Chemistry Dissertations
Department of Chemistry

Summer 7-1-2013

Macromolecular Interactions in West Nile Virus RNA-TIAR Protein Complexes and of Membrane Associated Kv Channel Peptides

Jin Zhang

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

MACROMOLECULAR INTERACTIONS IN WEST NILE VIRUS RNA-TIAR PROTEIN

COMPLEXES AND OF MEMBRANE ASSOCIATED KV CHANNEL PEPTIDES

by

JIN ZHANG

Under the Direction of Markus W. Germann

ABSTRACT

Macromolecular interactions play very important roles in regulation of all levels of bio-logical processes. Aberrant macromolecular interactions often result in diseases. By applying a combination of spectroscopy, calorimetry, computation and other techniques, the protein-protein interactions in the system of the Shaw2 Kv channel and the protein-RNA interactions in West Nile virus RNA-cellular protein TIAR complex were explored. In the former system, the results shed light on the local structures of the key channel components and their potential interaction

mediated by butanol, a general anesthetic. In the later studies, the binding modes of TIAR RRM2 to oligoU RNAs and West Nile virus RNAs were investigated. These findings provided insights into the basis of the specific cellular protein–viral RNA interaction and preliminary data for the development of strategies on how to interfere with virus replication.

INDEX WORDS: Macromolecular interaction, Thermodynamics, *tran-cis* Isomerization, Exchange spectroscopy, Cyclization, Shaw2 Kv channel, S4-S5 linker, S6, Anesthetics, DPC micelles, Paramagnetic relaxation enhancement, West Nile virus, RRM2, Binding interactions, Molecular docking, Isothermal titration calorimetry, Adenine H2, Homonuclear coupling, Long-range coupling

MACROMOLECULAR INTERACTIONS IN WEST NILE VIRUS RNA-TIAR PROTEIN

COMPLEXES AND OF MEMBRANE ASSOCIATED KV CHANNEL PEPTIDES

by

JIN ZHANG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2013

MACROMOLECULAR INTERACTIONS IN WEST NILE VIRUS RNA-TIAR PROTEIN

COMPLEXES AND OF MEMBRANE ASSOCIATED KV CHANNEL PEPTIDES

by

JIN ZHANG

Committee Chair: Markus W. Germann

Committee: Margo A. Brinton

W. David Wilson

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2013

DEDICATION

This dissertation is dedicated to my parents Shenggui and Yonglan, my brother Yu, my husband Yongzheng and my baby girl Ruohan. It is because of you that I never give up.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Markus Germann for his patience, trust, inspiration and friendship during the entire period of my graduate studies. He has given me precious opportunities to work on challenging and extremely interesting projects and to explore my own professional interests. He has always made himself available for help and advice. It has been a great pleasure to work with and learn from him.

I owe a great amount of gratitude to Dr. Margo Brinton and Dr. David Wilson for their input, availability, helpful discussions, for sparing their invaluable time reviewing and providing insightful comments on my dissertation. I would also like to thank Dr. Margo Brinton for providing tremendous support during our collaboration.

I am deeply grateful to my parents, Shenggui Zhang and Yonglan Cai, for their unconditional love, help and faith in me at every stage of my personal and academic life. I am thankful to my elder brother Yu, for his love, encouragement and understanding. My deepest love goes to my husband Yongzheng, who has always provided me his unwavering support, encouraged me not to give up, stood by me during the highs and lows, and been tolerant toward my occasional foul mood. I also thank my baby-girl Ruohan who brought me the happiest time while writing my dissertation.

A special thanks to Alex Spring, Xiaoguang Qu and Subrata Mishra for thoughtful ideas and countless hours of discussions. I thank Xinjian (John), Dunay and Wei for their technical assistance; Chris Johnson, Yoshiko, Kate, Marina and Chris for their help, kindness and wonderful time working together. Thanks to Hsuan (Sharon) for her collaboration and input. Additionally, I would like to thank staff members, Will, Stephanie and Rita for administrative help over the

vi

years. Lastly, I would like to acknowledge the Brain and Behavior program for the financial support.

Thanks to all of you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The ability of one molecule to recognize another through binding interactions underlies all levels of biological processes in living cells. The binding partners include small molecule-protein, protein-protein, protein-nucleic acids, small molecule-nucleic acids and so on. These interactions are specific and of high affinity; aberrant macromolecular interactions often result in diseases, including cancer (*1*). Recent technical advances have provided unprecedented insight into these interactions on a structural and thermodynamic basis. This has led to a better understanding of the biology mediated by these interactions and how their alteration causes diseases, therefore contributing to pharmaceutical science.

1.1 Protein-Protein Interaction and its Modulation by Small Molecules

Protein-protein interactions (PPIs) are of major importance to biological systems. They are involved in almost every biological process such as the transport machinery across the various biological membranes, packaging of chromatin, signal transduction (*2*), regulation of gene expression, and so on.

Intramolecular intereactions of proteins are also a widely recognized phenomenon. For example, the interactions of the side-chains on the protein surface (*3-6*) are important for protein conformational changes which occur upon binding (*7-8*). These intramolecular rearrangements often facilitate tight binding. Another type of intramolecular interaction occurs between protein domains (*9-13*). For example, an interaction exists between the serine protease motif in the passenger domain and the β-helical domain of serine protease autotransporters of the family *Entero-*

bacteriaceae (SPATE). The interaction is important for both the proteolytic and adhesive activities of SPATE. Disrupting the intramolecular interaction by mutation can result in an inactive protein (*12*).

Conventional drug design has primarily focused on the inhibition of enzymes, since the ligand-binding site on the protein can be often utilized to guide the development of small molecule drugs. However, due to the importance of PPIs in many biological processes, small molecules capable of modulating PPIs have become more attractive in pharmaceutical science (*14-20*).

In the voltage gated potassium (Kv) channel, the intramolecular interaction between the voltage sensing domain and the pore domain is thought to be responsible for the opening and closing of channel (*21-24*). This interaction may also be responsible for inhibition of the channel by general anesthetics (*25*). Chapter 3 demonstrates the interaction of the S4-S5 linker peptide of the voltage sensing domain and S6 termini of the pore domain with butanol, respectively. In addition, a butanol mediated intramolecular interaction in the channel's active site is proposed.

1.2 Protein-RNA Interaction

Interactions between proteins and RNAs (Protein–RNA interactions; PRIs) are essential events that control a variety of cellular processes. They are the major player in post-transcriptional control of RNAs, such as: alternative splicing, RNA editing, polyadenylation, RNA export, mRNA stabilization, mRNA localization and translation (*26-27*). RNA binding proteins (RBPs) achieve highly specific recognition of their RNA targets by recognizing both their sequences and structures (*28*). However, studies on PRIs are more challenging then protein-DNA interactions due to the complexity of RNA structure and recognition modes.

It has been generally accepted that the RBPs recognize their target RNAs using several known binding motifs, such as: RNA Recognition Motif (RRM), also known as RNA-binding domain (RBD) or ribonucleoprotein domain (RNP), K Homology (KH) domain, Zinc finger (mainly C-x8-C-x5-C-x3-H type), double-stranded RNA binding domain (DS-RBD), RGG box, DEAD/DEAH box, Pumilio/FBF (PUF) domain, Piwi/Argonaute/Zwille (PAZ) domain and Sm domain. Some RBPs may contain one or multiple copies of the same RNA binding domain, others may contain several distinct domains (*29*). The presence of some RNA binding domains is suggestive of the protein function; for example, the existence of a DEAD/DEAH box is indicative of RNA helicase activity of the protein. However, other domains may only predict an RNA binding property of the protein such as RRMs.

RRM is one of the most abundant protein domains in eukaryotes. It also has been found in prokaryotes and viruses. Prokaryotic RRM proteins have only one single copy of the RRM domain while in eukaryotic protein, RRMs are often found as multiple copies (two to six). This protein domain is usually comprised of 90 amino acids and contains an eight amino acid consensus sequence RNP-1 and a less well conserved six residue consensus sequence RNP-2 (*30-31*). The typical RRM topology consists of four anti-parallel β-strands and two α-helices arranged in a βαββαβ fold. The β sheet is most commonly used to interact with single-stranded RNA. However, in the case of proteins containing multiple RRMs, the N- and C-terminal extensions of the domain and the inter-domain linker are also involved in RNA binding (*32-34*), reflecting the remarkable adaptability of this domain in order to achieve high affinity and specificity.

Chapter 4 studies a case of a RRM containing cellular protein, T cell restricted intracellular antigen-1 related protein (TIAR), which interacts with a West Nile virus RNA segment. The findings provide insight into the binding mode and the basis for developing potential approaches

to interfere with the protein-viral RNA interaction, therefore contributing to designing anti-viral drugs or therapies.

1.3   Thermodynamics of Macromolecular Interactions

The thermodynamics of a macromolecular interaction describe the heat interchange that arises from the changes of structure and freedom of motion of the binding partners during a binding event.

Molecular binding interactions are only favored when they are coupled with a negative Gibbs' free energy (ΔG) that consists of both enthalpic (ΔH) and entropic ($T\Delta S$) components. The enthalpic contribution reflects the specificity and strength of the interactions between both partners. This includes ionic, hydrogen bonding, electrostatic, and van der Waals interactions and so on. The entropic contribution is related to the disorder/structure motion of the overall system. Solvation effects, such as solvent re-organization, or the displacement of tightly bound water or ions upon ligand binding can also contribute significantly to the entropic term of the binding free energy (*35*).

Macromolecular binding interactions can be enthalpy-driven or entropy-driven processes. Understanding the forces driving partner recognition is very important for elucidation of the binding mechanism. In addition, because molecular activity is highly structure dependent, understanding the correlation of the thermodynamic parameters with the structures of interacting partners provides valuable information for medicinal chemistry.

