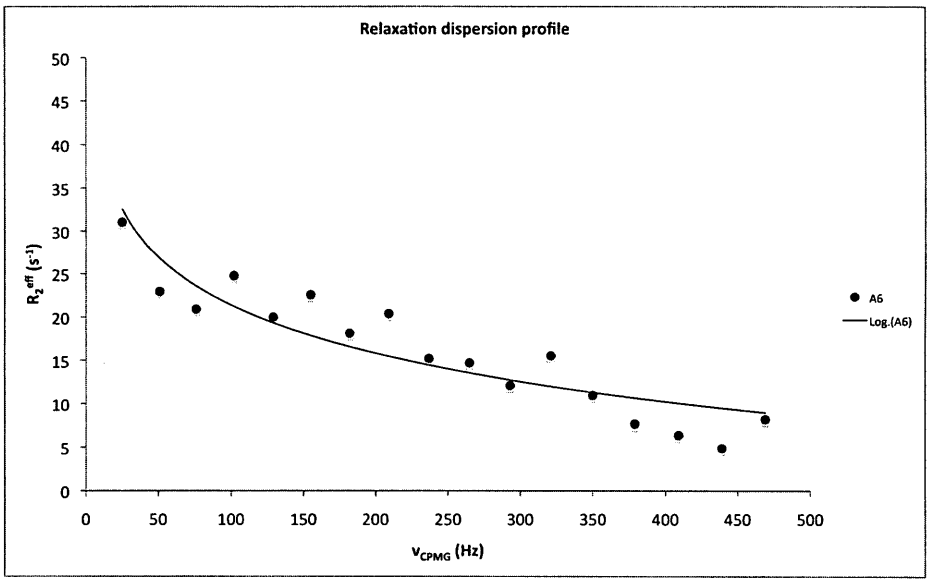


Figure 3.22 b. Relaxation dispersion profile of residue A6 of DHFR bound to inhibitor **1** (ternary).

3.3.6 Estimation of Global Correlation Time, τ_m

The global correlation time (τ_m) for the binary and ternary complexes was estimated with the program 'r2r1_tm' using the ^{15}N T_1 and T_2 relaxation data. Only residues with T_1/T_2 ratios within one standard deviation unit of the mean were used for estimation of τ_m. In the binary complex for which accurate relaxation data could be obtained, residues A6, L8, A9, D11, Q17, W30, N34, L36, V40, R57, G67-R71, D87, V88, I94, G97, G121, Y128, W133, S135, D142, E157, and R158 had T_1/T_2 ratios greater or smaller than one standard deviation unit of the mean. The estimated τ_m for the binary complex was 9.9 ns ± 0.6 ns. For the ternary complex, residues I2, A6, D11, V13, N23, L28, N34, L36, M42, K58, S64, G67-D70, D87, V88, M92, I94, Y128, D142, and R158 were eliminated for having T_1/T_2 ratios greater or smaller than one standard deviation unit of the mean. The ternary complex had an estimated τ_m of 10.8 ns ± 0.5 ns. In cases where the T_1/T_2 ratio is smaller by one standard deviation from the mean and the hetNOE value is low, it can be concluded that fast motions make significant contributions to T_1 and/or T_2 [129]. If the T_1/T_2 ratio is greater than one standard deviation unit from the mean, then slow motions give rise to low T_2 values as a result of chemical exchange [129].

The program 'quadric_diffusion' computed the diffusion tensor for isotropic, axially symmetric and fully anisotropic models for the binary (DHFR:NADPH) and ternary (DHFR:NADPH:**1**) complexes (Table 3.2 and Table 3.3). Only residues with T_1/T_2 ratios to one standard deviation unit of the mean were used to compute the

diffusion tensor for each model. F-statistical testing showed the largest F value for the axially symmetric model, indicating that the relaxation data for the binary and ternary complexes is best described by the axially symmetric diffusion tensor. This is in agreement with other relaxation studies of DHFR that also describe DHFR by the axially symmetric diffusion tensor [122]. Upon comparing the binary and ternary complexes, there is little difference between the diffusion parameters. This indicates that binding of inhibitor **1** to DHFR does not affect the rotational diffusion tensor significantly.

Table 3.2: Isotropic, axial symmetric and anisotropic rotational diffusion tensor parameters for DHFR:NADPH complex.

Tensor	**D_{iso} (10^{-7} s^{-1})**	**$2D_{zz}(D_{zz}+D_{yy})$**	**D_{xx}/D_{yy}**	**θ (rad)**
Isotropic	1.638 ± 0.005	-----	----	----
Axial	1.640 ± 0.005	1.107 ± 0.019	----	0.344 ± 0.115
Anisotropy	1.640 ± 0.005	1.108 ± 0.019	0.986 ± 0.018	0.336 ± 0.118

Tensor	**Φ (rad)**	**Ψ (rad)**	**χ^2**	**F**
Isotropic	----	-----	233.74	----
Axial	3.487 ± 0.333	----	197.18	5.56
Anisotropy	(-)2.753 ± 0.433	1.511 ± 0.834	196.82	0.08

Table 3.3: Isotropic, axial symmetric and anisotropic rotational diffusion tensor parameters DHFR:NADPH:**1** complex.

Tensor	**D_{iso} (10^{-7} s^{-1})**	**$2D_{zz}(D_{zz}+D_{yy})$**	**D_{xx}/D_{yy}**	**θ (rad)**
Isotropic	1.540 ± 0.004	-----	----	----
Axial	1.539 ± 0.004	0.898 ± 0.014	----	1.551 ± 0.112
Anisotropy	1.538 ± 0.004	1.085 ± 0.015	0.913 ± 0.015	0.133 ± 0.360

Tensor	**Φ (rad)**	**Ψ (rad)**	**χ^2**	**F**
Isotropic	----	-----	240.99	----
Axial	6.265 ± 0.106	----	191.54	6.89
Anisotropy	1.605 ± 1.599	(-)1.551 ± 1.118	187.45	0.85

3.3.7 Determining ΔC_p of DHFR:NADPH:Inhibitor by ITC

The heat capacity of proteins plays a significant role in establishing the energetics of protein stabilization, including protein folding and molecular recognition [130]. The heat capacity change with binding at constant pressure, ΔC_p, can influence the change in enthalpy (ΔH) and entropy (ΔS), which in turn will affect the Gibbs free energy of stabilization parameter considerably [130]. The term C_p directly measures the magnitude of enthalpy changes of the system at a constant pressure. Studies indicate that the enthalpy fluctuations are caused mainly by obstructed internal rotations by functional groups such as methyl groups, high frequency stretching and bending modes, as well as low frequency conformational variations. In addition, non-covalent interactions and solvent accessibility of groups on the protein surface affect the heat capacity of proteins [130].

The goal of this experiment was to determine if the crystal structure of DHFR:NADPH in the ground state is in the same conformation when bound to the ligand compared to the structure in solution. Determining the change in heat capacity of binding of the complex can achieve this. In addition, the observed ΔC_p is compared with the theoretical ΔC_p.

To determine the heat capacity of DHFR when bound to inhibitor **1** and when bound to the established inhibitor MTX, the enthalpy of binding at various temperatures is measured. Initial ITC runs that were performed with 1 injection of 2 μL followed by 35 injections of 8 μL volumes resulted in isotherms with high C-

values, represented by sharp isotherm curves indicating fast ligand saturation that did not allow for an accurate enthalpy value to be acquired. Since the enthalpy value is simply determined by the y-intercept, it is possible to obtain an enthapy value at various temperatures by performing 3 injections of 8 μL volumes at each desired temperature and then calculating the average (Figure 3.23-3.24). All data is fit to a one-site binding model and the enthalpy values are obtained. By only allowing for 3 injections at each temperature and not saturating the binding site, a curve is not generated by the data points, but rather a straight line is obtained from which the slope of the line can be extrapolated to the y-intercept to determine the associated enthalpy value. Enthalpy values were determined at a temperature range of 10°C to 30°C.

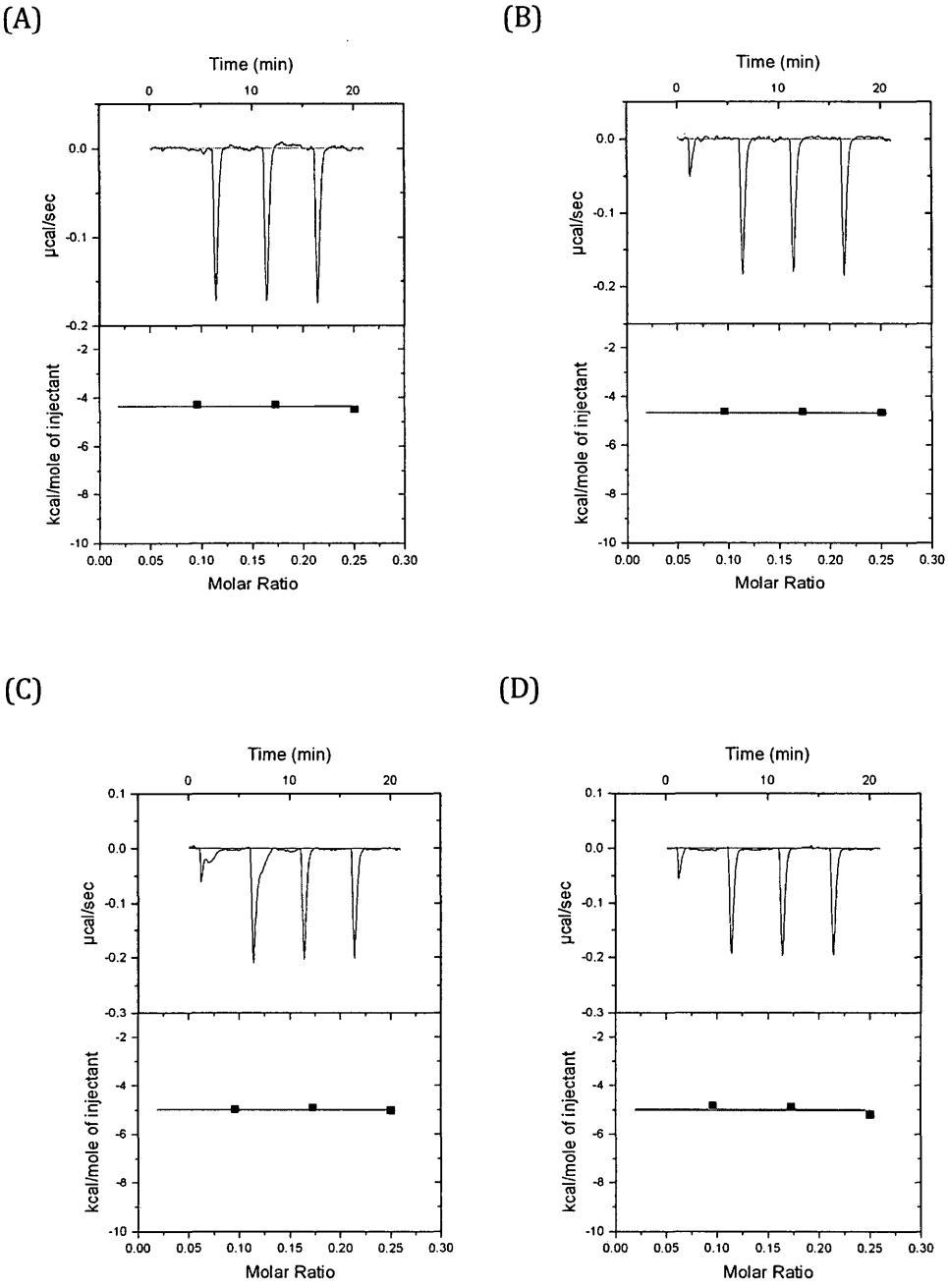
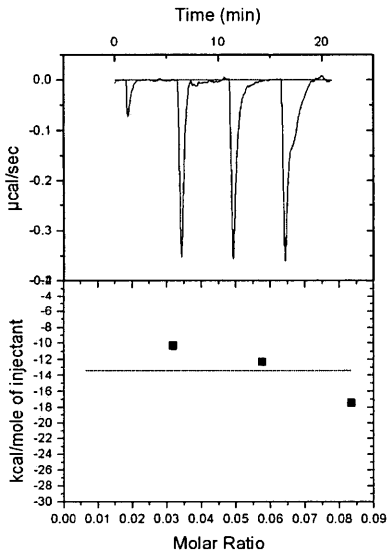
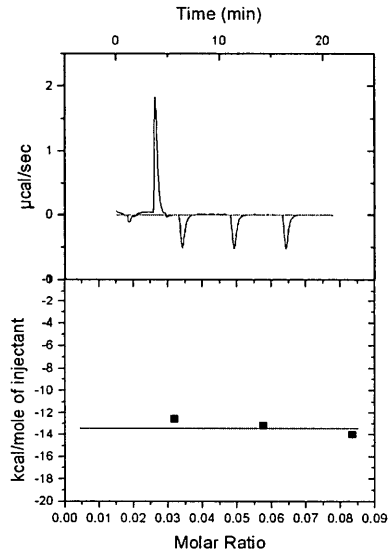


Figure 3.23: The resulting ITC chromatograms (top panels) and binding isotherm curves (bottom panels) of titrations of DHFR:NADPH to **1** taken over 3 injections of 8 µL. Titrations were performed at temperatures of (A) 10°C, (B) 15°C, (C) 20°C, and (D) 25°C.