Many experimental techniques have been developed to study various aspects of macromolecular binding thermodynamics. For example, X-ray crystallography depicts the static structures of ligand-receptor complexes. Thus, it provides very valuable information about the enthal-

pic contribution (hydrogen bonds, electrostatic interactions, etc). Nuclear magnetic resonance (NMR) (*36*) and fluorescence (*37*) spectroscopy are both powerful techniques for studying ligand-receptor dynamics, thereby helping the understanding of entropic components. In addition, NMR spectroscopy assesses the enthalpic contribution by providing structural information. Surface plasmon resonance (SPR) is also frequently used to study binding events and allows straightforward determination of equilibrium binding constants (*38*). Other widely used techniques in macromolecular association studies are calorimetric techniques, which include differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). The later technique can provide all the relevant thermodynamics parameters of the binding events in a single experiment (*39*).

1.4   Chapter Outline

In the following chapters, protein structure and inter- and intra-molecular interactions in various biological systems are studied by NMR combined with other physical or biological techniques. Chapter 2 shows an example of the use of NMR to explore the thermodynamics and kinetics of peptide bond *trans-cis* isomerization in various peptide systems. This work provides insight into the effect of solvents and flanking sequences as well as structural constraints on *trans-cis* isomerization. Chapter 3 details the structure and the orientation of the linker peptide of S4 and S5 (L45) of the Shaw2 voltage gated potassium channel in a membrane-like environment. A possible intra-molecular interaction between L45 and the C-terminus of S6 mediated by general anesthetic butanol is predicted. This study helps to understand the local conformation and the active site of the channel in response to anesthetics. In chapter 4, the specific interaction of 3′ terminus of the West Nile virus minus strand RNA and the cellular protein TIAR is investigated

to reveal the binding mode and the thermodynamics of this interaction. The results will shed light on the basis for the high affinity of the viral RNA-cellular protein, thus providing information for new therapeutic approaches for WNV infections. Chapter 5 presents a novel 2D NMR technique by which adenosine H2 and H8 resonances can be easily assigned for nucleic acid NMR structure determination.

1.5   References

1.   Garner, A. L., and Janda, K. D. (2011) Protein-protein interactions and cancer: targeting the central dogma., *Curr Top Med Chem. 11*, 258-280.

2.   Rodbell, M. (1980) The role of hormone receptors and GTP-regulatory proteins in membrane transduction, *Nature 284*, 17–22.

3.   Barlow, D. J., and Thornton, J. M. (1983) Ion-pairs in proteins., *J. Mol. Biol. 168*, 867-885.

4.   Heringa, J., and Argos, P. (1991) Side-chain clusters in protein structures and their role in protein folding., *J. Mol. Biol. 220*, 151-171.

5.   van den Burg, B., Dijkstra, B. W., Vriend, G., Van der Vinne, B., and Venema, G. E., V.G. . (1994) Protein stabilization by hydrophobic interactions at the surface, *Eur. J. Biochem. 220*, 981-985.

6.   Schindler, T., Perl, D., Graumann, P., Sieber, V., Marahiel, M. A., and Schmid, F. X. (1998) Surface-exposed phenylalanines in the RNP1/RNP2 motif stabilize the cold-shock protein CspB from Bacillus subtilis, *Proteins 30*, 401-406.

7.   Kranz, J. K., and Hall, K. B. (1999) RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface, *J. Mol. Biol. 285*, 215-231.

8.   Kranz, J. K., and Hall, K. B. (1998) RNA-binding mediates the local cooperativity between the β-sheet and the C-terminal tail of the human U1A RBD1 protein, *J. Mol. Biol. 275*, 465-481.

9.   Shin, H., Hsueh, Y. P., Yang, F. C., Kim, E., and Sheng, M. (2000) An Intramolecular Interaction between Src Homology 3 Domain and Guanylate Kinase-Like Domain Required for Channel Clustering by Postsynaptic Density-95/SAP90, *J Neurosci 20*, 3580-3587.

10.   Totaro, A., Paris, S., Asperti, C., and Curtis, I. d. (2007) Identification of an Intramolecular Interaction Important for the Regulation of GIT1 Functions, *Mol Biol Cell 18*, 5124-5138.

11.   Feiken, E., Etten, I. v., Gebbink, M. F. B. G., Moolenaar, W. H., and Zondag, G. C. M. (2000) Intramolecular Interactions between the Juxtamembrane Domain and Phosphatase Domains of Receptor Protein-tyrosine Phosphatase RPTPm, *J. Biol. Chem. 275*, 15350–15356.

12.   Tsang, C., Malik, H., Nassman, D., Huang, A., Tariq, F., Oelschlaeger, P., and Stathopoulos, C. (2010) Intramolecular Interactions between the Protease and Structural Domains Are Important for the Functions of Serine Protease Autotransporters, *Infect Immun 78*, 3335–3345.

13.   Calleja, V., Alcor, D., Laguerre, M., Park, J., Vojnovic, B., Hemmings, B. A., Downward, J., Parker, P. J., and Larijani, B. (2007) Intramolecular and Intermolecular Interactions of Protein Kinase B Define Its Activation In Vivo, *PLoS Biol 5*, e95.

14.   Toogood, P. L. (2002) Inhibition of protein-protein association by small molecules: approaches and progress, *J Med Chem 45*, 1543-1558.

15.   Sharma, S. K., Ramsey, T. M., and Bair, K. W. (2002) Protein-protein interactions: lessons learned, *Curr Med Chem: Anticancer Agents*, 311-330.

16.   Arkin, M. (2005) Protein-protein interactions and cancer: small molecules going in for the kill, *Curr Opin Chem Biol 9*, 317-324.

17.   Fry, D. C., and Vassilev, L. T. (2005) Targeting protein-protein interactions for cancer therapy, *J Mol Med 83*, 955-963.

18.   Yin, H., and Hamilton, A. D. (2005) Strategies for targeting protein-protein interactions with synthetic agents, *Angew Chem Int Ed Engl 44*, 4130-4163.

19.   Chene, P. (2006) Drugs targeting protein-protein interactions. , *Chem Med Chem 1*, 400-411.

20.   White, A. W., Westwell, A. D., and Brahemi, G. (2008) Protein–protein interactions as targets for small-molecule therapeutics in cancer, *Expert Rev Mol Med 10*, 1-14.

21.   Labro, A. J., Raes, A. L., Grottesi, A., Hoorick, D. V., Sansom, M. S. P., and Snyders, D. J. (2008) Kv channel gating requires a compatible S4-S5 linker and bottom part of S6, constrained by non-interacting residues, *J. Gen. Physiol. 132*, 667-680.

22.   Han, M., and Zhang, J. Z. H. (2008) Molecular dynamic simulation of the Kv1.2 voltage-gated potassium channel in open and closed state conformations, *J. Phys. Chem. B 112*, 16966-16974.

23.   Lee, S.-Y., Banerjee, A., and MacKinnon, R. (2009) Two separate interfaces between the voltage sensor and pore are required for the function of voltage-dependent K+ channels, *PLoS Biol. 7*, 0676-0686.

24.   Jensen, M. Ø., Jogini, V., Borhani, D. W., Leffler, A. E., Dror, R. O., and Shaw, D. E. (2012) Mechanism of voltage gating in potassium channels, *Science Vol. 336*, 229-233.

25.   Barber, A. F., Liang, Q., Amaral, C., Treptow, W., and Covarrubias, M. (2011) Molecular Mapping of General Anesthetic Sites in a voltage-Gated Ion Channel, *Biophys. J. 101*, 10.

26. Glisovic, T., Bachorik, J. L., Yong, J., and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation, *FEBS letters 582*, 1977-1986.

27. Matera, A. G., Terns, R. M., and Terns, M. P. (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs, *Nature Rev Mol Cell Biol 8*, 209-220.

28. Lee, M.-H., and Schedl, T. (200) RNA-binding proteins, *WormBook*, 1-13.

29. Stefl, R., Skrisovska, L., and Allain, F. H.-T. (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle, *EMBO Reports 6* 33-38.

30. Dreyfuss, G., Swanson, M. S., and Pinol-Roma, S. (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation, *Trends Biochem Sci 13*, 86-91.

31. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence, *Mol Cell Biol 6*, 2932–2943.

32. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J. 272*, 2118-2131.

33. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature 372*, 432–438.

34. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein, *Nature 398*, 579–585.

35. Holde, K. E. v., Johnson, W. C., and Ho, P. S. Principles of Physical Biochemistry.

36. Meyer, B., and Peters, T. (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors, *Angew Chem. Int. Ed. 42*, 864-890.

37. Weiss, S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy, *Nat Struct Biol. 7*, 724-729.

38. Alves, I. D., Park, C. K., and Hruby, V. J. (2005) Plasmon resonance methods in GPCR signaling and other membrane events, *Curr Protein Pept Sci. 6*, 293-312.

39. Perozzo, R., Folkers, G., and Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects, *J. Recept. Signal Transduct. Res. 24*, 1-52.

2 CHARACTERIZATION OF SECONDARY AMIDE PEPTIDE BOND ISOMERIZATION: THERMODYNAMICS AND KINETICS FROM 2D NMR SPECTROSCOPY

2.1 Preface

The initial assignments for Ac-IGGN-NH$_2$ peptide were obtained by Darrien James. This work was published in *Biopolymers* (2011) 95: 755-762.

2.2 Introduction

2.2.1 *The natural abundance of secondary and tertiary cis peptide bonds and their experimental detection*

The planar peptide bond can adopt two conformations, *cis* and *trans*, that are separated by a large rotational barrier of approximately 54-84 kJ/mol (*1-3*). Examination of existing protein structures reveals that with the exception of proline all other natural amino acids nearly exclusively occur as the *trans* isomer (99.95-99.97%) (*4-6*). In the *cis* isomer, the Cα positions (Cα and side chains) of preceding and following amino acids clash while the amide hydrogen offers much less steric repulsion rendering the *trans* conformation highly preferred. The tertiary amide peptide bond experiences clashes in both the *cis* and *trans* isomers which are energetically similar. Consequently, the proportion of the *cis* isomer in the population is significantly higher (5%-9%) (*4-6*).

Previous studies on the frequencies of *cis* peptide bonds vary widely (*4-7*). There is a discrepancy between the fraction of *cis* conformation detected from structural analysis and what

would be expected from the free energy difference obtained from model systems (*4*). For a ΔG

of 10 kJ/mol, the *cis* population is estimated to be 1.5% at 293 K, which is 50 times larger than

apparent in protein data bank (PDB) data (*6*). This discrepancy may be in part due to the difficul-

ty in detecting *cis* peptide bonds by lower resolution x-ray crystallography. Indeed, higher reso-

lution structures reveal a higher frequency of *cis* peptide bonds (*5-6*). Dominant *cis* peptide

bonds in proteins are reliably detected by NMR from characteristic Nuclear-Overhauser Effect

(NOE) proton-proton distances (*8-11*). However, if a *cis* peptide bond occurs only as a minor

species or in unstructured peptides, then the NOE's are not effective and the detection is clearly

more difficult. It is also noted that the NMR coupling constants $^3J_{CA-HN}$ and $^3J_{CA-CA}$ depend on

ω. However, the variations are very modest which limits their usefulness. In addition, ^{13}C and

^{15}N-enriched samples are required for such experiments (*12*).

The effect of flanking sequence on the stability of the *cis* conformation has been explored

in a number of studies focusing on the role of local interactions within neighboring residues (*13-

15*). Specifically, aromatic residues were observed to be able to stabilize *cis* conformations by

local interactions (*4, 14, 16-17*). In addition, protein folding constraints and distant residues can

also be a source of stabilization. This is supported by mutational studies, where a *cis* peptide

bond was retained even after the proline was replaced by a non-proline residue (*10, 18-22*).

Structural statistics have shown that *cis* peptide bonds are located preferably in the region with

strained backbones such as coils, bends and VI type turns (*5, 16, 23*). Interestingly, *cis* tertiary

amide bonds are generally located on the protein surface while cis secondary amide bonds are

frequently found in the protein interior (*16*).

2.2.2   Function of protein secondary cis peptide bonds and trans-cis isomerization in biologi-cal systems

Although rare, *cis* peptide bonds are functionally important. Secondary amide *cis* peptide bonds often occur near or at the active sites of proteins and are believed to engage in regulating biochemical properties and participate in metal binding (*19, 24-28*). For example, the Asp117-Asp118 *cis* peptide bond (*28*) is involved in Zn^{2+} binding at the active site of *Aeromonas Proteo-lytica* aminopeptidase. (PDB ID: 1amp). In addition, the strain inherent in a *cis* peptide bond can potentially serve as an energy reservoir which could be used to facilitate conformation changes or contribute to reactions (*4*). As a consequence of the large rotational barrier, the *trans-cis* isom-erization of both tertiary and secondary amide bonds is a slow process in aqueous solution (*5*). For tertiary amide bonds, the *cis* to *trans* rate constant is $1\text{-}5 \cdot 10^{-3}$ s^{-1} and the reverse process is slightly slower and depends on the flanking sequence (*29-30*). The *cis* to *trans* conversion is faster with 0.2-2.4 s^{-1} for secondary amide bonds while trans to *cis* reaction rate is $0.6\text{-}3.3 \cdot 10^{-3}$ s^{-1} (*3, 31*). As a result, the isomerization of both tertiary (*32-34*) and secondary amide (*35-38*) bonds contribute to rate-determining steps in protein folding, especially when the rarer *cis* form is pre-sent in the folded protein. This is the case for the refolding of RNase T_1 where a rapid initial folding step is followed by an extremely slow *trans* to *cis* conversion of the Tyr38-Ala39 bond (*35*). To facilitate folding, organisms possess both tertiary and secondary amide isomerases (*32, 39-40*).

2.2.3   Methods used for the detection of cis peptide bond and trans-cis isomerization

Several experimental (*18, 37, 41*) and theoretical (*42-46*) studies have explored the role of tertiary amide bond isomerization in regulating the kinetics and thermodynamics of protein

folding and function. Secondary amide *cis* peptide bonds have received less attention in part due to the fact that their low frequency and low abundance greatly hampers direct experimental studies in peptides and proteins and necessitates indirect approaches. For example, fluorescence was used to monitor the *cis-trans* conversion of Ala39 in the refolding of a RNase T_1 mutant (*35*). NMR was utilized early on to characterize *cis-trans* isomers of simple molecules containing N-substituted amide bonds (*47-48*). In addition, UV/visible and UV Raman spectroscopy have been used to study the *cis-trans* conversion of a GG dipeptide exploiting the different spectral properties of the two isomers (*49-50*). Small peptides have been investigated by the Fischer and Rabenstein groups using NMR in aqueous solution (*3, 31*). In these studies, the target peptide bonds were flanked by aromatic residues which permitted the use of the alanine methyl group as a convenient monitor to study the low abundance of *cis* conformation. In addition, recent theoretical calculations of simple dipeptide model systems have provided information on thermodynamic properties of *cis* conformers (*46, 51*). Encouraged by these studies we investigated peptides without aromatic side chains but with multiple peptide bonds to probe for all possible conformations.

In the present work, the quantum mechanical calculations on the dipeptide Ac-G-G-NH-Me were followed up by NMR experimental studies on tripeptides Ac-G-G-G-NH$_2$, Ac-I-G-G-NH$_2$, and an extended tetrapeptide Ac-I-G-G-N-NH$_2$. The formation of *cis* peptide bond conformers and the exchange process between two conformations were monitored with two-dimensional exchange spectroscopy (EXSY) (*52*). This permits the identification and study of minor species (<1%) in the presence of impurities which would obscure resonances of minor conformations in 1D spectra. Thermodynamic properties were calculated based on temperature dependent population changes. The *cis-trans* isomerization rates and activation energies were

obtained using Perrin & Dwyer's approach to extract exchange data from EXSY spectra (*52*). In addition, a cyclic peptide with the same sequence as the tetrapeptide IGGN was also investigated to explore whether the cyclization constraints would stabilize or inhibit *cis* conformer formation.

2.3   Materials and Methods

2.3.1 Chemicals

All linear peptides are acetylated on the N-terminus and amidated on the C-terminus. Both linear and cyclic peptides (purity > 95%) were purchased from GenScript USA Inc. (NJ, USA). Dimethyl sulfoxide-d6 (DMSO-d6), Acetone-d6 and CDCl$_3$ (isotopic purity > 99.6%) were purchased from Cambridge Isotope Laboratories (Andover, MA). N-methylacetamide (NMA, purity > 99%) was obtained from the Sigma Chemical Co. (St. Louis, MO).

2.3.2 NMR samples preparation

All linear peptides (20 mM) and the cyclic peptide (30 mM) were prepared in DMSO-d6. NMA samples (100 mM) were prepared in DMSO-d6, CDCl$_3$ and Aceton-d6.

2.3.3 Chemical shift quantum calculations

All *ab initio* Hartree Fock and density functional calculations were carried out using the Gaussian 03 package. Local minima for the Ac-G-G-NHMe peptides were optimized at the density functional B3LYP/6-31G(d) level. NMR calculations were carried out using HF/6-311+g (2d,p) (*53-54*). All calculations were carried out in the gas phase and DMSO was used as a reference.

2.3.4 NMR experiments

The NMR experiments were recorded on a Bruker Avance 600 MHz spectrometer with a 5 mm QXI probe. Two-dimensional phase sensitive NOESY/EXCSY spectra (6K×1K) were recorded with a series of mixing time from 15 ms to 250 ms, typically 16 scans were acquired. TOCSY (2K×400), COSY (2K×1400) and ROESY (1K×512) spectra were recorded with a sweep width of 10 ppm and 8 scans.

2.3.5 Exchange rate calculation

Impurities, satellite peaks and signal overlap limit the use 1D methods. Chemical exchange rates were determined from 2D Exchange Spectroscopy (EXSY) which provides a map of the exchanging species (*55*). Considering the greatly unequal population between the two exchanging species, as well as a similar spin-lattice relaxation time T_1 of the two species, (Supplementary Material Table S2.1) we found the approach outlined by Perrin & Dwyer (*52*) convenient. For a simple two site exchange, the total exchange rate k ($k = k_{c \to t} + k_{t \to c}$) is given by the equation:

$$k = \frac{1}{t_m} \ln \frac{r+1}{r-1} \tag{1}$$

The term r accounts for unequal populations and is defined as:

$$r = 4 X_c X_t \frac{I_c + I_t}{I_{c \to t} + I_{t \to c}} - (X_c - X_t)^2 \tag{2}$$

where I_c and I_t are the diagonal peak intensities of two exchangeable resonances in the EXCSY and $I_{c \to t}$ and $I_{t \to c}$ are the intensities of the exchange cross peaks, t_m is the mixing time and X_c and X_t are the mole fractions of the *cis* and *trans* forms.

2.3.6 Temperature coefficients of amide protons

The temperature coefficients $\Delta\delta/\Delta T$ of amide protons were obtained from the slopes of plots of chemical shift against temperature. Intra-molecular hydrogen bonds form when $\Delta\delta/\Delta T \leq$ 3 ppb/°C. Hydrogen bonds form with solvent when $\Delta\delta/\Delta T > 5$ ppb/°C.