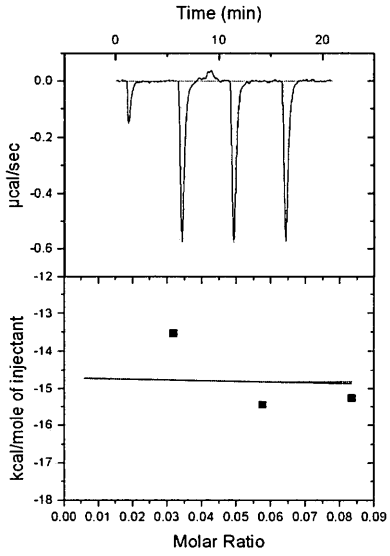
(A)


Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio

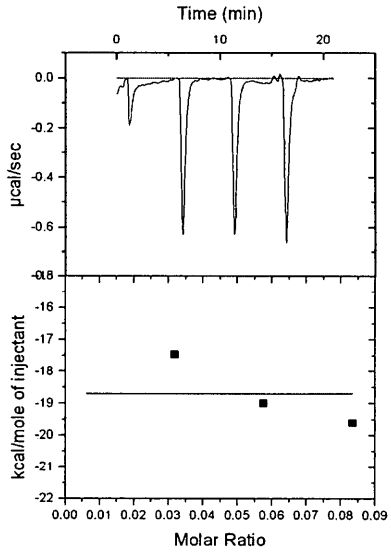
(B)


Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio

(C)


Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio

(D)


Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio

(E)

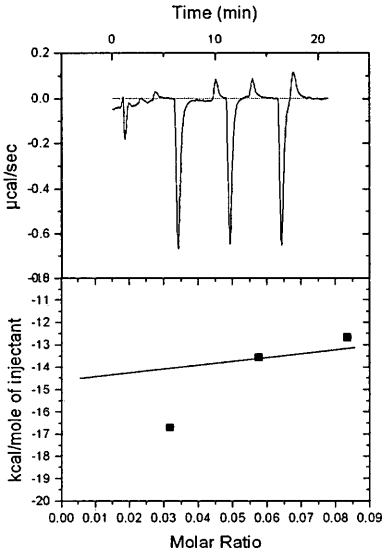
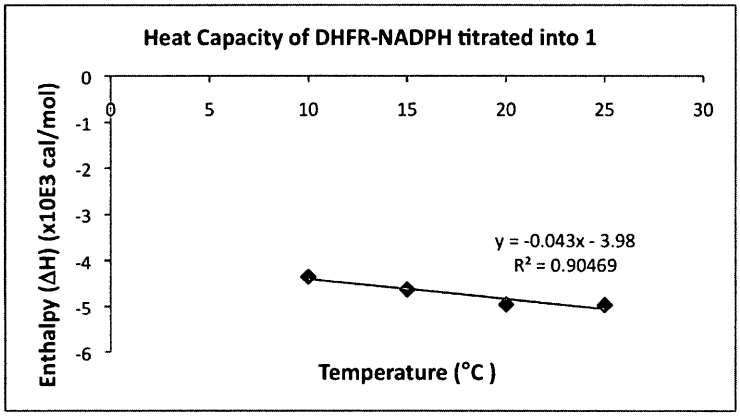


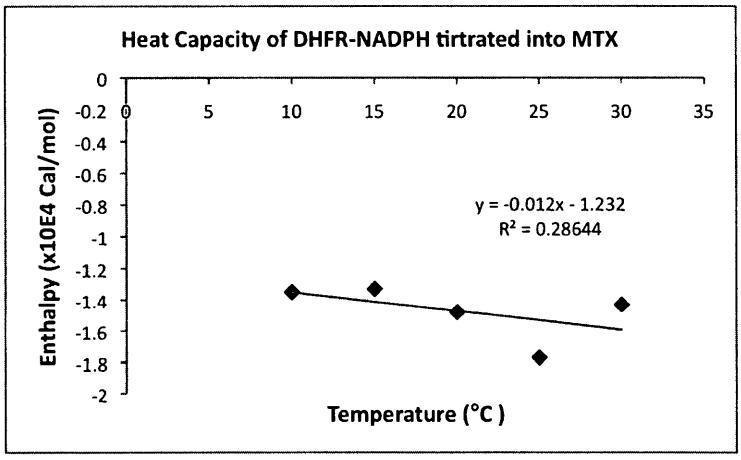
Figure 3.24: The resulting ITC chromatograms (top panels) and binding isotherm curves (bottom panels) of titrations of DHFR:NADPH to MTX taken over 3 injections of 8 µL. Titrations were performed at temperatures of (A) 10°C, (B) 15°C, (C) 20°C, (D) 25°C, and (E) 30°C.

Plotting the acquired enthalpy values against temperature resulted in a negative straight line that is indicative of an exothermic binding reaction of DHFR:NADPH binding to **1** that is enthalpically favored. Similarly, DHFR:NADPH binding to MTX is also found to be an exothermic process that is enthalpically driven. Heat capacity can be determined using the equation $C=\Delta H/\Delta T$ where C is the ratio of heat energy that is transferred in the reaction that causes the resulting increase in temperature. Therefore, by determining the slope of the line of the enthalpy-temperature plot, the heat capacity of DHFR:NADPH:**1** is found to be -43 ± 10 cal/mol·K (Figure 3.25) and the heat capacity of DHFR:NADPH:MTX is found to be -120 ± 109 cal/mol·K (Figure 3.26).



T (°C)	ΔH (cal mol^{-1})
10	(-4.36 ± 0.01) x 10^3
15	(-4.64 ± 0.01) x 10^3
20	(-4.96 ± 0.01) x 10^3
25	(-4.97 ± 0.01) x 10^3

Figure 3.25: Enthalpy-temperature plot displaying the heat capacity of DHFR:NADPH:**1**.



T (°C)	ΔH (cal mol^{-1})
10	(-1.35 ± 0.01) x 10^4
15	(-1.33 ± 0.01) x 10^4
20	(-1.48 ± 0.01) x 10^4
25	(-1.77 ± 0.01) x 10^4
30	(-1.43 ± 0.01) x 10^4

Figure 3.26: Enthalpy-temperature plot displaying the heat capacity of DHFR:NADPH:MTX.

Since noncovalent interactions contribute positively to heat capacity of the native protein, disrupting these interactions upon unfolding will contribute negatively to ΔC_p with a large negative ΔC_p that is characteristic of site-specific binding [131]. These results suggest that at higher temperatures the DHFR:NADPH complex is unfolding as the noncovalent interactions are disrupted and the protein surface area is exposed to hydration effects.

It has been recognized that heat capacity is influenced by the hydration of polar and apolar groups of the protein, where the changes in solvent accessibility is associated with protein folding and unfolding events [130]. Therefore, considering the role of the solvent is imperative for understanding protein energetics. To

determine the Polar Surface Area (PSA) and Apolar Surface Area (ASA) of the biomolecule in the free and bound forms, the computer software 'Vega' [132] was used to calculate the heat capacity of DHFR:NADPH when free and when bound to inhibitor. To do this, the crystal structure of DHFR:NADPH and the inhibitor involved in the binding reaction must be known. The PSA of a molecule is the surface sum of all the polar atoms in the biomolecule, which are usually oxygen and nitrogen atoms. Conversely, the ASA of the molecule is the surface sum of all the apolar atoms in the biomolecule [131]. This calculation is most often used to determine how well a molecule can permeate cells [133], however for the purpose of this experiment, Vega calculations are used to determine the heat capacity of the complex as a crystal structure in the ground state when bound to ligand. These heat capacity values can then be compared to heat capacity values of the complex in the solution state obtained by ITC. Using crystal structures from the Protein Data Bank (PDB), surface area values for DHFR, DHFR:NADPH, and DHFR:NADH:**1** were generated (Table 3.4). For comparison purposes, the program Marvin Draw [134] was also used to obtain surface area values for the inhibitors **1** an MTX in the solution.

Table 3.4: PSA, APA, and total surface area values of DHFR and DHFR complexes using Vega software [132] and Marvin Draw software [134].

Molecule	PDB ID	Polar Surface Area (PSA), (Å^2)	Apolar Surface Area (ASA), (Å^2)	Total Surface Area (Å^2)
DHFR	2ANQ	4570.9	3977.0	8547.9
DHFR	1RH3	4674.7	3927.0	8601.7
DHFR:NADPH	1RX1	4607.2	3703.4	8310.6
DHFR:NADPH:**1**	2ANQ	4419.5	3836.4	8255.9
1 (CRYSTAL)	2ANQ	340.1	247.5	587.6
1 (SOLUTION)	n/a*	372.4	257.9	630.3
DHFR:NADPH:MTX	1RH3	4499.2	3521.5	8020.7
MTX (CRYSTAL)	1RH3	366.8	310.1	676.9
MTX (SOLUTION)	n/a*	392.4	284.0	676.4

The change in heat capacity, ΔC_p, upon protein folding is dependent on the water accessible polar and nonpolar surface areas as given by Eq. (12) [131]

$$\Delta C_p = (0.32 \pm 0.04 \times \Delta\text{nonpolar}) - (0.14 \pm 0.04 \times \Delta\text{polar}) \qquad (12)$$

where the units for the constants 0.32 and 0.14 are cal/mol·K·Å^2.

Using the PSA and APA values generated in Table 3.4, the contribution of Δpolar and Δnonpolar for DHFR:NADPH binding to ligand **1** can be calculated:

$\Delta polar_{(ligand\ 1)}$ = DHFR:NADPH + **1** $\rightleftharpoons$ DHFR:NADPH:**1**

= 4607.2 Å^2 + 340.1 Å^2 $\rightleftharpoons$ 4419.5 Å^2

= -527.8 Å^2

$\Delta nonpolar_{(ligand\ 1)}$ = DHFR:NADPH + **1** $\rightleftharpoons$ DHFR:NADPH:**1**

= 3703.4 Å^2 + 247.5 Å^2 $\rightleftharpoons$ 3836.4 Å^2

= -114.5 Å^2

Substituting the values for Δpolar and Δnonpolar into Eq. (12), the predicted change in heat capacity of DHFR:NADPH:**1** is determined to be 37.252 ± 25.692 cal/mol•K.

Similarly, the PSA and APA values generated in Table 3.4 is used to calculate the contribution of Δpolar and Δnonpolar for DHFR:NADPH binding to ligand MTX:

$\Delta polar_{(ligand\ MTX)}$ = DHFR:NADPH + MTX $\rightleftharpoons$ DHFR:NADPH:MTX

= 4607.2 Å^2 + 366.8 Å^2 $\rightleftharpoons$ 4499.2 Å^2

= -474.8 Å^2

$\Delta nonpolar_{(ligand\ MTX)}$ = DHFR:NADPH + MTX $\rightleftharpoons$ DHFR:NADPH:MTX

= 3703.4 Å^2 + 310.1 Å^2 $\rightleftharpoons$ 3521.5 Å^2

= -492.0 Å^2

Substituting the values for Δpolar and Δnonpolar into Eq. (12), the predicted change in heat capacity of DHFR:NADPH:MTX is determined to be -90.968 ± 38.672 cal/mol•K.

The small difference seen between the observed and theoretical heat capacity values of DHFR binding to MTX suggests that the binding mechanism in solution is similar to the binding mechanism that yields the ground state conformation observed in the crystal structure of the complex. The binding of MTX to *E. coli* DHFR is known to be a complex process that involves multiple conformational states. The crystal structure of DHFR bound to MTX shows that MTX sits in the active site of the enzyme and the Met20 loop closes over the MTX molecule (Figure 3.27) [117, 126]. In this ground state conformation the Met20 loop is flexible since MTX does not directly interact or bind to the Met20 loop, suggesting that any conformational changes involves an opening and closing of the loop over MTX.

Unlike MTX and other DHFR inhibitors known to date, inhibitor **1** interacts with the Met20 loop (Figure 3.28). This interaction is particularly interesting because the Met20 loop has an important role in the structural changes that occur in

the binding mechanism [116]. The observed heat capacity of DHFR:NADPH binding to **1** is significantly different from the theoretical heat capacity, indicating that the mode of binding may be considerably different in solution compared to that achieved for the conformation of the complex in the crystal structure. The structure of the DHFR:NADPH complex represented in the X-ray structure (1RX1) may not be the structure that binds inhibitor **1**. Rather, inhibitor 1 may be bound by a different structure that did not crystallize or is only transiently populated. Since the inhibitor makes contacts with the Met20 loop, this loop is held in place and becomes less flexible and more rigid in structure. This stability contributes positively to heat capacity when the complex is in the crystal structure. However, in solution, disruption of these interactions and changes to the how the contacts are made or how many contacts are made will affect the magnitude of enthalpy changes of the system and contribute negatively to the heat capacity of the complex formation such that the heat capacity has a more negative value.

Figure 3.27: Crystal structure illustrating the Met20 loop (pink) of *E. coli* DHFR closing over inhibitor MTX (orange) (PDB ID: 1RH3) [117].

Figure 3.28: Crystal structure illustrating the Met20 loop (pink) of *E. coli* DHFR interacting with inhibitor **1** (blue) (PDB ID: 2ANQ) [126].

3.4 CONCLUSIONS

DHFR is dynamic in nature and conformational changes are necessary for substrate binding and product release to be executed efficiently. In order to determine how inhibitor binding affects the catalytically important motions of *E. coli* DHFR, ^{15}N relaxation parameters were measured upon binding of inhibitor **1** by employing well established NMR relaxation methods that use T_1, T_2, ^1H-^{15}N NOE, and CPMG pulse sequences. In the presence of inhibitor **1**, the catalytic domain of DHFR binds **1** in the substrate-binding pocket and an occluded conformation is assumed as residues in and around the Met20 loop exhibit significant motions. In both the DHFR:NADPH and DHFR:NADPH:**1** complexes mobility is observed in the Met20, F-G and G-H loops as well as the adenosine binding loop (residues 67-69) as reflected by low ^1H-^{15}N-NOE values, non-average T_1/T_2 values and significant τ_m values. These results are in good agreement with the general dynamic features of DHFR presented in studies from the Wright laboratory that analysed the dynamics of DHFR:folate, DHFR:DHNADPH:folate and DHFR:$NADP^+$:folate complexes [121, 122]. The Wright laboratory reported that significant motions were found to occur in the occluded complexes (DHFR:folate and DHFR:DHNADPH:folate), whereas in the closed complex (DHFR:$NADP^+$:folate) motion in the Met20 and F-G loops is reduced. The relaxation data presented here in this dissertation show that the catalytically important Met20 and F-G loops in both DHFR:NADPH and DHFR:NADPH:**1** complexes most closely resembles that of the closed DHFR complex.

These findings are consistent with the analysis of the X-ray structure [126]. DHFR exhibits motions on the microsecond-millisecond timescale that is indicative to motions that can be attributed to either conformational exchange within the protein and biochemical events such as ligand binding. Future studies include determining if this effect on DHFR dynamics is seen with other DHFR inhibitors. It is possible that compounds that favor the formation of the closed conformation could lead to more efficient DHFR inhibitors.

The estimated τ_m for both binary and ternary complexes was in the nanosecond timescale. Although the relaxation data could not be fit by the FAST-ModelFree automated model-fitting program, information about the diffusion tensor dynamics of the binary and ternary complexes were determined. Relaxation data for the binary and ternary complexes is best described by the axially symmetric diffusion tensor, which is consistent with previous relaxation studies of DHFR [121, 122]. Upon comparing the binary and ternary complexes, there is little difference between the diffusion parameters, indicating that binding of inhibitor **1** to DHFR does not significantly affect the rotational diffusion tensor. An observed increase in τ_m is indicative of a restriction of rotational freedom of the entire molecule upon formation of the ternary complex, which is consistent with a large molecular weight complex.

Heat capacity changes of DHFR:NADPH upon binding to the inhibitor MTX provided an understanding of the energetic contributions to the stability of the protein-inhibitor complexes of DHFR that is currently not known. Inhibitor binding

with DHFR is shown to be an exothermic process as protein folding and structural stability is increased and protein flexibility is reduced. Since the MTX inhibitor does not interact with the Met20 loop, the loop is flexible and the binding mechanism in solution is similar to that observed in the crystal state where the presence of the inhibitor demonstrates positive binding cooperativity and the affinity of DHFR for NADPH is enhanced upon inhibitor binding [126, 135]. Conversely, inhibitor **1** does interact with the Met20 loop. Inhibitor **1** in the presence of the cofactor also leads to positive binding cooperativity and the affinity of DHFR:NADPH for **1** is comparable to inhibitors of DHFR such as MTX whose Kd ranges from 10^{-8}M to less than 10^{-11}M depending on the presence of cofactors and the source of DHFR [135-138]. However, differences in ΔC_p of DHFR binding to inhibitor **1** compared to that of MTX indicate that the mode of binding to **1** is different from what is observed in the crystal structure of the complex. It is likely that the structure that binds to **1** did not crystallize or that the structure is transiently populated.

CHAPTER 4

THE DEVELOPMENT OF AN AFFINITY ELECTROPHORESIS METHOD FOR SCREENING RNA-SMALL MOLECULE INTERACTIONS

PREFACE

This chapter presents work from two separate papers - one published and one accepted at the time of writing this dissertation. The published paper illustrates that the affinity electrophoresis method [139] *described in this chapter works successfully. The work outlined in the accepted paper demonstrates how the affinity electrophoresis method previously described can be made quantifiable by cross-linking the ligand to the gel matrix.*

4.0 AMINOGLYCOSIDE ANTIBIOTICS

The ability of RNA to interact with small molecules is the basis for a wide assortment of biochemical functions. In nature, RNA-small molecule interactions in riboswitches regulate gene expression [140, 141] while in biotechnology applications, RNA aptamers can be selected as biosensors for small molecules [142]. In addition, the interactions between RNA and small molecules govern the action of a number medicinally important molecules such as the aminoglycoside class of antibiotics (Figure 4.1) [143, 144]. The chemical and structural diversity of antibiotics makes these molecules particularly interesting for understanding the binding specificity of RNA for its target ligands [24, 145-147]. Aside from its interactions with small molecules, it is clear that RNA has many other important functions in cellular processes [35, 39, 148]. Controlling the biological function of RNA through the binding of a small molecule would be a powerful biochemical tool. For this reason, developing a method of quickly and easily identifying compounds that bind to RNA of a defined size or shape is very beneficial.

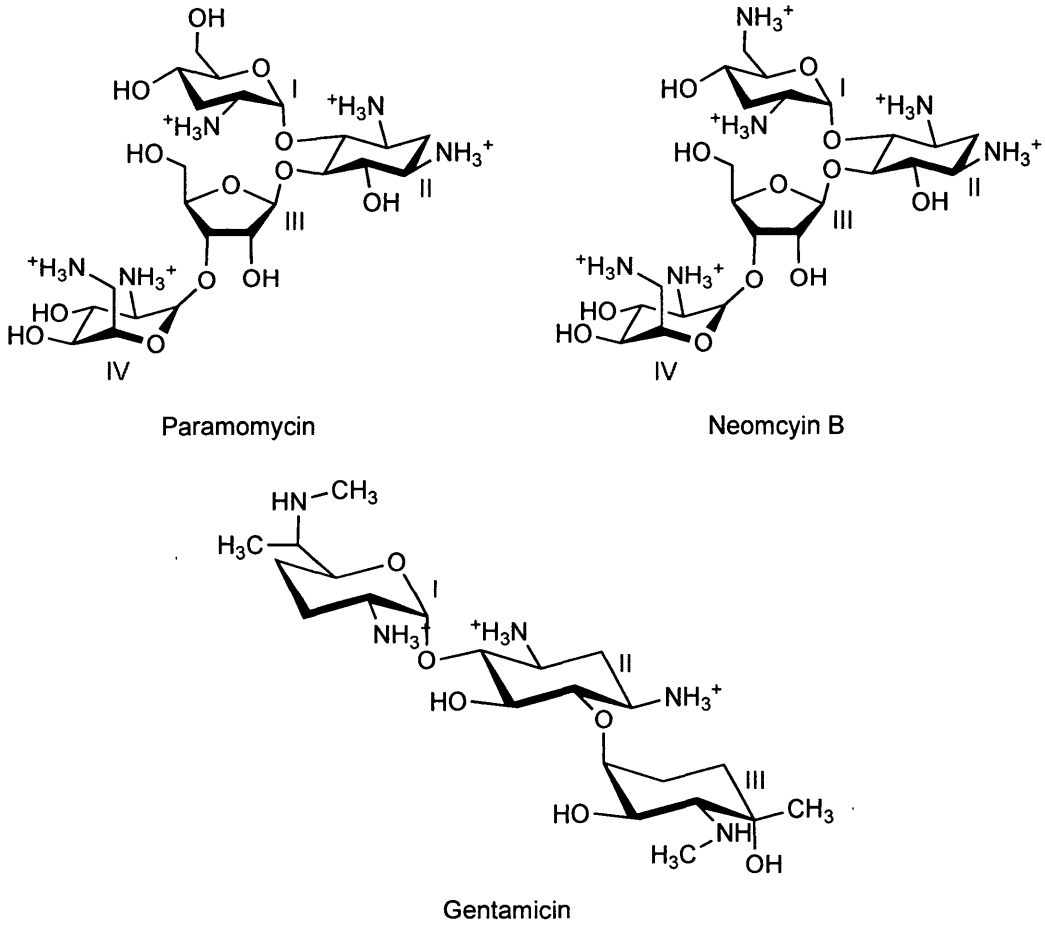


Figure 4.1: Chemical structures of the aminoglycoside antibiotics used in this affinity electrophoresis study.

There are many biophysical methods available to study RNA-small molecule interactions including isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence anisotropy. However, many of these techniques require specialized and expensive equipment. While the interaction of RNA with

proteins is often readily investigated using a simple electrophoretic mobility shift assay (EMSA) or gel shift assay [149, 150], this technique is limited when applied to RNA-small molecule systems because of the negligible difference in mobility between free and bound RNA. An alternate gel-based method for determining binding affinity is affinity electrophoresis. In this method, the gel, usually acrylamide, is polymerized in the presence of a ligand so as to embed the ligand into the gel matrix. A biomolecule is then run on the gel, and binding is gauged by the difference in mobility between gels run with and without the ligand or between a lane containing a nonbinding standard and the biomolecule being studied. This technique has been particularly widely used in the area of protein-carbohydrate interactions to identify which carbohydrates are bound by a particular protein and also to quantify binding [151-154]. For RNA-peptide interactions, a modified affinity electrophoresis method, polyacrylamide gel coelectrophoresis (PACE), was developed [155]. In the PACE method, a gel containing bands of different concentrations of peptide is prepared, and then the RNA is run through these different bands. Binding is assayed by the decrease in mobility of the RNA in bands of higher concentration of peptide. Although several examples have been published that detects the binding of molecules using affinity electrophoresis, in each case this technique has been limited to interactions involving proteins or peptides with another molecule [91, 153-156].

4.1 RESEARCH GOALS

This project aimed to show that affinity electrophoresis could be used to detect RNA-small molecule interactions [139]. In this method, the acrylamide gel is polymerized in the absence and presence of ligand such that the ligand is embedded into the gel matrix during polymerization. The RNA molecule is then run through the gel and binding is gauged by the difference in mobility of the bands between the gels polymerized with and without ligand. To demonstrate the validity of this method, the ribosomal A-site aminoglycoside interaction is used as the model system (Figure 4.2). In addition, this method is used to determine the range of aminogylcoside molecules that can be bound by the CC mismatch-containing RNA molecule from the thymidylate synthase mRNA construct (TSMC) (Figure 4.2). To achieve these objectives, the A-site RNA and TSMC RNA was used to guage the affinity to aminoglycoside ligands paromomycin, neomycin B, and gentamicin. As a control, mobility is also compared with a non-binding internal standard. While this method can identify interactions between small molecules and RNA, the binding is not quantifiable as the small molecule ligand can move in the gel during electrophoresis. For the ligand to be stationary during electrophoresis it must either be uncharged or too large to move through the gel. Aminoglycosides are neither.

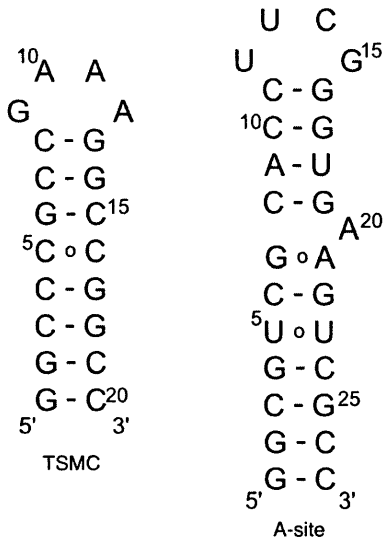


Figure 4.2: Sequence and secondary structures of the two RNA molecules used in this study. TSMC is the thymidylate synthase mRNA construct containing the CC mismatch and A-site is the *E. coli* ribosomal A-site.

To make this method quantifiable, the affinity electrophoresis method developed was modified to make the technique quantifiable for RNA-small molecule interactions. To achieve this, a chemical modification of the aminoglycoside was made by attaching an acrylamide monomer onto the structure. This modified aminoglycoside was then added to the acrylamide mixture during polymerization with the results so that the ligand becomes covalently attached to the gel, and cannot move during electrophoresis (Figure 4.3). By measuring the relative

migration values (R_f) and plotting $1/R_f$ versus ligand concentration, a straight line is obtained from which an apparent K_d can be determined.

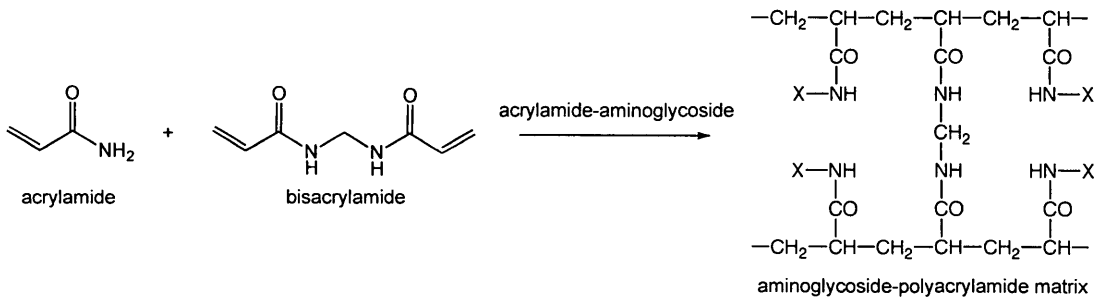


Figure 4.3: Diagram of our scheme for generating a polyacrylamide gel cross-linked with an aminoglycoside. X in the diagram is an aminoglycoside molecule.

4.2 MATERIALS AND METHODS

4.2.1 RNA Preparation

The TSMC RNA was prepared by *in vitro* transcription [157] in a manner previously described [139, 158]. The A-site RNA was purchased from the University of Calgary DNA Services and the same sequence used previously for nuclear magnetic resonance (NMR) structural studies by the Puglisi group [159] and by us, the Johnson group, for studying the binding of new non-aminoglycoside A-site binding compounds [160].

4.2.2 Synthesis of Cross-Linked Aminoglycoside-Acrylamide

In a flame-dried vial containing a stirrer, aminoglycoside sulfate salt (0.5 mmol, 1.0 equiv.) was added before it was sealed with a septum. The vial was purged under high vacuum and backfilled with argon (g) three times before a ddH_2O solution (totaling 1 mL) containing Na_2CO_3 (46 mg, 0.55 mmol, 1.1 equiv.) was added via syringe. The resultant yellow suspension was sonicated for 5 min until the mixture was solubilised completely. To the clear, yellow solution was added acryloyl chloride (50 μL, 0.55 mmol, 1.1 equiv.) in one portion *via* microlitre syringe. The resultant solution was left to stir at room temperature for 24 h and was carried forward for polymerization reactions without further purification.