2.4   Results and Discussion

2.4.1 Chemical shift calculations

To assess the impact of *cis* peptide bonds on the chemical shift of the amide protons, a dipeptide model was evaluated (Figure 2.1A). In this model, we focused on the peptide bond between two glycines. GtG and GcG describe two models with either a *trans* or *cis* peptide bond (ω_1); while cGG, GGc represent models with *cis* conformations adopted by the N (ω_0) and C (ω_2) terminal peptide-like bonds,respectively. The amide proton of G2 in the GcG model was upfield shifted by 3.31 ppm compared to the GtG control, indicating that the amide proton in *cis* conformation is more shielded (Table 2.1), which is consistent with previous experimental data (*3*). These calculations also predicted that a *cis* conformation can cause an upfield shift of an adjacent *trans* peptide amide proton as predicted for both cGG and GGc.
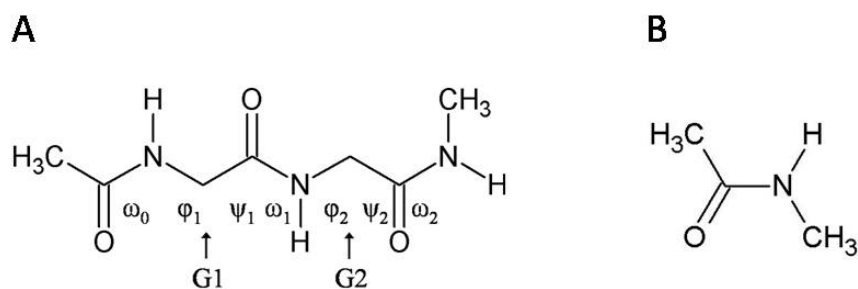
Figure 2.1 (A) Definition of torsion angles of the Ac-G-G-NH-Me dipeptide model. (B) N-methyl acetamide.

Table 2.1 Torsion angles and Gly2 H^N chemical shifts computed for Ac-G-G-NHMe.

	ω_0	φ_1	ψ_1	ω_1	φ_2	ψ_2	ω_2	G2$\delta_{H(N)}$ (ppm)
G*t*G	177.6	81.2	-67.0	-178.1	-80.2	67.5	-178.6	8.2
G*c*G	-176.9	-120.7	45.0	7.5	-98.7	-11.6	-177.3	4.9
*c*GG	11.1	-96.9	-2.1	-177.2	-81.2	65.1	-179.4	6.7
GG*c*	-175.6	-83.1	63.5	176.1	120.7	-65.7	2.0	7.7

2.4.2 NMR assignment of 1H resonances

Proton resonances of the peptides were assigned using standard two-dimensional TOCSY, NOESY and ROESY experiments. In addition to regular peptide amide protons which resonate around 8 ppm, we also observed several minor signals (7.3-7.8 ppm) at less than 1% of the main amide proton signals. At this low signal intensity, resonances of minor conformers were identified from EXSY and ROESY experiments which distinguish minor conformers from potential impurities. As shown in Figure 2.2, exchange peaks clearly correlate with the major and minor amide proton (H^N) resonances. Two minor resonances were observed to be associated with Gly1 and Gly2 of Ac-G-G-G-NH$_2$, Gly2 of Ac-I-G-G-NH$_2$ as well as Gly2 and Gly3 of Ac-I-G-G-N-NH$_2$ (Figure 2.3). It is highly unlikely that this indicates the presence of multiple simultaneous *cis* peptide conformations because this would be energetically strongly disfavored, especially in an unconstrained system. Therefore, only one of the two signals can be from the *cis* isomer, while the other is suspected to arise from the *trans* isomer that is affected by a neighboring *cis* peptide bond. This would be consistent with the chemical shift calculations. However, similar findings were also reported previously for peptides containing proline, which had a much higher percentage of *cis* isomers (*56-60*). Such an assignment would require that the kinetic and thermodynamic property of a minor *trans* form is the same as the neighboring *cis* conformer. As

shown in Table 2.2, the *cis* conformers (*c*) are characterized by a greater upfield shift which is consistent with reported data (*3*). In addition, the three bond coupling $^3J_{HN\alpha H}$ of the minor and major conformers are different, signifying a difference in the backbone torsion angle φ and allowing us to distinguish two minor conformers. Specifically, the *cis* forms exhibit a larger $^3J_{HN\alpha H}$ coupling compared to the major *trans* forms in agreement with previously reported data (*3*). However, the other minor conformers (*t*) had similar $^3J_{HN\alpha H}$ couplings to the major *trans* conformers (*T*) and were assigned as *trans* conformation coexisting with an adjacent *cis* peptide bond. For Ac-I1 in Ac-I-G-G-NH$_2$ and Ac-I-G-G-N-NH$_2$ only one minor conformer, the *cis* conformer, could be detected (Table 2.2).
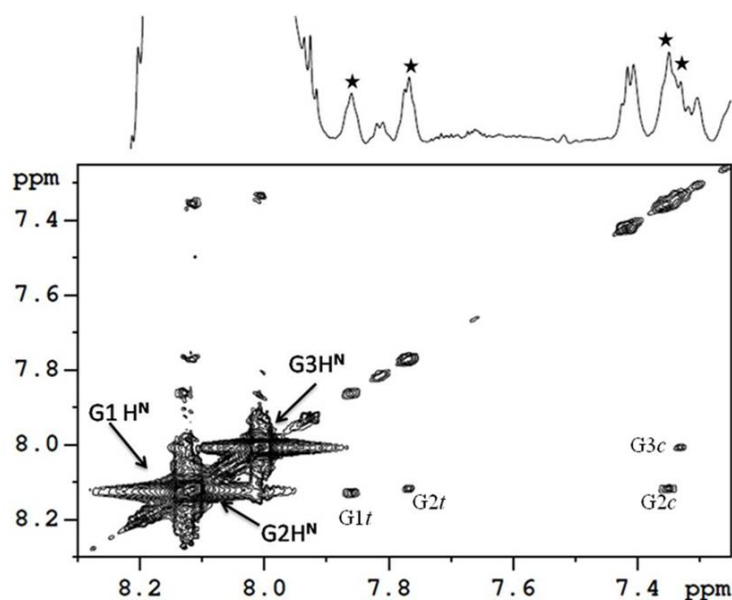
Figure 2.2 H^N region of 2D EXSY with corresponding 1D spectrum of Ac-GGG-NH$_2$ at 303 K. A data matrix of 6 K ×1 K was recorded with relaxation delay of 3 s, mixing time of 100 ms. Minor conformers are labeled in the 1D spectrum.
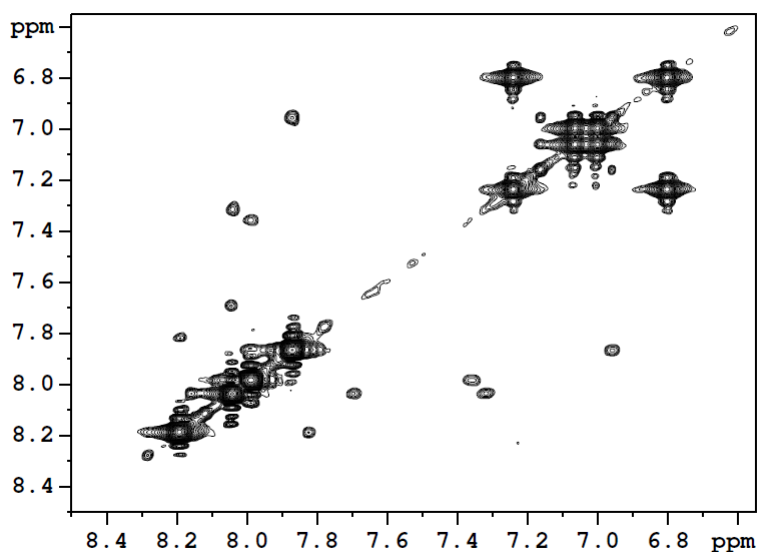
Figure 2.3 H^N region of 2D EXSY with corresponding 1D spectrum of Ac-I-G-G-N-NH$_2$ at 303 K. A data matrix of 6 K ×1 K was recorded with relaxation delay of 3 s and mixing time of 100 ms.

Table 2.2 Amide proton chemical shifts and $^3J_{HN\alpha H}$ couplings for major (*trans*) and minor peptide bond isomers recorded at 303K. T indicates major and m minor H^N resonances. (m1: minor *trans* conformer. m2: *cis* conformer) ΔCS represents the chemical shift difference between major and minor forms. In all cases, the minor forms resonate upfield from the major forms. OL: coupling data could not be extracted due to overlap. *: only one minor form could be detected.

Pept. Seq.		Gly1(Ile1)			Gly2			Gly3			Asn4	
		T	*t*	*c*	*T*	*t*	*c*	*T*	*t*	*c*	*T*	*c*
Ac-G-G-G-NH$_2$	ΔCS	-	0.27	1.0	-	0.35	0.76	-	*	0.69		
	$^3J_{NHaH}$	5.8	OL	6.7	5.6	OL	6.7	5.9	*	6.2		
Ac-I-G-G-NH$_2$	ΔCS	-	*	0.95	-	0.4	1.1	-	*	0.66		
	$^3J_{NHaH}$	6.5	*	9.5	5.6	OL	6.6	5.6	*	6.4		
Ac-I-G-G-N-NH$_2$	ΔCS	-	*	0.72	-	0.37	1.0	-	0.35	0.72	-	0.6
	$^3J_{NHaH}$	8.3	*	9.9	5.6	OL	OL	5.6	4.9	6.6	8.3	9.8

2.4.3 Temperature dependence of the cis-trans isomerization

An average population of 0.18% *cis* peptide bonds was found for Ac-G-G-G-NH$_2$ at 296 K (Table 2.3). Asn of Ac-I-G-G-N-NH$_2$ shows a slightly higher percentage of 0.25% at 293 K. Due to signal overlap, no reliable data were obtained for other residues of Ac-I-G-G-N-NH$_2$ and Ac-I-G-G-NH$_2$. The equilibrium constants results in an average free energy difference of 16 kJ/mol (Table 2.3). This value is in line with the previous results of computational and experimental study on N-methylacetamide (NMA) (*2, 61*). Similar enthalpy changes of isomerization were obtained for the minor *trans* conformer of G1, (referred to as Ac-*t*G1) and the *cis* conformer of Gly2 (G1-*c*G2), as well as minor *trans* conformer G1-*t*G2 and *cis* conformer G2-*c*G3. These data are consistent with the presence of the following conformers for this peptide: Ac-*t*G-*t*G-*t*G-NH$_2$ (~99.3%), Ac-*t*G-*c*G-*t*G-NH$_2$ (0.14%), Ac-*t*G-*t*G-*c*G-NH$_2$ (0.22%) and Ac-*c*G-*t*G-*t*G (~0.3%).