4.2.3 Affinity Gel Electrophoresis Preparation

Method 1 – Detection of RNA-Small Molecule Interactions

Polyacrylamide non-denaturing electrophoresis gels of the A-site RNA and TSMC RNA were run in the absence and presence of varying amounts of the aminoglycosides paromomycin, gentamicin, and neomycin B (Sigma). Gels were prepared using the Bio-Rad Mini-PROTEAN 3 apparatus using a 10 cm x 7.5 cm x 0.75 mm gel. All gels were run in a 0.5x Tris/Borate/EDTA (TBE) buffer (1 x TBE: 9 mM TRIS, 9 mM boric acid, 0.3 mM EDTA) at room temperature for 30 minutes at 300 V using the constant current setting. Bromophenol blue/xylene cyanol (blue/green dye) was run as an internal standard. RNA was detected using Stains-All (Sigma). The concentration of RNA used was 0.13 mM for the TSMC RNA and 0.14 mM for the A-site RNA.

Method 2 – Quantification of RNA-Small Molecule Interactions

For both the cross-linked and embedded gel preparation, 20% polyacrylamide (19:1 acrylamide:bisacrylamide) non-denaturing electrophoresis gels of the TSMC RNA were run along with a control gel containing varying amounts of the appropriate aminoglycosides paromomycin and neomycin B (Sigma-Aldrich). Gels were prepared using the Bio-Rad Mini-PROTEAN 3 apparatus using a 10 cm x 7.5 cm x 0.75 mm gel. All gels were run in a 0.5x Tris/Borate/EDTA (TBE) buffer (1 x TBE: 9 mM TRIS, 9 mM boric acid, 0.3 mM EDTA) at room temperature for 30

minutes at 300 V using the constant current setting. Bromophenol blue/xylene cyanol (blue/green dye) was run as an internal standard. RNA was detected using Stains-All (Sigma-Aldrich). In each lane, 1 μL of an RNA solution was loaded. The RNA solution was composed of a 1:1 mixture of 0.13 mM TSMC RNA:50% glycerol (v/v).

4.2.4 Affinity Gel Electrophoresis Analysis

The affinity electrophoresis gels were analysed according to the method of Takeo [91] in order to quantify the average apparent binding affinity. This method is summarized here. For the equilibrium between a macromolecule (M) and a ligand (L):

$$M + L \leftrightarrows ML \tag{13}$$

The apparent dissociation constant is given by Eq. (14):

$$K_d = \frac{[M][L]}{[ML]} \tag{14}$$

If the distance migrated by the macromolecule in the absence of ligand is given by R_o and the distance migrated in the presence of a particular concentration of ligand is r, then:

$$\frac{R_o}{r} = \frac{[M]_t}{[M]} = \frac{[ML]+[M]}{[M]} = \frac{[ML]}{[M]} + 1 \tag{15}$$

where $[M]_t$ is the total concentration of the macromolecule. If Eq. (14) is rearranged and substituted into Eq. (15), then:

$$\frac{R_o}{r} = \frac{[L]}{K_d} + 1 \qquad (16)$$

In our graphs we define the relative migration R_f as:

$$R_f = \frac{r}{R_o}$$

Therefore Eq. (16) becomes:

$$\frac{1}{R_f} = \frac{[L]}{K_d} + 1$$

A plot of $\frac{1}{R_f}$ versus [L] will give a straight line with a slope of $\frac{1}{K_d}$ under conditions where the total concentration of the ligand (L) is much greater than the total concentration of macromolecule.

4.3 RESULTS

4.3.1 Detection of RNA-Small Molecule Interactions

Affinity electrophoresis gels were used to show binding between aminoglycosides and the A-site RNA and the TSMC RNA (Figure 4.4). Both the A-site RNA and TSMC RNA samples were run in a polyacrylamide gel prepared in the absence of aminoglycosides and gels polymerized in the presence of 0.5% (w/v) of the aminoglycosides paromomycin, gentamicin, and neomycin B. Clearly, in the gels containing aminoglycosides, the mobility of the RNA is greatly reduced compared with the gel run in the absence of ligand. This decreased migration indicated that binding is occurring in the gel between the RNA and the aminoglycoside, and this binding reduces the mobility of the RNA in the gel.

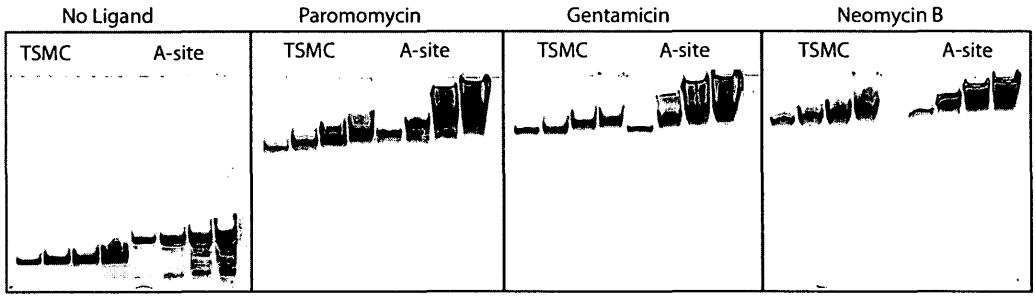


Figure 4.4: Affinity electrophoresis demonstrating the interaction of the TSMC RNA and A-site RNA molecules with aminoglycoside ligands. Shown are nondenaturing 20% polyacrylamide gels containing no ligand and 0.5% (w/v) paromomycin, gentamicin, and neomycin B. Four different volumes (1, 2, 4, and 6 μL) of each RNA (0.13-0.14 mM) are loaded. All four gels were run in parallel at room temperature for the same total time.

As a control, gels were also run where 0.2% (w/v) maltotriose and sucrose were present in the gel and compared these gels with ones where no ligand and 0.2% (w/v) paromomycin were added (Figure 4.5). Maltotriose and sucrose were chosen as controls because they, like paromomycin, are oligosaccharides. The TSMC RNA in the gels with maltotriose and sucrose ran a similar distance as in the gel with no ligand, whereas the mobility of the RNA was greatly reduced in the gel with paromomycin (Figure 4.5).

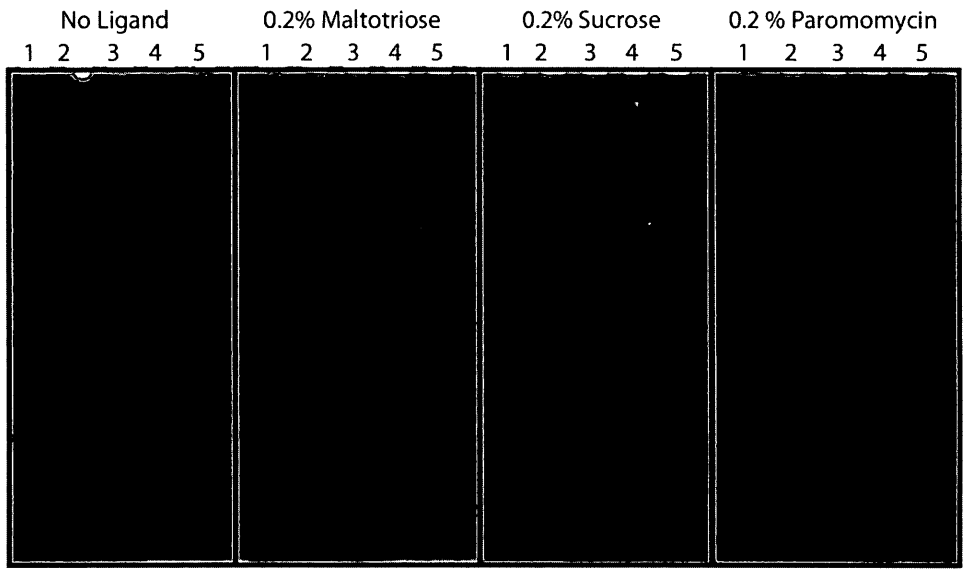


Figure 4.5: Native gel electrophoresis of TSMC RNA in the absence of ligand added to the gel and in the presence of 0.2% maltotriose, 0.2% sucrose, and 0.2% paromomycin. Shown in lane 1 is the bromophenol blue dye marker, and lanes 2, 3, 4, and 5 contain 1, 2, 4, and 6 μL of 0.13 mM TSMC RNA, respectively. The gel is 20% acrylamide. Relative to the bromophenol blue dye, the gels with sucrose and maltotriose show no retardation of the RNA. A decrease in mobility is seen in the gel with paromomycin.

When performing affinity electrophoresis, a number of experimental factors needed to be optimized to obtain the best results. As shown in Figure 4.4, the amount of material loaded into the gel is one of the factors that needed to be

optimized. Enough RNA needed to be loaded to ensure detection of the complex in the gel, whereas too much RNA leads to increased band size caused by streaking or smearing of the band and decrease in migration of the RNA. For the two RNA samples analyzed here, the best results were obtained by loading 0.13 nmol (1 μL of a 0.13 mM RNA solution) of RNA. An additional factor we tested was the acrylamide concentration in the gel. Affinity electrophoresis in gels were performed with both 12% and 20% acrylamide (data not shown). In all cases, a clearer distinction between bound and free RNA was obtained with 20% gels, although the time need to run the gel increased at higher acrylamide concentrations.

Both the A-site RNA and TSMC RNA display different affinities for the different aminoglycoside ligands. This different affinity is reflected in a different mobility of these RNA molecules in the gels containing different aminoglycosides. Figure 4.6 shows gels of the A-site RNA and TSMC RNA molecules run in the presence of 0.2% (w/v) paromomycin, neomycin B, and gentamicin. The affinity of these two RNA molecules for these ligands has been studied in solution with both the A-site RNA and TSMC RNA binding neomycin B with higher affinity than paromomycin [144], [161], [162, 163]. This known difference in solution binding affinity is reflected in the affinity electrophoresis technique. In the gel containing neomycin, the RNA migrates a significantly shorter distance than in the gel containing paromomycin (Figure 4.6). For the A-site RNA, there is a previous report that gentamicin has a slightly lower IC_{50} value than paromomycin [164]. This likely reflects a higher affinity for gentamicin compared with paromomycin. Again these

affinity electrophoresis results are consistent with the known solution affinity, with both RNA molecules tested migrating slower in the presence of gentamicin than with paromomycin in the gel. Although the affinity of the TSMC RNA for gentamicin has not been reported previously, from the slower migration of TSMC in the presence of gentamicin compared with paromomycin, it is predicted that the TSMC binds paromomycin weaker than gentamicin and that TSMC binds neomycin B the tightest among these three ligands.

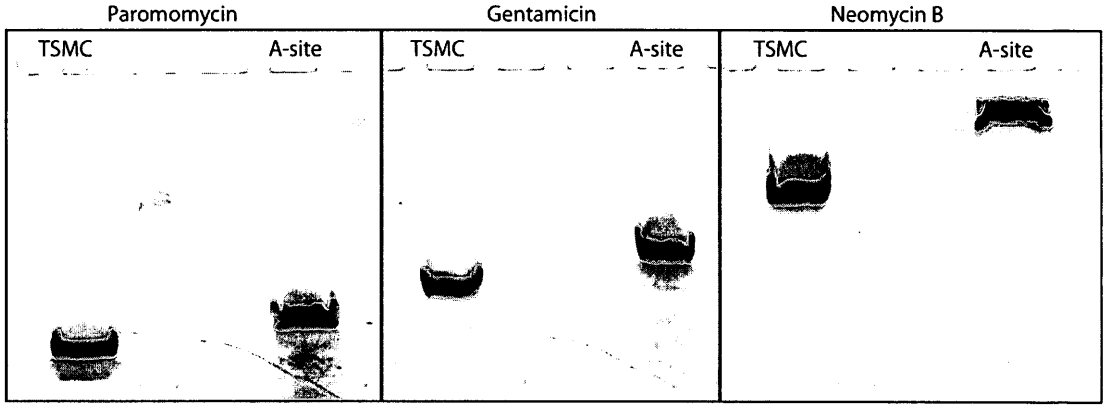


Figure 4.6: Affinity electrophoresis demonstrating the difference in interaction of the TSMC RNA and A-site RNA molecules with aminoglycoside ligands. Shown are nondenaturing 20% polyacrylamide gels containing 0.2% (w/v) paromomycin, gentamicin, and neomycin B. All gels were run in parallel for the same time at room temperature. To best differentiate between the different ligand/RNA combinations, the RNA in the gel containing no ligand was allowed to run out of the gel; therefore, it is not shown.

The affinity electrophoresis screen presented here displays dose-dependent binding between the ligand and RNA. The ligand concentration embedded in the gel was varied and the relative migration of the same amount of TSMC RNA in the different gels was measured. Figure 4.7 shows that in the presence of a lower concentration of ligand, the RNA molecules migrate faster than at higher ligand concentrations.

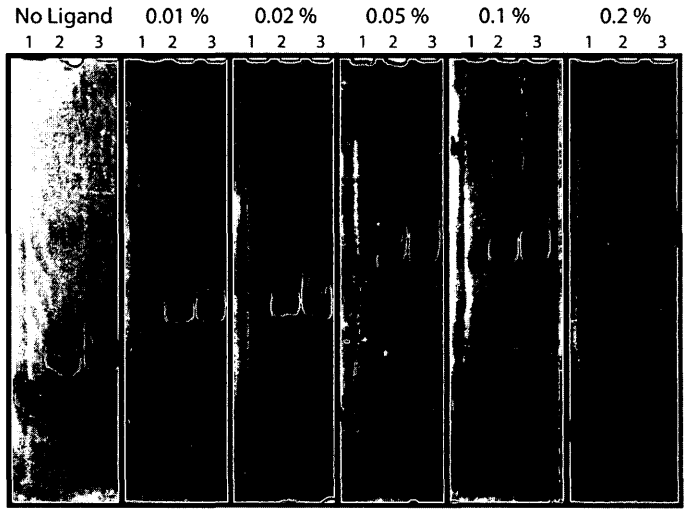


Figure 4.7: Affinity electrophoresis demonstrating the difference in mobility of the TSMC RNA with different concentrations of ligand in the gel. Shown are 20% polyacrylamide gels containing 0, 0.01, 0.02, 0.05, 0.1 and 0.2% (w/v) paromomycin. Shown in lane 1 is the bromophenol blue dye marker, and lanes 2 and 3 contain 1 and 2 μL of 0.13 mM TSMC RNA respectively. All gels were run at room temperature.