Table 2.3 Thermodynamic parameters for the peptide bonds *trans/cis* isomerization of Ac-G-G-G-NH$_2$ peptide at different temperatures. Data for *cis* Ac-*c*G1 could not be determined due to signal overlap.

	T (K)	% *cis*	$\Delta G_{t\text{-}c}$ (kJ/mol)	$\Delta H_{t\text{-}c}$ (kJ/mol)
Ac-*t*G1	296	0.13 ± 0.005	16.3 ± 0.1	
	303	0.12 ± 0.020	16.9 ± 0.4	-14.7 ± 2.4
	310	0.10 ± 0.008	17.8 ± 0.2	
G1-*t*G2	296	0.23 ± 0.015	14.9 ± 0.2	
	303	0.18 ± 0.004	15.9 ± 0.1	-18.6 ± 0.8
	310	0.16 ± 0.012	16.5 ± 0.2	
G1-*c*G2	296	0.15 ± 0.007	15.9 ± 0.3	
	303	0.16 ± 0.013	16.3 ± 0.3	-14.0 ± 3.9
	310	0.12 ± 0.021	17.3 ± 0.5	
G2-*c*G3	296	0.20 ± 0.012	15.3 ± 0.2	
	303	0.16 ± 0.013	16.2 ± 0.2	-16.4 ± 6.0
	310	0.15 ± 0.027	16.7 ± 0.5	

2.4.4 Solvent effects on the cis-trans isomerization

Surprisingly, the percentage of *cis* conformer detected decreased upon increasing the temperature signifying that the *trans* to *cis* conversion in DMSO is exothermic, with the possible exception of G3-*c*N4 in Ac-I-G-G-N-NH$_2$ for which we obtained an enthalpy of -0.4 kJ/mol. This is in contrast to previous studies which report that the *trans* to *cis* conversion is endothermic in aqueous solution (*3*). To further investigate this, NMA (Figure 2.1B) was examined in different solvents to explore solvent effects on peptide bond isomerization. We note that earlier studies concluded that both the isomerization equilibria and enthalpy of NMA are insensitive to solvent changes (*47*). However, we observed a higher percentage of *cis* conformers in solvents with larger dipole moments (Table 2.4). In addition, a higher amount of *cis* isomers at higher temperature results in a positive enthalpy of 1.8 kJ/mol, 7.1 kJ/mol and 15.3 kJ/mol in DMSO, acetone and H$_2$O, respectively. Therefore, the enthalpy of peptide bond isomerization is dependent on the solvent. This result also demonstrates that the thermochemistry of peptide bond isomerization depends on the local environment which will change upon protein folding.

Table 2.4 Thermodynamic parameters for the NMA *trans/cis* isomerization.

solvent	T (K)	$K_{t\text{-}c} \times 10^2$	$\Delta G_{t\text{-}c}$ (kJ/mol)	$\Delta H_{t\text{-}c}$ (kJ/mol)
DMSO	293	0.92	11.4	
	303	0.95	11.7	1.8
	313	0.96	12.1	
Acetone	283	0.85	11.2	
	293	0.98	11.3	7.1
	303	1.03	11.5	
H_2O	293	1.51	10.2	
	303	2.01	9.8	15.3
	313	2.25	9.9	

2.4.5 Kinetics of the cis-trans isomerization

Temperature-dependent exchange rates are shown in Table 2.5. At 296 K, the *trans* to *cis* rate constant is approximately $4\text{-}9 \cdot 10^{-3}\,\text{s}^{-1}$, while the *cis* to *trans* conversion is about 1000 times faster. Both forward and reverse rates are faster than those reported previously for peptide bonds in an aqueous system (*3, 31*). This might be due to solvent effects considering the limited hydrogen bond capability of DMSO compared to H_2O and also to the sequence context which could affect the kinetics of isomerization (discussed in the next section). The activation energies (E_a) of both forward and reverse processes were derived based on the Arrhenius equation. Similar kinetics and activation energies were obtained for Ac-*t*G1 and G1-*c*G2, as well as for G1-*t*G2 and G2-*c*G3 confirming our conclusion drawn from thermodynamic data.

Table 2.5 Isomerization rate constants and activation energies Ea $_{t-c}$ (trans-to-cis) and Ea$_{c-t}$ (*cis-to-trans*) for the Ac-G-G-G-NH$_2$ peptide. k_{t-c} (*trans-to-cis*) and k_{c-t} (cis-to-trans) were obtained from: $k_{t-c}/ k_{c-t} = $ K and $k_{t-c} + k_{c-t} = k$.

	T (K)	$k_{t-c} \times 10^3$ (s^{-1})	k_{c-t} (s^{-1})	Ea$_{t-c}$ (kJ/mol)	Ea$_{c-t}$ (kJ/mol)
	296	7.9 ± 1.0	6.0 ± 0.4		
Ac-*t*G1	303	15.2 ± 2.3	12.3 ± 5.0	86.5 ± 16.2	100.6 ± 23.5
	310	38.4 ± 1.3	38.0 ± 0.7		
	296	4.9 ± 0.7	2.3 ± 0.5		
G1-*t*G2	303	13.1 ± 1.3	7.3 ± 0.7	98.7 ± 6.1	113.6 ± 10.7
	310	29.9 ± 3.9	18.1 ± 2.1		
	296	8.4 ± 0.8	5.5 ± 1.1		
G1-*c*G2	303	21.7 ± 4.0	13.8 ± 0.7	81.3 ± 18.8	95.3 ± 3.7
	310	37.5 ± 6.5	31.4 ± 2.1		
	296	4.8 ± 0.8	2.4 ± 0.8		
G2-*c*G3	303	12.1 ± 2.8	7.4 ± 1.7	97.0 ± 2.1	113.1 ± 8.8
	310	28.3 ± 2.3	18.8 ± 5.8		
	293	1.6 ± 0.3	0.2 ± 0.09		
HN-CO in NMA	303	5.4 ± 1.5	0.6 ± 0.1	98.0 ± 10.9	96.2 ± 12.3
	313	22.3 ± 5.0	2.3 ± 0.3		

2.4.6 Effect of flanking sequence on isomerization

To address potential flanking sequence effects, Ac-I-G-G-NH$_2$ data were compared to Ac-G-G-G-NH$_2$ data by plotting exchange crosspeaks intensities of each peptide bond against mixing times (Figure 2.4). The exchange peaks between the *trans* and *cis* isomer of I1-G2 were of much lower intensity than for G1-*c*G2 of Ac-G-G-G-NH$_2$ at all mixing times investigated. These results demonstrate that the Ile side chain slows down the *trans-cis* isomerization com-pared to Gly at the same position and constitutes a specific sequence effect, which has not been addressed in previous studies. We also note that the *cis-trans* isomerization rate of NMA is slow-

er than our peptides in DMSO (Table 2.5), which may indicate that longer sequences enhance the

isomerization rate as was also suggested previously (*3*).
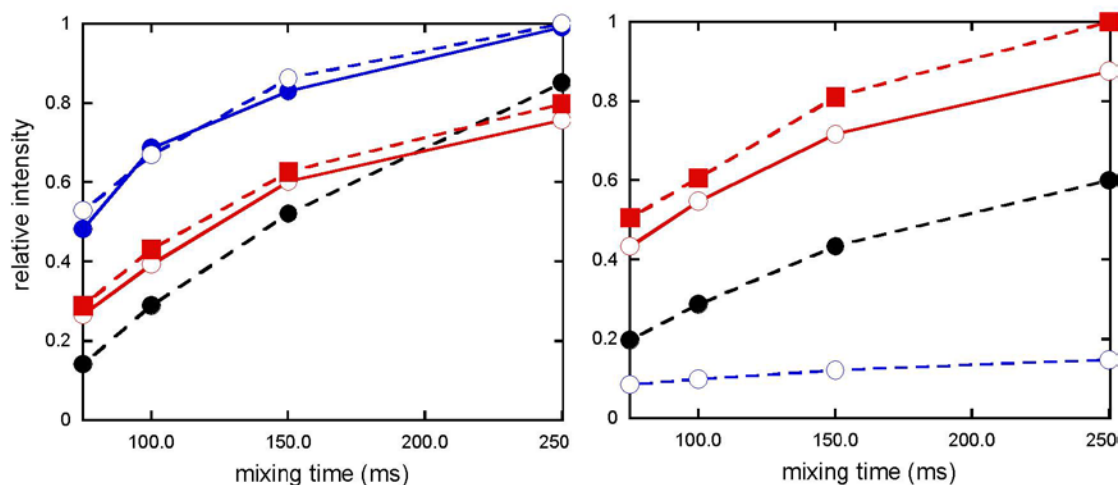
Figure 2.4 Build-up of exchange peaks between major and minor conformers. Intensities are scaled based on the highest intensity crosspeak. Ac-G-G-G-NH$_2$ (left panel): Dashed black line: exchange between major conformer Ac-*T*G1and minor conformers Ac-*c*G1. Blue line: exchange between Ac-*T*G1 and minor conformer Ac-*t*G1. Dashed blue line: exchange between G1-*T*G2 and G1-*c*G2. Red line: exchange between G1-*T*G2 and G1-*t*G2; dashed red line: exchange between G2-*T*G3 and G2-*c*G3. Ac-I-G-G-NH$_2$ (right panel): Dashed black line: exchange between major conformer Ac-*T*I1 and minor conformers Ac-*c*I1. Dashed blue line: exchange between I1-*T*G2 and I1-*c*G2. Red line: exchange between I1-*T*G2 and I1-*t*G2; dashed red line: exchange between G2-*T*G3 and G2-*c*G3. A dashed line indicates a *cis* conformer and solid line signifies (*t*). Residue 1 (circle), Residue 2 (open circle), Residue 3 (square).
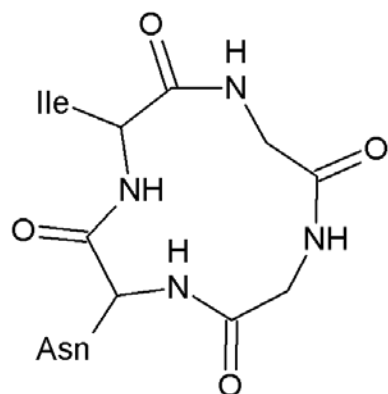
Figure 2.5 The IGGN cyclic peptide

2.4.7 Effect of constraints on isomerization

Constraining a peptide could conceptually encourage or inhibit *cis* peptide bond formation. We have also investigated the cyclic peptide I-G-G-N (Figure 2.5). Similar to its linear counterpart, we observed several minor conformers in the exchange spectra. However, the intensities of the exchange crosspeaks are much weaker at the same mixing times compared to the linear peptides, indicating that the *trans-cis* isomerization, while still possible, is significantly slower.