As an internal standard the migration of the RNA compared with the bromophenol blue dye was measured. As seen in Figure 4.8 (a), if the relative migration is plotted against paromomycin concentration, the distance the TSMC moves plateaus at the highest concentrations of paromomycin.

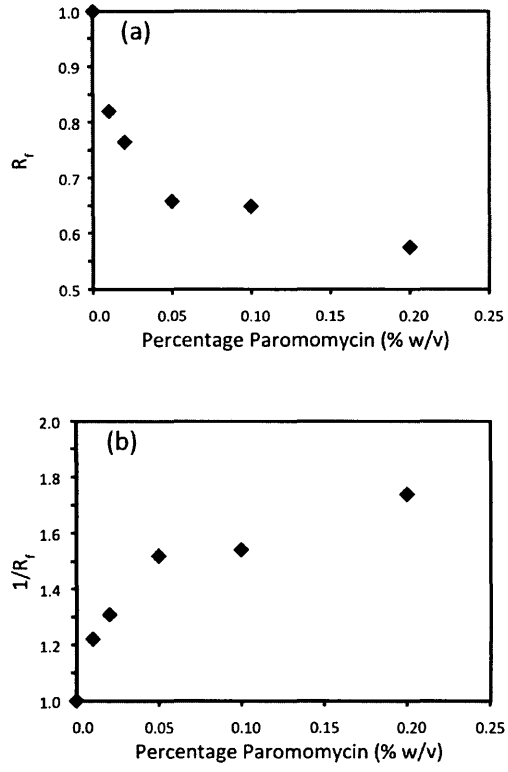


Figure 4.8: Affinity plots for the interaction of the CC mismatch-containing RNA hairpin from the TSMC with the aminoglycoside paromomycin. (A) The relative migration (R_f) of the TSMC RNA is plotted against the amount of paromomycin present in the gel. (B) The reciprocal of the relative migration ($1/R_f$) of the TSMC RNA is plotted against the amount of paromomycin.

4.3.2 Quantification of RNA-Small Molecule Interactions

In order to prevent the aminoglycoside ligand from moving in the gel during electrophoresis, paromomycin and neomycin B aminoglycosides were reacted with acryloyl chloride to form an acrylamide-aminoglycoside (Figure 4.9). Formation of the acrylamide-aminoglycoside is straightforward. With the conditions used, there is on average one acrylamide added per aminoglycoside molecule. However, there is no control over the location of the acrylamide monomer and the acryloyl chloride could react with any of the amine or alcohol groups on the aminoglycoside. A mixture of acrylamide-aminoglycoside products that differ by the location of the acrylamide monomer are produced.

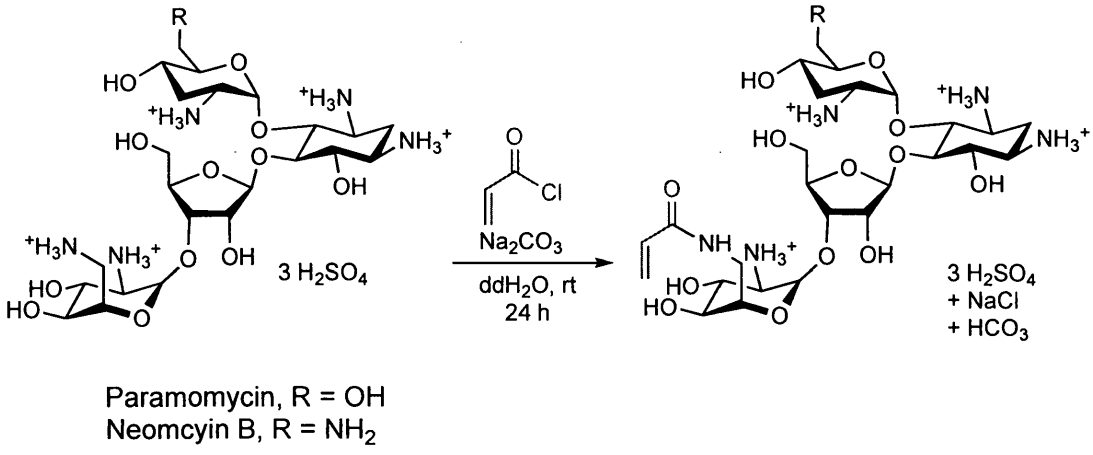


Figure 4.9: Reaction scheme for cross-linked acrylamide-paromomycin and acrylamide-neomycin B, when R=OH and R=NH_3^+ respectively. One product is shown from a potential range of cross-linked acrylamide products.

Confirmation of acrylamide-aminoglycoside synthesis was made by infrared (IR) and high-resolution mass spectrometry (HRMS). ^1H and ^{13}C NMR was also performed on the products in D_2O, however, these results were inconclusive due to extensive peak overlap. IR spectra indicate a loss of the O-H stretch and N-H stretch for both cross-linked acrylamide-aminoglycosides compared to the free aminoglycosides. This loss of the O-H stretch and N-H stretch suggests that the acryloyl chloride is reacting with both the alcohol and amine groups on the aminoglycoside. HRMS provided exact molecular weights of the expected acrylamide-aminoglycoside products. For the cross-linked paromomycin, the determined value of 638.3015 m/z matched the theoretical HRMS value of 638.3010

m/z. Similarly, the cross-linked neomycin had a determined value of 637.3170 m/z that exactly matched the theoretical value of 637.3170 m/z. Since it was only necessary that the acrylamide monomer be attached to the aminoglycoside, purification and isolation of the cross-linked acrylamide-aminoglycoside was not required. As a result, yields were not determined.

Cross-linking of the acrylamide-aminoglycoside was performed by adding the desired amount of the acrylamide-aminoglycoside to the acrylamide mixture prior to adding tetramethylethylenediamine (TEMED) and ammonium persulfate (APS) polymerizing agents. Affinity gel electrophoresis was performed using either cross-linked ligands of acrylamide paromomycin or acrylamide neomycin B. Addition of the acrylamide-aminoglycoside had no observable effect on gel polymerization. As a control, gels were also prepared with either embedded paromomycin or embedded neomycin B [139].

RNA binding to the cross-linked acrylamide-aminoglycoside was analysed using the interaction of the TSMC RNA with aminoglycoside molecules as a model system [139, 144, 158]. The interaction of the TSMC RNA with paromomycin was previously characterized by ITC methods. In all cases where an aminoglycoside is present, the mobility of the RNA is greatly reduced compared with the gel run in the absence of ligand (Figure 4.10 and Figure 4.11). Moreover, the gels containing cross-linked paromomycin show that RNA mobility is significantly reduced compared with the embedded paromomycin (Figure 4.10 and Figure 4.12). Similarly, the gels containing cross-linked neomycin B show decreased RNA mobility compared with

the embedded neomycin B (Figure 4.11 and Figure 4.12). This decreased migration indicates that binding is occurring in the gel between the RNA and the aminoglycoside, and that cross-linking the aminoglycoside to the acrylamide results in a significant decrease of RNA mobility compared with the embedded ligand.

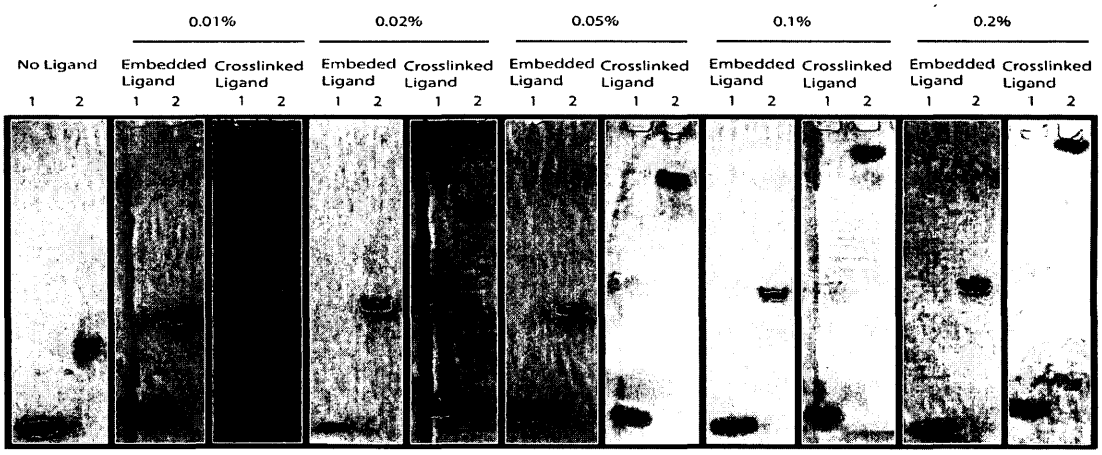


Figure 4.10: Affinity electrophoresis demonstrating the difference in mobility of the TSMC RNA between embedded and cross-linked paromomycin. Shown are separate gels containing different concentrations of paromomycin when co-polymerized within the gel matrix (embedded) versus when the ligand is cross-linked to the gel matrix. Compared to the gels with embedded aminoglycoside, decreased mobility of the TSMC RNA is observed when paromomycin is cross-linked to the gel matrix. Shown are 20% polyacrylamide gels containing 0, 0.01, 0.02, 0.05, 0.1, 0.2% (w/v) paromomycin (embedded and cross-linked). In each gel, lane 1 is the bromophenol blue internal standard and lane 2 contains the TSMC RNA.

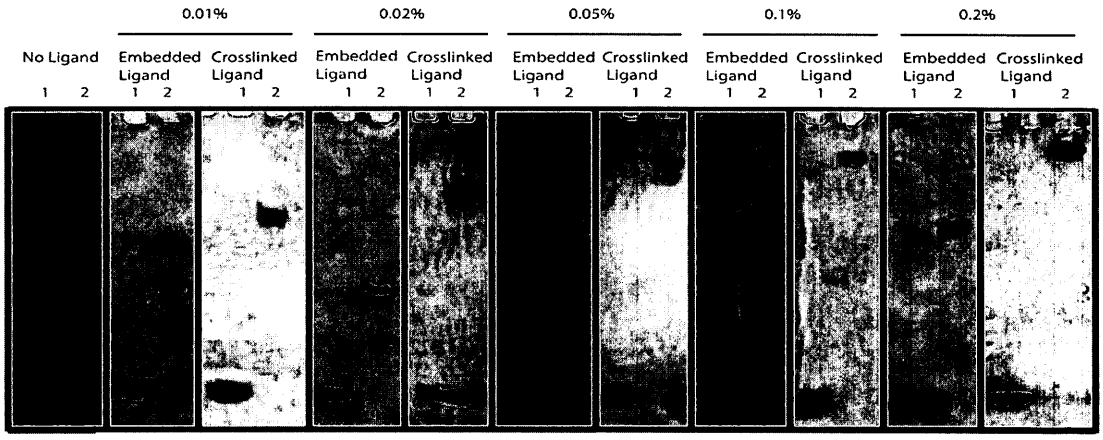


Figure 4.11: Affinity electrophoresis demonstrating the difference in mobility of the TSMC RNA between embedded and cross-linked neomycin B. Shown are separate gels containing different concentrations of neomycin B when co-polymerized within the gel matrix (embedded) versus when the ligand is cross-linked to the gel matrix. Compared to the gels with embedded aminoglycoside, decreased mobility of the TSMC RNA is observed when neomycin B is cross-linked to the gel matrix. Shown are 20% polyacrylamide gels containing 0, 0.01, 0.02, 0.05, 0.1, 0.2% (w/v) neomycin B (embedded and cross-linked). In each gel, lane 1 is the bromophenol blue internal standard and lane 2 contains the TSMC RNA.

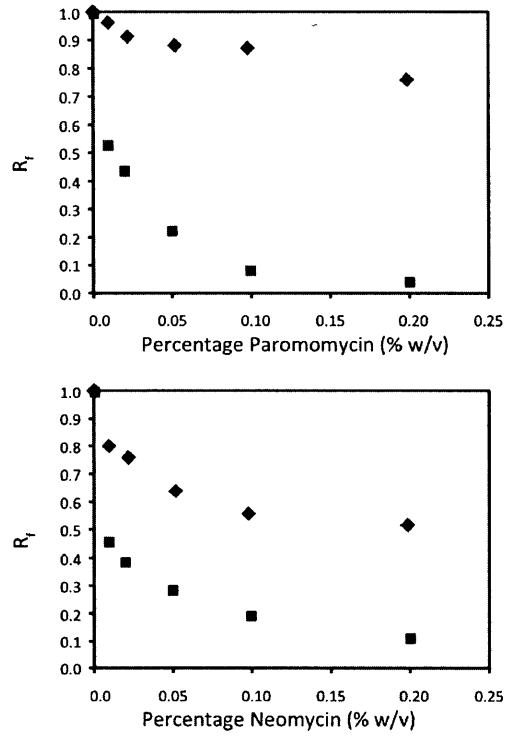


Figure 4.12: Plots comparing the relative migration (R_f) between TSMC RNA interacting with the cross-linked paromomycin and cross-linked neomycin B (red squares) and TSMC RNA interacting with the embedded paromomycin and embedded neomycin B (black diamonds).