Further study on the amide proton temperature coefficient of cyclic IGGN in DMSO solutions revealed the existence of an intramolecular hydrogen bond involving H^N of Ile and Gly2 (Table 2.6). This may further conformationally lock the cyclic peptide and explain the slower *trans-cis* isomerization. In addition to slowing the kinetics of the isomerization, cyclization also limits the appearance of minor conformers. An approximate occurrence of 0.03% was obtained for the minor conformer of the G2-G3 peptide bond, which is 5-8 times less than that for the linear forms (Ac-G-G-G-NH_2 and Ac-I-G-G-N-NH_2). Interestingly, using two cysteine residues to

cyclize a tetrapeptide was previously reported to have little effect on the population of *cis* isomers compared to the linear form (*31*).

Table 2.6 Cyclic IGGN amide proton chemical shifts at different temperatures and their temperature coefficients in DMSO solutions.

Temperature(K)	Ile	Δδ	Gly2	Δδ	Gly3	Δδ	Asn	Δδ
298	7.667	0	7.743	0	8.668	0	8.33	0
303	7.657	0.01	7.737	0.006	8.636	0.032	8.309	0.021
308	7.647	0.02	7.733	0.01	8.6	0.068	8.285	0.045
313	7.638	0.029	7.731	0.012	8.565	0.103	8.261	0.069
318	7.63	0.037	7.728	0.015	8.531	0.137	8.238	0.092
Temp Coeff. (ppb/°C)		1.9		0.8		6.9		4.6

2.5 Conclusions

Peptide bond isomerization is a fundamental property of natural peptides and proteins and is of importance in protein folding/refolding and biochemical activities. We examined the peptide bond isomerization of several linear oligopeptides and a cyclic form. For linear peptides, 0.13% - 0.23% *cis* conformer was found in DMSO at 296K. In the cyclic form, the occurrence is even lower, demonstrating that backbone constraints affect the *cis* to *trans* isomer ratio, in this case negatively. The observation of two minor conformers associated with the same peptide bond signifies the presence of an additional conformer in addition to the expected *cis* conformer. Thermodynamic and kinetic properties reveal that the minor conformer with smaller upfield shift and similar $^3J_{HN\alpha H}$ to the major *trans* conformer is actually also a trans conformer that is impacted by the *cis* peptide bond on the C terminal side.

The *cis-trans* equilibrium and enthalpy were shown to be solvent dependent, suggesting that peptide bonds located in the interior of folded proteins experience a different thermochemis-

try from those exposed to an aqueous environment. As a result, the thermodynamics of peptide bond isomerization will change upon protein folding.

Also, higher isomerization rates were obtained in DMSO compared to aqueous solutions demonstrating solvent effects. Small residues also favor isomerization while neighboring bulky side chains may slow the *trans-cis* isomerization, as seen for the I1-*c*G2 of Ac-I-G-G-NH$_2$. In addition, the isomerization was slowed upon backbone cyclization, demonstrating that local structural elements also affect the kinetics of *trans-cis* isomerization.

2.6   References

1.   Drakenberg, T., and Forsén, S. (1970) The Barrier to Internal Rotation in Amides. I. Formamide, *J. Phys. Chem. 74(1)*, 1-7.
2.   Drakenberg, T., and Forsén, S. (1971) The barrier to internal rotation in monosubstituted amides, *J. Chem. Soc. D*, 1404 - 1405.
3.   Scherer, G., Kramer, M. L., Schutkowski, M., Reimer, U., and Fischer, G. (1998) Barriers to Rotation of Secondary Amide Peptide Bonds, *J. Am. Chem. Soc. 120*, 5568-5574.
4.   Jabs, A., Weiss, M. S., and Hilgenfeld, R. (1999) Non-proline Cis Peptide Bonds in Proteins, *J. Mol. Biol. 286*, 291-304.
5.   Stewart, D. E., Sarkar, A., and Wampler, J. E. (1990) Occurrence and Role of Cis Peptide Bonds in Protein Structure, *J. Mol. Biol. 214*, 253-260.
6.   Weiss, M. S., Jabs, A., and Hilgenfeld, R. (1998) Peptide Bonds revisited, *Nat. Struct. Biol. 5*, 676.
7.   MacArthur, M. W., and Thornton, J. M. (1991) Influence of proline residues on protein conformation, *J. Mol. Biol. 218*, 397-412.
8.   Billeter, M., Braun, W., and Wuthrich, K. (1982) Sequential Resonance Assignments in Protein 1H Nuclear Magnetic Resonance Spectra, *J Mol Biol 155*, 321-346.
9.   Wüthrich, K., Billeter, M., and Braun, W. (1984) Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances, *J Mol Biol 180*, 715-740.
10.   Xiong, Y., Juminaga, D., Swapna, G. V. T., Wedemeyer, W. J., Scheraga, H. A., and Montelione, G. T. (2000) Solution NMR evidence for a cis Tyr-Ala peptide group in the structure of [Pro93Ala] bovine pancreatic ribonuclease A, *Protein Sci. 9(2)*, 421-426.

11. Montelione, G. T., Hughes, P., Clardy, J., and Scheraga, H. A. (1986) Conformational Properties of 2,4-Methanoproline(2-Carboxy-2,4-methanopyrrolidine) in Peptides: Determination of Preferred Peptide Bond Conformation in Aqueous Solution by Proton Overhauser Measurements, *J Am Chem Soc 108*, 6765-6773.

12. Hu, J.-S., and Bax, A. (1997) Determination of - and ø1 Angles in Proteins from 13C-13C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond?, *J. Am. Chem. Soc. 119*, 6360-6368.

13. Lorenzen, S., Peters, B., Goede, A., Preissner, R., and Frömmel, C. (2005) Conservation of cis prolyl bonds in proteins during evolution, *Proteins: Structure, Function, and Bioinformatics 58*, 589-595.

14. Meng, H. Y., Thomas, K. M., Lee, A. E., and Zondlo, N. J. (2006) Effects of i and i+3 residue identity on Cis–Trans isomerism of the aromatici+1–prolyli+2 amide bond: Implications for type VI β-turn formation, *Biopolymers (Peptide Science), 84*, 192-204.

15. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects on peptidyl-prolyl cis/trans isomerization, *J. Mol. Biol. 279*, 449-460.

16. Pal, D., and Chakrabarti, P. (1999) Cis Peptide Bonds in Proteins: Residues Involved, their Conformations, Interactions and Locations, *J. Mol. Biol. 294*, 271-288.

17. Wu, W.-J., and Raleigh, D. P. (1998) Local Control of Peptide Conformation: Stabilization of cis Proline Peptide Bonds by Aromatic Proline Interactions, *Biopolymers 45*, 381-394.

18. Dodge, R. W., and Scheraga, H. A. (1996) Folding and Unfolding Kinetics of the Proline-to-Alanine Mutants of Bovine Pancreatic Ribonuclease A, *Biochemistry 35*, 1548-1559.

19. Guana, R.-J., Xianga, Y., Hea, X.-L., Wanga, C.-G., Wanga, M., Zhanga, Y., and Sundberg, E. J. (2004) Structural Mechanism Governing Cis and Trans Isomeric States and an Intramolecular Switch for Cis/Trans Isomerization of a Non-proline Peptide Bond Observed in Crystal Structures of Scorpion Toxins, *J. Mol. Biol. 341*, 1189-1204.

20. Birolo, L., Malashkevich, V. N., Capitani, G., Luca, F. D., Moretta, A., Jansonius, J. N., and Marino, G. (1999) Functional and Structural Analysis of cis-Proline Mutants of Escherichia coli Aspartate Aminotransferase, *Biochemistry 38*, 905-913.

21. Jin, L., Stec, B., and Kantrowitz, E. R. (2000) A cis-Proline to Alanine Mutant of E. coli Aspartate Transcarbamoylase: Kinetic Studies and Three-Dimensional Crystal Structures, *Biochemistry 39*, 8058-8066.

22. Wu, Y., and Matthews, C. R. (2002) A Cis-Prolyl Peptide Bond Isomerization Dominates the Folding of the Alpha Subunit of Trp Synthase, a TIM Barrel Protein, *J. Mol. Biol. 322*, 7-13.

23. Wathen, B., and Jia, Z. (2008) Local and Nonlocal Environments around Cis Peptides, *J. Proteome Res. 7*, 145-153.

24. Ludwig, M. L., Pattridge, K. A., Metzger, A. L., and Dixon, M. M. (1997) Control of Oxidation-Reduction Potentials in Flavodoxin from Clostridium beijerinckii: The Role of Conformation Changes, *Biochemistry 36*, 1259-1280.