The TSMC RNA interacted with the cross-linked aminoglycosides to produce quantifiable data. This interaction was analyzed by measuring the relative migration (R_f) of the TSMC RNA against the bromophenol blue internal standard and plotting the R_f value against the concentration of cross-linked and non-cross-linked paromomycin and to the concentration of cross-linked and non-cross-linked

neomycin B (Figure 4.12). Plotting of the reciprocal relative migration ($1/R_f$) against the concentration of cross-linked paromomycin and against the concentration of cross-linked neomycin B (Figure 4.13 (A) and (B)) generated straight lines from which the corresponding apparent dissociation constants were determined from the inverse of the slope of the lines [92]. From the data in Figure 4.13 an apparent K_d value of 156 ± 7 μM was calculated for TSMC RNA binding cross-linked paromomycin and an apparent K_d value of 450 ± 31 μM was calculated for TSMC RNA binding cross-linked neomycin B. In comparison, the use of embedded ligands is not quantifiable as a straight line does not result when plotting $1/R_f$ versus aminoglycoside concentration (Figure 4.13 C and D).

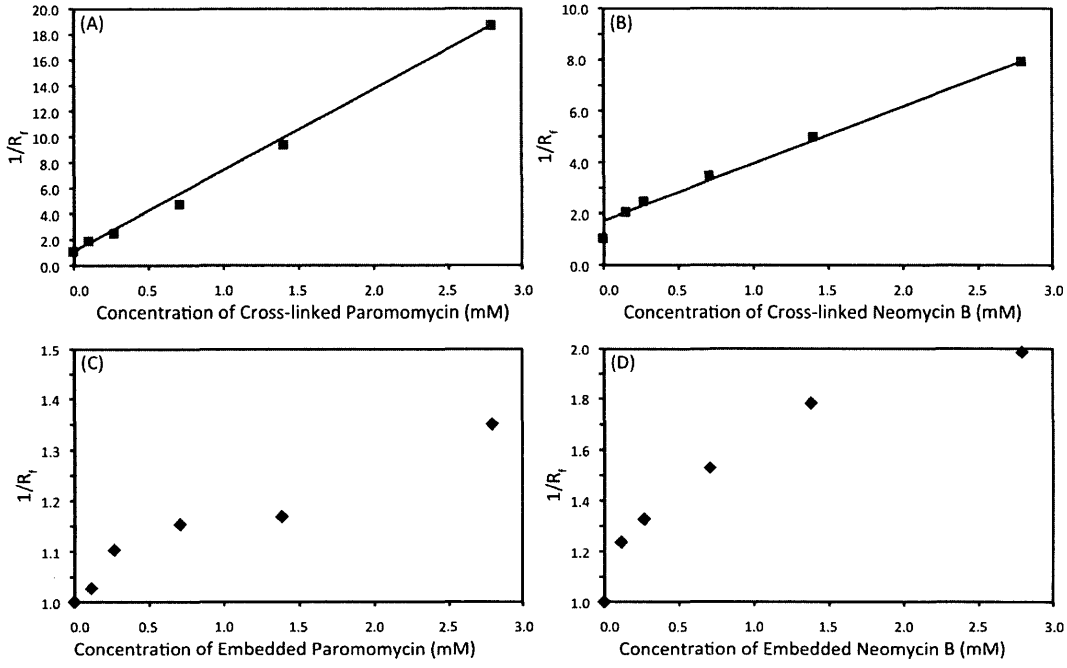


Figure 4.13: Plots showing the quantitative binding between cross-linked aminoglycosides and the TSMC RNA. Binding affinity is determined from the slope of the straight line given by the plot of the reciprocal of the relative migration ($1/R_f$) of TSMC RNA versus ligand concentration of (A) cross-linked paromomycin and (B) cross-linked neomycin B. In comparison, plots of the (C) embedded paromomycin and (D) embedded neomycin B do not result in a straight line.

4.4 DISCUSSION

4.4.1 Detection of RNA-Small Molecule Interactions

This study has shown that affinity electrophoresis can be used to screen for RNA-small molecule interactions. This assay is quick and easy to perform, and it can be carried out with readily available equipment that can be found in nearly any biochemistry laboratory. This project used affinity electrophoresis to look at RNA-aminoglycoside interactions, but it should be possible to perform affinity electrophoresis with any potential RNA binding ligand provided that it does not interfere with the gel polymerization. In addition, this screen could be run to assay different RNA sequences or structures to test their binding ability to different aminoglycsides or other RNA-binding ligands.

This method was verified using two known aminoglycoside binding RNA systems, the RNA run in the ligand-containing gel ran slower than in the absence of ligand, demonstrating that binding takes place within the gel matrix. This mobility decrease is dose dependent in that higher concentrations of ligand porogressively reduced the RNA mobility. The mobility also varied between different aminoglycoside ligands, and this mobility decrease tracked the known solution affinity between the different aminoglycosides.

Affinity electrophoresis can be applied in both a qualitative and quantitative fashion [156]. To be quantitative and obtain a dissociation constant for the biomolecule-ligand interaction from affinity electrophoresis, the ligand must not

migrate during electrophoresis. This is typically accomplished by the use of an uncharged ligand or a ligand with a large enough molecular size that it cannot migrate trough the gel during electrophoresis. Unfortunately, in the case of the RNA-aminoglycoside combination used here, these two requirements are not met. The aminoglycoside molecules are charged and are small enough to migrate through the gel during the course of the experiment. The putative migration of the ligand during electrophoresis does not prevent the observation of the occurrence of ligand binding, but it does preclude the determination of an absolute dissociation constant (K_d). In Figure 4.7, the TSMC RNA shows a dose response to an increase in the amount of paromomycin in the gel. This interaction was quantified by measuring the relative migration (R_f) of the TSMC RNA against the bromophenol blue dye and plotting the relative migration against the paromomycin concentration (Figure 4.12). For a quantitative analysis, the plot of the reciprocal of the R_f to the paromomycin concentration should result in a straight line. As seen in Figure 4.13 (C), this plot is not a straight line, confirming the non-quantitative nature of this analysis. Even though K_d values cannot be determined, it is still possible to determine relative affinity constants between two ligands of the same or very similar charge and size. One method to make the affinity electrophoresis of RNA-aminoglycoside interactions quantifiable would be to covalently cross-link the aminoglycoside to the polyacrylamide gel. This would prevent migration of the ligand during the course of the experiment, allowing the relative migration distances to be used to determine dissociation constants.

4.4.2 Quantification of RNA-Small Molecule Interactions

This research used affinity electrophoresis coupled with cross-linking of the ligand to the gel matrix to quantitatively study RNA-aminoglycoside interactions. Previously, using only embedded ligands, it was shown that binding occurred in a dose-dependent manner. However, when plotting $1/R_f$ values versus ligand concentration a straight line was not obtained with the embedded ligands [139] (Figure 4.13 (C) and (D)). This prevented binding being quantified. This non-linearity was attributed as being due to the ligand moving in the gel during electrophoresis. Here, this limitation was resolved by cross-linking the aminoglycoside directly to the gel by reacting the aminoglycoside with acryloyl chloride. During gel polymerization, the ligand becomes covalently attached to the gel. Using this method, the TSMC RNA that interacts with the aminoglycoside binds significantly tighter to the cross-linked ligand than to the embedded ligand and binding approaches saturation at the higher ligand concentrations examined (Figure 4.12). To the best of our knowledge this is the first example of cross-linking of a small-molecule ligand to the polyacrylamide gel to study RNA-small molecule interactions.

Using the cross-linked aminoglycosides an apparent K_d value was determined for the RNA-ligand interaction (Figure 4.13). The term apparent K_d was defined as when the acryloyl chloride reacted with the aminoglycoside to get a mixture of products, with the acryloyl group potentially reacting with each alcohol

and amino group. Rings I and II on the aminoglycoside (Figure 4.1) are the rings that most closely interact with the RNA [165]. When the acryloyl group reacts with alcohol or amine groups on ring I or II, binding will be obstructed and the apparent K_d value increased compared to when the acryloyl reacts at rings III or IV. This mixture of products, with some products binding weaker than others, results in the measured apparent K_d being weaker (156 ± 7 μM) than what was previously measured in solution (0.6 μM; [144]). However, this does not take away from the research's main conclusion that using the cross-linked aminoglycoside results in quantifiable data as shown by the linear plots in Figure 4.13.

The apparent K_d values showed tighter binding for the cross-linked paromomycin than for the cross-linked neomycin B (Figure 4.13). Though binding of neomycin B to the TSMC RNA construct used here has not been previously demonstrated, the binding of neomycin B to a similar CC mismatch was shown to be tighter than to paromomycin. This apparent discrepancy is likely due to the fact that neomycin B has more amino groups in ring I than has paromomycin (Figure 4.1). Since the pK_a of the amino is lower (approximate pK_a = 9-10) than that of the hydroxyl group (pK_a = 16-17), the one equivalent of sodium bicarbonate will preferentially deprotonate the amino to form a reactive amine with an available lone pair that immediately reacts with acryloyl chloride. Since neomycin B contains more amine groups at ring I, reaction with acryloyl chloride reduces the ligand binding ability of neomycin B for the TSMC RNA to a greater extent than reaction of paromomycin.

This method was primarily used to quantify RNA-aminoglycoside interactions, however it should be possible to apply this method to any ligand that binds RNA as long as the cross-linked ligand does not interfere with gel polymerization. In addition, this cross-linked ligand affinity electrophoresis method is particularly useful for comparing interactions between different RNA molecules to the same aminoglycosides by running a constant concentration of each RNA on the same gel.

4.5 CONCLUSIONS

We have presented an affinity electrophoresis-based method for screening RNA-small molecule interactions. This method is rapid and easy to perform with commonly available equipment. Although this research demonstrated the use of affinity electrophoresis using the well-established A-site-aminoglycoside and TSMC-aminoglycoside interactions, this method should be applicable for gauging the binding of any RNA molecule to any small molecule embedded in a gel and for screening the potential binding of RNA molecules of different sequences and structures to known RNA ligands.

Furthermore, this research has shown that an affinity electrophoresis based method can be made quantifiable for determining dissociation constants of RNA-aminoglycoside interactions. This method can easily be applied to other RNA-small molecules using equipment and reagents common to most laboratories. Furthermore, synthesis of other cross-linked aminoglycosides and subsequent testing to different RNA molecules with various sequences and sizes can also be achieved following our method to obtain quantitative binding information.

REFERENCES

1. Cusack, S., *RNA-protein complexes.* Curr. Opin. Struct. Biol., 1999. **9**: p. 66-73.
2. Boodram, S.N. and P.E. Johnson, *Finding the path in an RNA folding pathway.* Structure, 2010. **18**: p. 1550-1.
3. Bujnicki, J.M., *Prediction of Protein Structures, Functions and Interactions*, ed. J.M. Bujnicki. 2009, United Kingdom: John Wiley & Sons Ltd.
4. Jones, S., D.T.A. Daley, N.M. Luscombe, H.M. Berman, and J.M. Thornton, *Protein-RNA interactions: a structural analysis.* Nucleic Acids Res., 2001. **29**: p. 943-54.
5. Saunders, L.R. and G.N. Barber, *The dsRNA binding protein family: critical roles, diverse cellular functions.* FASEB J., 2003. **17**: p. 961-83.
6. Wang, Z., E. Hartman, K. Roy, G. Chanfreau, and J. Feigon, *Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs.* Structure, 2011. **19**: p. 999-1010.
7. Masliah, G., P. Barraud, and F. Allain, H.-T., *RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence.* Cell Mol. Life Sci., 2013. **70**: p. 1875-95.
8. Klotz, I.M., *Protein interactions with small molecules.* Acc. Chem. Res., 1974. **7**: p. 162-8.

9. Paoli, M., R. Liddington, J. Tame, A. Wilkinson, and G. Dodson, *Crystal structure of T state haemoglobin with oxygen bound at all four haems.* J. Mol. Biol., 1996. **256**: p. 775-92.

10. Lenz, T., J.J. Fischer, and M. Dreger, *Probing small molecule-proein interactions: A new perspective for functional proteomics.* J. Proteomics, 2011. **75**: p. 100-15.

11. Taylor, R.D., P.J. Jewsbury, and J.W. Essex, *A review of protein-small molecule docking methods.* J. Comput. Aided Mol. Des., 2002. **16**: p. 151-66.

12. Bantscheff, M., C. Hopf, U. Kruse, and G. Drewes, *Proteomics-based strategies in kinase drug discovery.* Ernst Schering Found. Symp. Proc., 2007. **3**: p. 1-28.

13. Rix, U. and G. Superti-Furga, *Target profiling of small molecules by chemical proteomics.* Nat. Chem. Biol., 2009. **5**: p. 616-24.

14. Greenbaum, D., K.F. Medzihradszky, A. Burlingame, and M. Bogyo, *Epoxide electophiles as activity-dependent cysteine protease profiling and discovery tools.* Chem. Biol., 2000. **7**: p. 569-81.

15. Kato, D., K.M. Boatright, A.B. Berger, T. Nazif, and e. al., *Activity-based probes that target diverse cysteine protease families.* Nat. Chem. Biol., 2005. **1**: p. 33-8.

16. Luo, Y., C. Blex, O. Baessler, M. Glinski, M. Dreger, M. Sefkow, and H. Koester, *The cAMP capture compound mass spectometry as a novel tool for targeting cAMP-binding proteins: from protein kinase A to potassium/sodium*

hyperpolarization-activated cyclic nucleotide-gated channels. Mol. Cell Proteomics, 2009. **8**: p. 2843-56.

17. Luo, Y., J.J. Fischer, O.Y. Baessler, A.K. Schrey, J. Ungewiss, M. Glinski, M. Sefkow, M. Dreger, and H. Koester, *GDP-capture compound-a novel tool for the profiling of GTPases in pro- and eukaryotes by capture compound mass spectrometry (CCMS).* J. Proteomics, 2010. **73**: p. 815-9.
18. Wang, W.U., C. Chen, K.-h. Lin, Y. Fang, and C. Lieber, *Label-free detection of small-molecule-protein interactions by using nanowire nanosensors.* PNAS, 2005. **102**(9): p. 3208-12.
19. Cheng, K.W., C.C. Wong, M. Wang, Q.Y. He, and F. Cheng, *Identification and characterization of molecular targets of natural products by mass spectrometry.* Mass Spectrom. Rev., 2010. **29**: p. 126-55.
20. Harvey, I., P. Garneau, and J. Pelletier, *Inhibition of translation by RNA-small molecule interactions.* RNA, 2002. **8**: p. 452-63.
21. Jovanovic, M. and M.O. Hengartner, *miRNAs and apoptosis: RNAs to die for.* Oncogene, 2006. **25**: p. 6176-87.
22. Fedor, M.J. and J.R. Williamson, *The catalytic diversity of RNAs.* Nat. Rev. Mol. Cell Biol., 2005. **6**: p. 399-412.
23. Sledz, C.A., M. Holko, M.J. de Veer, R.H. Silverman, and B.R. Williams, *Activation of the interferon system by short-interfering RNAs.* Nat. Cell Biol., 2003. **5**: p. 834-9.