25. Klabunde, T., Sharma, S., Telenti, A., Jr., W. R. J., and Sacchettini, J. C. (1998) Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing, *Nat Struct Biol. 5*, 31-36.

26. Aufhammer, S. W., Warkentin, E., Ermler, U., Hagemeier, C. H., Thauer, R. K., and Shima, S. (2005) Crystal structure of methylenetetrahydromethanopterin reductase (Mer) in complex with coenzyme F420: Architecture of the F420/FMN binding site of enzymes within the nonprolyl cis-peptide containing bacterial luciferase family, *Protein Sci. 14*, 1840-1849.

27. Héroux, A., White, E. L., Ross, L. J., Davis, R. L., and Borhani, D. W. (1999) Crystal Structure of Toxoplasma gondii Hypoxanthine-Guanine Phosphoribosyltransferase with XMP, Pyrophosphate, and Two Mg2+ Ions Bound: Insights into the Catalytic Mechanism, *Biochemistry 38*, 14495-14506.

28. Chevrier, B., Schalk, C., D'Orchymont, H., Rondeau, J.-M., Moras, D., and Tarnus, C. (1994) Crystal structure of Aeromonas proteolytica aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family, *Structure 2*, 283-291

29. Jacobson, J., Melander, W., Vaisnys, G., and Horvath, C. (1984) Kinetic Study on Cis-Trans Proline Isomerization by High-Performance Liquid Chromatography, *J. Phys. Chem. 88*, 4536-4542.

30. Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M., and Katchalski, E. (1960) The Configurational Changes of Poly-L-proline in Solution *J. Am. Chem. Soc. 82*, 5263-5279.

31. Nguyen, K., Iskandar, M., and Rabenstein, D. L. (2010) Kinetics and Equilibria of Cis/Trans Isomerization of Secondary Amide Peptide Bonds in Linear and Cyclic Peptides, *J. Phys. Chem. B 114* 3387–3392.

32. Schmida, F. X., Mayra, L. M., Muckea, M., and Schonbrunne, E. R. (1993) Prolyl Isomerases: Role in Protein Folding *Adv Protein Chem. 44*, 25-66.

33. Wedemeyer, W. J., Welker, E., and Scheraga, H. A. (2002) Proline Cis-Trans Isomerization and Protein Folding, *Biochemistry 41*, 14637-14644.

34. Dugave, C., and Demange, L. (2003) Cis-Trans Isomerization of Organic Molecules and Biomolecules:Implications and Applications, *Chem. Rev. 103*, 2475-2532.

35. Odefey, C., Mayr, L. M., and Schmid, F. X. (1995) Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding, *J. Mol. Biol. 245*, 69-78.

36. Svensson, A.-K. E., Jr, J. C. O. N., and Matthews, C. R. (2003) The coordination of the isomerization of a conserved non-prolyl cis peptide bond with the rate-limiting steps in the folding of dihydrofolate reductase, *J. Mol. Biol. 326*, 569-583.

37. Walkenhorst, W. F., Green, S. M., and Roder, H. (1997) Kinetic Evidence for Folding and Unfolding Intermediates in Staphylococcal Nuclease, *Biochemistry 36*, 5795-5805.

38. Pappenberger, G., Aygün, H., Engels, J. W., Reimer, U., Fischer, G., and Kiefhaber, T. (2001) Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics, *Nat. Struct. Biol. 8*, 452-458.

39. Schiene-Fischer, C., Habazettl, J., Schmid, F. X., and Fischer, G. (2002) The hsp70 chaperone DnaK is a secondary amide peptide bond cis-trans isomerase *Nat. Struct. Biol. 9*, 419-424.

40. Fischer, G. (1994) Peptidyl-Prolyl cis/trans Isomerases and Their Effectors, *Angewandte Chemie International Edition in English 33*, 1415-1436.

41. Cao, A., Welker, E., and Scheraga, H. A. (2001) Effect of Mutation of Proline 93 on Redox Unfolding/Folding of Bovine Pancreatic Ribonuclease A, *Biochemistry 40*, 8536-8541.

42. Kang, Y. K., and Park, H. S. (2005) Ab initio conformational study of N-acetyl-L-proline-N',N'-dimethylamide: a model for polyproline, *Biophys. Chem. 113*, 93-101.

43. Radzicka, A., Acheson, S. A., and Wolfenden, R. (1992) Cis/trans isomerization at proline: desolvation and its consequences for protein folding, *Bioorganic Chemistry 20*, 382-386.

44. Kang, Y. K. (2004) Ab Initio and DFT conformational study of proline dipeptide, *J Mol. Struct. (Theochem) 675*, 37-45.

45. Kang, Y. K., and Choi, H. Y. (2004) Cis-trans isomerization and puckering of proline residue, *Biophys. Chem. 111*, 135-142.

46. Kang, Y. K. (2006) Conformational Preference sof Non-Prolyl and Prolyl Residues, *J. Phys. Chem. B 110*, 21338-21348.

47. Radzicka, A., Pedersen, L., and Wolfenden, R. (1988) Influences of Solvent Water on Protein Folding: Free Engergies of Sovation of Cis and Trans Peptides Are Nearly Identical, *Biochemistry 27*, 4538-4541.

48. Goodman, M., Chen, F., and Lee, C.-Y. (1974) Conformational Aspects of Polypeptide Structure. XLV. Nuclear magnetic resonance Study of trans-Cis Isomerization in an N-methyl-L-alanine Derivative, *J. Am. Chem. Soc. 96*, 1479-1484.

49. Schiene-Fischer, C., and Fischer, G. (2001) Direct Measurement Indicates a Slow Cis/Trans Isomerization at the Secondary Amide Peptide Bond of Glycylglycine, *J. Am. Chem. Soc. 123*, 6227-6231.

50. Li, P., Chen, X. G., Shulin, E., and Asher, S. A. (1997) UV Resonance Raman Ground and Excited State Studies of Amide and Peptide Isomerization Dynamics, *J. Am. Chem. Soc. 119*, 1116-1120.

51. Gutierrez, L. J., Baldoni, H. A., and Enriz, R. D. (2009) Conformational and electronic study of cis-peptides (non-proline residues) occurring in natural proteins, *J. Mol. Struct. 934*, 103-111.

52. Perrin, C. L., and Dwyer, T. J. (1990) Application of Two-Dimensional NMR to Kinetics of Chemical Exchange, *Chem. Rev. 90*, 935-967.

53.   Wiberg, K. B., Hammer, J. D., Zilm, K. W., and Cheeseman, J. R. (1999) NMR Chemical Shifts. 3. A Comparison of Acetylene, Allene, and the Higher Cumulenes, *J. Org. Chem. 64*, 6394–6400.

54.   Cheeseman, J. R., Trucks, G. W., Keith, T. A., and Frisch, M. J. (1996) A comparison of models for calculating nuclear magnetic resonance shielding tensors, *J. Chem. Phys. 104*, 5497-5509.

55.   Jeener, J., Meier, B. H., Bachmann, P., and Emst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys. 71*, 4546.

56.   Grathwohl, C., and Wuthrich, K. (1976) NMR Studies of the Molecular Conformations in the Linear Oligopeptides H-(L-Ala)n-L-Pro-OH, *Biopolymers 15*, 2043-2057.

57.   Grathwohl, C., and Wuthrich, K. (1976) The X-Pro Peptide bond as an NMr Probe for Conformational Studies of Flexible Linear Peptides, *Biopolymers 15*, 2025-2041.

58.   Toma, F., Fermandjian, S., Low, M., and Kisfaludy, L. (1981) 13C-NMR Studies of ACTH: Assignment of Pesonances and Conformational Features, *Biopolymers 20*, 901-913.

59.   Toma, F., Fermandjian, S., Low, M., and Kisfaludy, L. (1978) *Biochim Biophys Acta 534*, 112-122.

60.   Stimson, E. R., Montelione, G. T., Meinwald, Y. C., Rudolph, R. K. E., and Scheraga, H. A. (1982) Equilibrium Ratio, *Biochemistry 21*, 5252-5262.

61.   Jorgensen, W. L., and Gao, J. (1988) Cis-Trans Energy Difference for the Peptide Bond in the Gas Phase and in Aqueous Solution, *J. Am. Chem. Soc. 110*, 4212-4216.

2.7   Supplementary Material

Table S1. Spin-lattice relaxation time T_1 of G3 amide proton in trans and cis conformation. *T*: *trans* conformer; *c*: *cis* conformer.

	G3(*T*)	G3(*c*)
T_1(ms)	762	750

3   STRUCTURAL AND FUNCTIONAL STUDIES OF THE KV CHANNEL S4-S5 LINKER PEPTIDES AND S6-C-TERMINUS PEPTIDES IN A MICELLULAR ENVIRONMENT

3.1   Preface

This project was collaboration with Dr. Covarrubias at Thomas Jefferson University, 2D NMR experiments including the assessment of the peptide proton relaxation properties using the paramagnetic reagent Gd-DTPA-BMA as well as the initial calculation of the 3D structure of S4-S5 linker peptide in micelles using the DYANA package were carried out by Xiaoguang Qu. This work was published in *Biochimica et Biophysica Acta, Biomembranes* (2013) *1828*: 595-601.