24. Thomas, J.R. and P.J. Hergenrother, *Targeting RNA with small molecules.* Chem. Rev., 2008. **108**: p. 1171-1224.

25. Vicens, Q. and E. Westhof, *Molecular Recognition of Aminoglycoside Antibiotics by Ribosomal RNA and Resistance Enzymes: An Analysis of X-Ray Crystal Structures.* Biopolymers, 2003. **70**: p. 42-57.

26. Blount, K.F., J.X. Wang, J. Lim, N. Sudarsan, and R.R. Breaker, *Antibacterial lysine analogs that target lysine riboswitches.* Nat. Chem. Biol., 2007. **3**: p. 44-9.

27. Schroeder, K.T., P. Daldrop, and D.M. Lilley, *RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn.* Structure, 2011. **19**: p. 1233-40.

28. Matthews, D.H., M.D. Disney, J.L. Childs, S.J. Schroeder, M. Zuker, and D.H. Turner, *Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.* Proc. Natl. Acad. Sci. USA, 2004. **101**: p. 7287-92.

29. Fire, A., S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, and C.C. Mello, *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**: p. 806-11.

30. Johnson, P.H., *RNA Interference: application to drug discovery and challenges to pharmaceutical development.* 2011, New Jersey: John Wiley & Sons, Inc. 310.

31. Zhang, B., X. Pan, G.P. Cobb, and T.A. Anderson, *Plant microRNA: a small regulatory molecule with big impact.* Dev. Biol., 2006. **289**: p. 3-16.

32. Sharp, P.A., *RNA-interference* Genes and Devel., 2001. **15**: p. 485-90.

33. Meister, G. and T. Tuschl, *Mechanisms of gene silencing by double-stranded RNA.* Nature, 2004. **431**: p. 343-49.

34. Zamore, P.D. and B. Haley, *Ribo-gnome: The Big World of Small RNAs.* Science, 2005. **309**: p. 1519-24.

35. Siomi, H. and M.C. Siomi, *On the Road to Reading the RNA-Interference Code.* Nature, 2009. **257**: p. 396-404.

36. Obbard, D.J., F.M. Jiggin, D.L. Halligan, and T.J. Little, *Natural selection drives extremely rapid evolution in antiviral RNAi genes.* Curr. Biol., 2006. **16**: p. 580-5.

37. Bennetzen, J.L., *Transposable element contributes to place gene and genome evolution.* Plant Mol. Biol., 2000. **42**: p. 251-69.

38. Ishizu, H., H. Siomi, and M.C. Siomi, *Biology of Piwi-interacting RNAs: new insights into biogenesis and function inside and outside of germlines.* Genes Dev., 2012. **26**: p. 2361-73.

39. Carthew, R.W. and E.J. Sontheimer, *Origins and mechanisms of miRNAs and siRNAs.* Cell, 2009. **136**: p. 642-55.

40. Cai, X., C.H. Hagedorn, and B.R. Cullen, *Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs.* RNA, 2004. **10**: p. 1957-66.

41. Andreassen, H., et al., *Analysis of the Secondary Structure of the Human Immunodeficiency Virus (HIV) Proteins p17, gp120, and gp41 by Computer Modeling Based on Neural Network Methods.* Journal of AIDS, 1990. **3**: p. 615-22.

42. Sontheimer, E. and R. Carthew, *Silence from within: endogenous siRNAs and miRNAs.* Cell, 2005. **122**: p. 9-12.

43. Sontheimer, E.J., *Assembly and function of RNA silencing complexes.* Nature Rev. Mol. Cell. Biol., 2005. **6**: p. 127-38.

44. Carrington, J.C. and V. Ambros, *Role of microRNAs in plant and animal development.* Science, 2003. **301**: p. 336-8.

45. Saumet, A. and C.-H. Lecellier, *Anti-viral RNA silencing: do we look like plants?.* . Retrovirology, 2006. **3**: p. 3.

46. Pfeffer, S., et al., *Identification of microRNAs of the herpesvirus family.* Nat. Methods, 2005. **2**: p. 269-76.

47. Sullivan, C.S., A.T. Grundhoff, S. Tevethia, J.M. Pipas, and D. Ganem, *SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells.* Nature, 2005. **435**: p. 682-6.

48. Lecellier, C., P. Dunoyer, K. Arar, J. Lehmann-Che, S. Eyquem, C. Himber, A. Saib, and O. Voinnet, *A cellular microRNA mediates antiviral defense in human cells.* Science, 2005. **308**: p. 557-60.

49. Calin, G., et al., *Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.* Proc. Natl. Acad. Sci. U.S.A., 2002. **99**: p. 15524-29.

50. Kim, K., G. Chadalapaka, S.-O. Lee, D. Yamada, X. Sastre-Garau, P.-A. Defossez, Y.-Y. Park, J.-S. Lee, and S. Safe, *Identification of oncogenic microRNA-17-92/ZBTB4/specificity protein axis in breast cancer.* Oncogene, 2012. **31**: p. 1034-44.

51. Zhang, B., X. Pan, G. Cobb, and T. Anderson, *microRNAs as oncogenes and tumor suppressors.* Developmental Biology, 2007. **302**: p. 1-12.

52. Johnson, S., et al., *RAS is regulated by the let-7 microRNA family.* Cell, 2005. **120**: p. 635-47.

53. Buechner, J., E. Tomte, B. Haug, J. Henriksen, C. Lokke, T. Flaegstad, and C. Einvik, *Tumor-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma.* Br. J. Cancer, 2011. **105**: p. 296-303.

54. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**: p. 834-8.

55. Jinek, M. and J.A. Doudna, *A three-dimensional view of the molecular machinery of RNA interference.* Nature, 2009. **457**: p. 405-12.

56. Keeler, J., *Understanding NMR Spectroscopy.* 2nd ed 2010, West Sussex: John Wiley & Sons Ltd.

57. Rule, G.S. and T.K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*, ed. R. Kaptein. Vol. 5. 2006, Dordrecht: Springer.

58. Sanders, J.K.M. and B.K. Hunter, *Modern NMR Spectroscopy: A guide for chemists*. 2nd ed 1993, Toronto: Oxford University Press.

59. Kanelis, V., J.D. Forman-Kay, and L.E. Kay, *Multidimensional NMR methods for protein structure determination.* IUBMB Life, 2001. **52**: p. 291-302.

60. Clore, G.M., L.E. Kay, A. Bax, and A.M. Gronenborn, *Four-dimensional 13C/13C-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: Application to interleukin 1b.* Biochemistry, 1991. **30**: p. 12-18.

61. Clore, G.M. and A.M. Gronenborn, *Structures of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy.* Science, 1991. **252**: p. 1390-9.

62. Gardner, K.H. and L.E. Kay, *The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins.* Annu. Rev. Biophys. Biomol. Struct., 1998. **27**: p. 357-406.

63. Sattler, M., J. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients.* Prog. Nucl. Mag. Res. Sp., 1999. **34**: p. 93-158.

64. Clore, G.M. and A.M. Gronenborn, *Determining the structures of large proteins and protein complexes by NMR.* Trends Biotechnol., 1998. **16**: p. 22-34.

65. Cavanagh, J., W.J. Fairbrother, A.G. Palmer III, M. Rance, and N.J. Skelton, *Protein NMR Spectroscopy: Principles and Practice.* 2nd ed 2007, California: Elsevier.

66. Akitt, J.W., *NMR and Chemistry: An Introduction to Modern NMR Spectroscopy.* 3rd ed 1992, New York: Chapman & Hall.

67. Derome, A.E., *Modern NMR Techniques for Chemistry Research.* 1987, Oxford: Pergamon Press. 295.

68. Grzesiek, S. *Notes on relaxation and dynamics.* in *EMBO Practical Course: Structure determination of biological macromolecules by solution NMR.* 2003. Heidelberg.

69. Renner, C., M. Schleicher, L. Moroder, and T.A. Holak, *Practical aspects of the 2D ^{15}N-{^{1}H}-NOE experiment.* J. Biomol. NMR, 2002. **23**: p. 23-33.

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

71. Atkinson, A.R. and B. Kieffer, *The role of protein motions in molecular recognition: insights from heteronuclear NMR relaxation measurements.* Prog. Nucl. Magn. Reson. Spectrosc., 2004. **44**: p. 141-87.

72. Farrow, N.A., O. Zhang, J.D. Forman-Kay, and L.E. Kay, *Comparison of the backbone dynamics of a folded and unfolded SH3 domain existing in equilibrium in aqueous buffer.* Biochemistry, 1995. **34**: p. 868-78.

73. Abragam, A., *The principles of nuclear magnetism.* The international series of monographs on physics, ed. N.F. Mott and D.H. Wilkinson 1961, Oxford: Oxford University Press.

74. Peng, J.W. and G. Wagner, *Mapping of spectral density functions using herteronuclear NMR relaxation measurements.* J. Magn. Reson., 1992. **98**: p. 308-32.

75. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity.* J. Am. Chem. Soc., 1982. **104**: p. 4546-4559.

76. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results.* J. Am. Chem. Soc., 1982. **104**: p. 4559-70.

77. Jardetzky, O. and J.-F. Lefèvre, eds. *Protein Dynamics, Function, and Design.* NATO Science Series A. Vol. 301. 1998, Plenum Press: New York. 222.

78. McPherson, A., *Introduction to Macromolecular Crystallography.* 2003, New Jersey: John Wiley & Sons Inc. 237.

79. Drenth, J., *Principles of protein X-ray crystallography.* 2007: Springer Advanced Texts in Chemistry.

80. Chayen, E. and Naomi, *Methods for separating nucleation and growth in protein crystallisation.* Prog. Biophys. Mol. Bio., 2005. **88**: p. 329-37.

81. Petoukhov, M.V. and D.I. Svergun, *Applications of small-angle X-ray scattering to biomacromolecular solutions.* Int. J. Biochem. Cell Biol., 2013. **45**: p. 429-37.

82. Mertens, H.D. and D.I. Svergun, *Structural characterization of proteins and complexes using small-angle X-ray solution scattering.* J. Struct. Biol., 2010. **172**: p. 128-41.

83. Tsutakawa, S.E., G.L. Hura, K.A. Frankel, P.K. Cooper, and J.A. Tainer, *Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography.* J. Struct. Biol., 2007. **158**: p. 214-23.

84. Pons, C., M. D'Abramo, D.I. Svergun, M. Orozco, P. Bernado, and J. Fernandez-Recio, *Structural characterization of protein-protein complexes by integrating cmputational docking with small-angle scattering data.* J. Mol. Biol., 2010. **403**: p. 217-30.

85. Yang, S., M. Parisien, F. Major, and R. Benoit, *RNA structure determination using SAXS data.* J. Phys. Chem. B, 2010. **114**: p. 10039-48.

86. Nienhaus, G.U., ed. *Protein-Ligand Interactions: Methods and Applications.* Methods in Molecular Biology, ed. J.M. Walker 2005, Humana Press Inc.: Totowa. 568.

87. Liang, Y., *Applications of isothermal titration calorimetry in protein science.* Acta Biochim. Biophys. Sin., 2008. **40**: p. 565-76.

88. Holdgate, G.A. and W.H.J. Ward, *Measurements of binding thermodynamics in drug discovery.* Drug Discov. Today, 2005. **10**: p. 1543-50.

89. Harding, S.E. and B.Z. Chowdhry, eds. *Protein-Ligand Interactions: hydrodynamics and calorimetry.* The Practical Approach Series. 2001, Oxford University Press Inc.: New York. 354.

90. Bog-Hansen, T.C., *Crossed immuno-Affinityelectrophoresis: An analytical Method to Predict the Result of Affinity Chromatography.* Anal. Biochem., 1973. **56**: p. 480-8.

91. Takeo, K., *Affinity electrophoresis: Principles and applications.* Electrophoresis, 1984. **5**: p. 187-95.

92. Takeo, K., *Advances in affinity electrophoresis.* J. Chromatogr., A, 1995. **698**: p. 89-105.

93. Anderot, M., M. Nilsson, A. Vegvari, E.H. Moeller, M. van de Weert, and R. Isaksson, *Determination of dissociation constants between polyelectrolytes and proteins by affinity capillary electrophoresis.* J. Chromatogr. B Analyt. Technol. Biomed. Life Sci., 2009. **877**: p. 892-6.

94. Kiessig, S., H. Bang, and F. Thunecke, *Interaction of cyclophilin and cyclosporins monitored by affinity capillary electrophoresis.* J. Chromatogr. A, 1999. **853**: p. 469-77.

95. Tabara, H., E. Yigit, H. Siomi, and C.C. Mello, *The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1 and a DexH-Box helicase to direct RNAi in C. elegans.* Cell, 2002. **109**: p. 861-71.

96. Parker, G., D.M. Eckert, and B.L. Bass, *RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA.* RNA, 2006. **12**: p. 807-18.

97. Wu, H., A. Henras, G. Chanfreau, and J. Feigon, *Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III.* Proc. Natl. Acad. USA, 2004. **101**: p. 8307-12.

98. Leulliot, N., et al., *A new a-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III.* EMBO J, 2004. **23**: p. 2468-77.

99. Chang, K.-Y. and A. Ramos, *The double-stranded RNA-binding motif, a versatile macromolecular docking platform.* FEBS J, 2005: p. 2109-17.

100. Stefl, R., M. Xu, L. Skrisovska, R.B. Emeson, and F.H. Allain, *Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs.* Structure, 2006. **14**: p. 345-55.

101. Nanduri, S., B.W. Carpick, Y. Yang, B.R.G. Williams, and J. Qin, *Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation.* EMBO J, 1998. **17**: p. 5458-65.

102. Sohn, S.Y., W.J. ae, J.J. Kim, K.-H. Yeom, V.N. Kim, and Y. Cho, *Crystal structure of human DGCR8 core.* Nature Struct. Mol. Biol., 2007. **14**: p. 847-53.

103. Aagaard, L. and J.J. Rossi, *RNAi therapeutics: principles, prospects and challenges.* Adv. Drug Deliv. Rev., 2007. **59**: p. 75-86.

104. Rydzanicz, R., X.S. Zhao, and P.E. Johnson, *Assembly PCR Oligo Maker: A tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production.* Nucleic Acids Res, 2005. **33**: p. W521-5.

105. Johnson, B.A. and R.A. Blevins, *NMRView: A computer program for the visualization and analysis of NMR data.* J. Biomol. NMR, 1994. **4**: p. 603-14.

106. Svergun, D.I., M.V. Petoukhov, and M.H.J. Koch, *Determination of domain structure of proteins from X-ray solution scattering.* Biophys. J., 2001. **80**: p. 2946-53.

107. Zhu, X.-Q., S.-X. Li, H.-J. He, and Q.-S. Yuan, *On-column refolding of an insoluble His(6)-tagged recombinant EC-SOD overexpressed in Escherichia coli.* Acta Biochim. Biophys. Sin., 2005. **37**: p. 265-9.

108. Chiliveri, S.C.K., Sonu; Marelli, Udaya Kiran; Deshmukh, Mandar V., *Backbone and sidechain methyl Ile (d1), Leu and Val chemical shift assignments of RDE-4 (1-243), an RNA interference inititiation protein in C. elegans.* Biomol. NMR Assign., 2012. **6**: p. 143-6.

109. Haasnoot, C.A.G., F.A.A.M. de Leeuw, and C. Altona, *The relationship between proton-proton NMR coupling constants and substituent electronegativities I.* Tetrahedron, 1980. **36**: p. 2783-92.

110. Parker, G.S., T.S. Maity, and B.L. Bass, *dsRNA Binding Properties of RDE-4 and TRBP Reflect Their Distinct Roles in RNAi.* J. Mol. Biol., 2008. **384**: p. 967-79.

111. Boehr, D.D., D. McElheny, H.J. Dyson, and P.E. Wright, *The dynamic energy landscape of dihydrofolate reductase catalysis.* Science, 2006. **313**: p. 1638-42.

112. Hawser, S., S. Lociuro, and K. Islam, *Dihydrofolate reductase inhibitors as antibacterial agents.* Biochem. Pharmacol., 2006. **71**: p. 941-8.

113. Boehr, D.D., H.J. Dyson, and P.E. Wright, *Conformational relaxation following hydride transfer plays a limited role in dihydrofolate reductase catalysis.* Biochemistry, 2008. **47**: p. 9227-33.

114. Mauldin, R.V., M.J. Carroll, and A.L. Lee, *Dynamic dysfunction in dihydrofolate reductase results from antifolate drug binding: modualtion of dynamics within a structural state.* Structure, 2009. **17**: p. 386-94.

115. McElheny, D., J.R. Schnell, J.C. Lansing, H.J. Dyson, and P.E. Wright, *Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis.* Proc. Nat. Acad. Sci. U.S.A., 2005. **102**: p. 5032-7.

116. Batruch, I., E. Javasky, E.D. Brown, M.G. Organ, and P.E. Johnson, *Thermodynamic and NMR analysis of inhibitor binding to dihydrofolate reductase.* Bioorg. Med. Chem., 2010. **18**: p. 8485-92.

117. Sawaya, M.R. and J. Kraut, *Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence.* Biochemistry, 1997. **36**: p. 586-603.

118. Osborne, M.J., R.P. Venkitakrishnan, H.J. Dyson, and P.E. Wright, *Diagnostic chemical shift markers for loop conformation and substrate and cofactor binding in dihydrofolate reductase complexes.* Protein Science, 2003. **12**: p. 2230-8.

119. Venkitakrishnan, R.P., E. Zaborowski, D. McElheny, S.J. Benkovic, H.J. Dyson, and P.E. Wright, *Conformational changes in the active site loops of*

dihydrofolate reductase during the catalytic cycle. Biochemistry, 2004. **43**: p. 16046-55.

120. Schnell, J.R., H.J. Dyson, and P.E. Wright, *Structure, dynamics, and catalytic function of dihydrofolate reductase.* Annu. Rev. Biophys. Biomol. Struct., 2004. **33**: p. 119-40.

121. Epstein, D.M., S.J. Benkovic, and P.E. Wright, *Dynamics of the dihydrofolate reductase-folate complex: catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features.* Biochemistry, 1995. **34**: p. 11037-48.

122. Osborne, M.J., J.R. Schnell, S.J. Benkovic, H.J. Dyson, and P.E. Wright, *Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism.* Biochemistry, 2001. **40**: p. 9846-59.

123. Schnell, J.R., H.J. Dyson, and P.E. Wright, *Effect of cofactor binding and loop conformation on side chain methyl dynamics in dihydrofolate reductase.* Biochemistry, 2004. **43**: p. 374-83.

124. Li, L., C.J. Falzone, P.E. Wright, and S.J. Benkovic, *Functional role of a mobile loop of Escherichia coli dihydrofolate reductase in transition-state stabilization.* Biochemistry, 1992. **31**: p. 7826-33.

125. Delagio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J. Biomol. NMR., 1995. **6**: p. 277-93.

126. Summerfield, R.L., et al., *A 2.13Å structure of E. coli Dihydrofolate reductase bound to a novel competitive inhibitor reveals a new binding surface involving the M20 loop region.* J. Med. Chem., 2006. **49**: p. 6977-86.

127. Neudecker, P., P. Lundstrom, and L.E. Kay, *Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding.* Biophys. J., 2009. **96**: p. 2045-54.

128. Abraham, R.J., J. Fisher, and P. Loftus, *Introduction to NMR Spectroscopy.* 1988, New York: John Wiley & Sons Ltd.

129. Clore, G.M., P.C. Driscoll, P.T. Wingfield, and A.M. Gronenborn, *Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy.* Biochemistry, 1990. **29**: p. 7387-7401.

130. Gomez, J., V.J. Hilser, D. Xie, and E. Freire, *The heat capacity of proteins.* Proteins Struct. Funct. Genet., 1995. **22**: p. 404-12.

131. Spolar, R.S. and M. Record Jr., *Coupling of local folding to site-specific binding of proteins to DNA.* Science, 1994. **263**: p. 777-84.

132. Pedretti, A., L. Villa, and G. Vistoli, *VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming.* J. Comput.-Aided Mol. Des. , 2004. **18**: p. 167-73.

133. Ertl, P., B. Rohde, and P. Selzer, *Fast calculation of molecular polar surface area as a sum of fragment based contributions and its application to the prediction of drug transport properties.* J. Med. Chem., 2000. **43**: p. 3714-7.

134. Baujard, O., V. Baujard, S. Aurel, C. Boyer, and R.D. Appel, *MARVIN, multi-agent softbot to retrieve multilingual medical information on the Web.* Med. Inform. (Lond), 1998. **23**: p. 187-91.

135. Birdsall, B., A.S.V. Burgen, and G.C.K. Roberts, *Binding of coenzyme analogs to Lactobacillus casei dihydrofolate reductase: binary and ternary complexes.* Biochemistry, 1980. **19**: p. 3723-31.

136. Fierke, C.A., K.A. Johnson, and S.J. Benkovic, *Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli.* Biochemistry, 1987. **26**: p. 4085-92.

137. Schweitzer, B.I., A.P. Dicker, and J.R. Bertino, *Dihydrofolate reductase as a therapeutic target.* FASEB J., 1990. **4**: p. 2441-52.

138. Bystroff, C. and J. Kraut, *Crystal structure of unliganded Escherichia coli dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding.* Biochemistry, 1991. **30**: p. 2227-39.

139. Boodram, S.N., C.M. Cho, T.J. Tavares, and P.E. Johnson, *Identification of RNA-ligand interactions by affinity electrophoresis.* Anal. Biochem., 2011. **409**: p. 54-9.

140. Montange, R.K. and R.T. Batey, *Riboswitches: emerging themes in RNA structure and function.* Annu. Rev. Biophys., 2008. **37**: p. 117-33.

141. Roth, A. and R.R. Breaker, *The structure and functional diversity of metabolite-binding riboswitches.* Annu. Rev. Biochem., 2009. **78**: p. 305-34.

142. Mayer, G., *The chemical biology of aptamers.* Angew. Chem., Int. Ed., 2009. **48**: p. 2672-89.

143. Magnet, S. and J.S. Blanchard, *Molecular Insights into Aminoglycoside Action and Resistance.* Chem. Rev., 2005. **105**: p. 477-97.

144. Tok, J.B.-H., J. Cho, and R.R. Rando, *Aminoglycoside Antibiotics Are Able To Specifically Bind the 5'-Untranslated Region of Thymidylate Synthase Messenger RNA.* Biochemistry, 1999. **38**: p. 199-206.

145. Tor, Y., *Targeting RNA with Small Molecules.* ChemBioChem, 2003. **4**: p. 998-1007.

146. Chow, C.S. and F.M. Bogdan, *A Structural basis for RNA-ligand interactions.* Chem. Rev., 1997. **97**: p. 1489-1513.

147. Aboul-ela, F., *Strategies for the design of RNA-binding small molecules.* Future Med. Chem., 2010. **2**: p. 93-119.

148. Sharp, P.A., *The centrality of RNA.* Cell, 2009. **136**: p. 577-80.

149. Hellman, L.M. and M.G. Fried, *Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions.* Nat. Proctoc., 2007. **2**: p. 1849-61.

150. Gagnon, K.T. and E.S. Maxwell, *Electrophoretic mobility shift assay for characterizing RNA-protein interaction.* Methods Mol. Biol., 2011. **703**: p. 275-91.

151. Tomme, P., A.L. Creagh, D.G. Kilburn, and C.A. Haynes, *Interaction of polysaccharides with the N-terminal cellulose-binding domain of Cellulomonas*

fimi CenC. 1. Binding specificity and calorimetric analysis. Biochemistry, 1996. **35**: p. 13885-94.

152. Boraston, A.B., P. Tomme, E.A. Amandoron, and D.G. Kilburn, *A novel mechnism of xylan binding by a lectin-like module from Streptomyces lividans xylanase 10A.* Biochem. J., 2000. **350**: p. 933-41.

153. Kormos, J.M., P.E. Johnson, E. Brun, P. Tomme, L.P. McIntosh, C.A. Haynes, and D.G. Kilburn, *Binding site analysis of cellulose binding domain CBD_{N1} from endoglucanase C of Cellulomonas fimi by site-directed mutagenesis.* Biochemistry, 2000. **39**: p. 8844-52.

154. Tomme, P., A. Boraston, J.M. Kormos, R.A.J. Warren, and D.G. Kilburn, *Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules.* Enzyme Microb. Technol., 2000. **27**: p. 453-8.

155. Cilley, C.D. and J.R. Williamson, *Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE).* RNA, 1997. **3**: p. 57-67.

156. Horejsi, V., *Affinity electrophoresis.* Anal. Biochem., 1981. **112**: p. 1-8.

157. Milligan, J.F., D.R. Groebe, G.W. Witherell, and O.C. Uhlenbeck, *Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates.* Nucleic Acids Res., 1987. **15**: p. 8783-98.

158. Tavares, T.J., A.V. Beribisky, and P.E. Johnson, *Structure of the Cytosine-Cytosine mismatch in the thymidylate synthase mRNA binding site and analysis*

of its interaction with the aminoglycoside paromomycin. RNA, 2009. **15**: p. 911-22.

159. Fourmy, D., M.I. Recht, S.C. Blanchard, and J.D. Puglisi, *Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic.* Science, 1996. **274**: p. 1367-71.

160. Maddaford, S.P., M. Motamed, T. K., M.S.K. Choi, J. Ramnauth, S. Rakhit, R. Hudgins, D. Fabris, and P.E. Johnson, *Identification of a novel non-carbohydrate molecule that binds to the ribosomal A-site RNA.* Bioorg. Med. Chem. Lett., 2004. **14**: p. 5987-90.

161. Ryu, D.H. and R.R. Rando, *Aminoglycoside binding to human and bacterial A-site rRNA decoding region constructs.* Bioorg. Med. Chem., 2001. **9**: p. 2601-8.

162. Kaul, M. and D.S. Pilch, *Thermodynamics of Aminoglycoside-rRNA Reocognition: The Binding of Neomycin-Class Aminoglycosides to the A Site of 16 rRNA.* Biochemistry, 2002. **41**: p. 7695-7706.

163. Kaul, M., C.M. Barbieri, and D.S. Pilch, *Aminoglycoside-Induced Reduction in Nucleotide Mobility at the Ribosomal RNA A-Site as a Potentially Key Determinant of Antibacterial Activity* J. Am. Chem. Soc., 2006. **128**: p. 1261-71.

164. Kondo, J., M. Hainrichson, I. Nudelman, D. Shallom-Shezifi, C.M. Barbieri, D.S. Pilch, E. Westhof, and T. Baasov, *Differential selectivity of natural and synthetic aminioglycosides towards the eukaryotic and prokaryotic decoding A sites.* ChemBioChem, 2007. **8**: p. 1700-9.

165. Vicens, Q. and E. Westhof, *Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site.* Structure, 2001. **9**: p. 647-58.