3.2   Overview and Background

3.2.1   *Biological importance of voltage-gated potassium (Kv) channel*

Voltage-gated potassium (Kv) channels comprise a large family of potassium channels, which are essential components of the nervous system. They occur in nerve cell (neuron) membranes and are responsible for the generation and propagation of nerve impulses (or action potentials). Open Kv channels stabilize the resting potential of membranes and dampen the effectiveness of excitatory inputs to a cell; closure of Kv channels, on the other hand, enhances the excitability. More specifically, the membranes of all nerve cells have a potential difference (resting membrane potential) across them when they are not being stimulated or conducting impulses. The interior of the cell is negative with respect to the exterior due to an unequal distribution of

two ions, sodium (Na$^{+)}$ and potassium (K^{+}) on the two sides of a nerve cell membrane. In neurons, stimuli (such as touch, sound, light, etc) can alter this potential difference by opening sodium channels in the membrane and letting sodium ions flow into the cell, therefore reducing the voltage across the membrane and completely depolarizing the membrane. This opens more voltage-gated ion channels in the adjacent membrane, and results in a wave of depolarization (the action potential) along the cell. As the action potential nears its peak, the sodium gates close, and potassium gates open, allowing ions to flow out of the cell to restore the normal potential of the membrane (*1*).

Utilizing a combination of genomic and biophysical techniques, a growing number of discoveries have linked Kv channel mutations with various diseases of heart, kidney, pancreas and central nervous system, such as myokymia, which was the first mutated mammalian Kv channel human disease (*2-6*).

3.2.2 Structure-functional feature of Kv channel

Kv channels share a similar structural topology. They are tetramers formed from four identical subunits each having six transmembrane-helices (S1-S6). The first four helices (S1-S4) form a voltage sensing domain (VSD), that detects voltage difference across cell membranes via charged amino acids. The S5 and S6 segments of all four subunits form a pore domain where the S5-S6 linkers (P-loop) act as an extracellular selectivity filter while the S6 C-termini of all subunits make up the intracellular pore portal (*7*). Numerous studies suggest that opening and closing of the Kv channel are carried out by coupling the conformational changes of the VSD and S6 C-termini through the S4-S5 linker (*8-18*). This necessitates conformational malleability and cor-

rect positioning of the S4-S5 linker in order to transmit the coupling between the VSD and S6 C-terminus.

3.2.3 Drug binding sties on Kv channel

Their involvement in vital biological processes and disease states render Kv channels important targets for drug therapy (*5, 19*). Numerous small molecules and drugs have been found to interact with Kv channels. On the extracellular side, the P-loop and adjacent residues of S5 and S6 are binding sites for toxins and channel blockers (*20-21*). Binding sites for small molecules are also located on the intracellular side of S5 and S6 (*22-23*).

3.2.4 Structure-function correlation of Shaw2 Kv channel in 1-alkanol inhibition

The *Drosophila* Shaw2 is a neuronal Kv channel that is closely related to the mammalian Kv3 channels (*24*). It is reported that the Shaw2 channel is selectively inhibited by 1-alkanols and halothane at pharmacologically relevant concentrations (*25-27*). The action of the inhibitors is consistent with binding to an intracellular site and the stabilization of the channel's close state (*28*). The energetics and kinetics of this inhibition have been investigated by applying a combination of biochemical, electrophysiological and structural methods (*29-30*).

It also has been demonstrated that the S4-S5 linker of Shaw2 is required for 1-alkanol inhibition. Transplanting just a thirteen amino acid segment from the Shaw2 S4-S5 linker into Kv3.4 causes this modified human channel to become 1-alkanol responsive (*26*). The Shaw2 S4-S5 linker peptide (L45) was shown later on to readily adopt an α-helical structure in solution and in the membrane environment (phospholipid micelles), while the corresponding Kv3.4 peptide does not (*29-30*). This links the 1-alkanol response to the α-helical propensity of L45.

CD spectra revealed that in the absence of TFE, L45 is only partially structured. Upon increasing the TFE concentration, the peptide gradually assumes an α-helical conformation (*30*). The three distinct sets of TOCSY cross peaks between H_α and side chain H_β protons for both Gln320 and Phe322 in 44 ms TOCSY spectrum reveals the coexistence of 3 different conformations (α helix, β sheet and random coil) of Shaw2 L45 in the presence of 20% TFE (*31*).
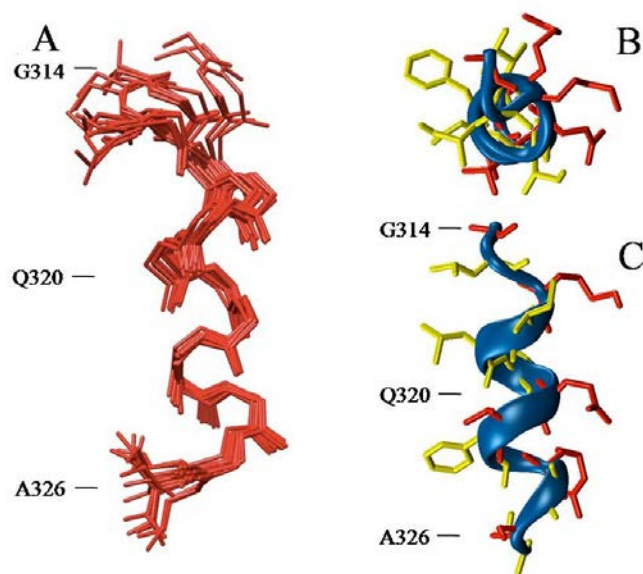
Figure 3.1 Shaw2 L45 NMR structure in DPC micelles generated by DYANA. A total of 40 structures were generated. The mean backbone RMSD is 0.96±0.33 Å, and 1.44±0.27 Å for the mean heavy atom. The resulting structures were visualized with Pymol. A): The 20 conformers with the lowest target function (from $3.13*10^{-2}$ to $9.07*10^{-2}$). B): Top view of the lowest energy conformer in cartoon mode. Hydrophobic residues are shown in yellow and polar side chains are in red. C): Side view

The NMR structure of the Shaw2 L45 in 30 mM DPC micelles was calculated with the DYANA macro anneal (*32*) using the distance and torsion angle information extracted from the NMR spectra (Supplementary Materials, Table S3.2). It is revealed that a regular α-helix is formed from residues Ile317 to Ser325 while the helix is less ordered at the C-terminus and is disrupted at the N-terminus (Figure 3.1). This is either due to dynamics or a consequence of few-

er constraints at that location. In contrast to the structure of L45 in 20% TFE, no evidence NMR

suggests the presence of more than one conformation of L45 in DPC micelles (Figure 3.2).
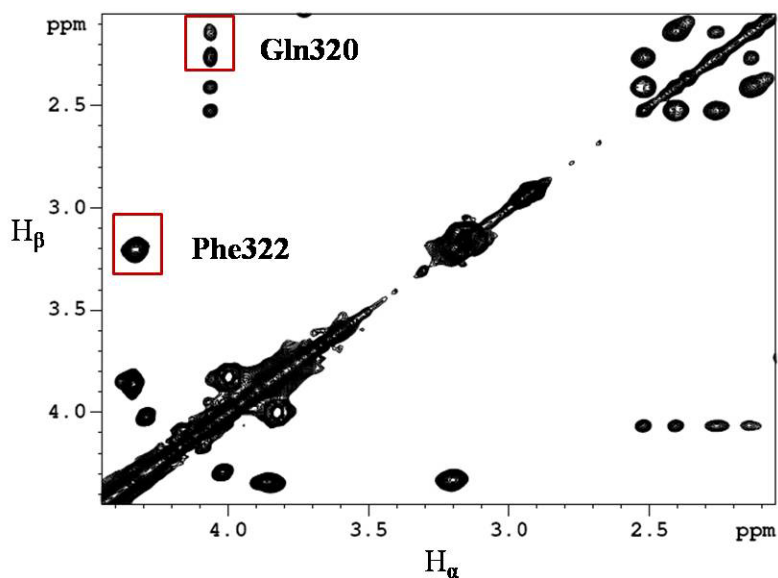
Figure 3.2 H_α-H_β region TOCSY spectrum (45 ms mixing time) of 1 mM Shaw2 L45 in 30 mM DPC,10 mM sodium phosphate, pH* = 5.8 at 302 K. The box marks the H_α-H_β cross peaks of Gln320 and Phe322. The correlations of Gln320 α proton to δ and γ protons are also visible below the box.

3.2.5 Solvent exposure and orientation of Shaw2 L45 in DPC micelles

The solvent accessibility of Shaw2 L45 residues in DPC micelles was probed by taking

advantage of the differences in residue relaxation properties when paramagnetic reagents were

present. These reagents effectively relax exposed NMR active nuclei, while nuclei that are buried

in a micelle are protected. In the presence of the paramagnetic reagent Gd-DTPA-BMA, Gly314,

Lys316, Ile317, Ile319, Gln320 and Arg323 showed a large decrease in the T_1 value, indicative

of solvent accessibility. In contrast, Leu315, Leu318, Phe322 and Ala326 were only marginally

affected, suggesting they are buried in the micelle (*31*) (Supplementary Materials Figure S3.1). The solvent accessibility was further explored using the paramagnetic attenuation method presented previously (*33-34*). As expected the H_α of charged residues Lys316 and Arg323 are highly accessible while the H_α of Leu315 is considered protected. All other residues with $1.39 \geq Ai \geq 0.61$ are intermediate (*31*) (Supplementary Materials Figure S3.1).

As described previously (*35*), the orientation of a regular α-helical peptide in the micelles can be defined by its immersion depth A, rotation ρ and tilt angle τ. The dependence of PRE on A, ρ and τ was simulated (Supplementary Materials Figure S3.2). Fitting the PRE data from residues Leu318 to Ser325 and the residue numbers n gives the immersion depth A of 7.0 ± 1.5 Å for the H_α proton of Leu318, a rotation angle ρ of $206 \pm 4.0°$ and a helix tilt angle τ of $6.9 \pm 2.8°$. Considering the surface curvature of a DPC micelle with a diameter of ~ 40Å (*36*), we can conclude that L45 is bound to the surface and nearly parallel to it as shown in Figure 3.4. According to the orientation and residue accessibility data, Leu315, Leu318, Phe322 and Ala326, are buried inside the micelle, while the charged residues, Lys316 and Arg323, and the polar residue Gln320 face towards the solution and Thr321 and Ser325 are at the interface.
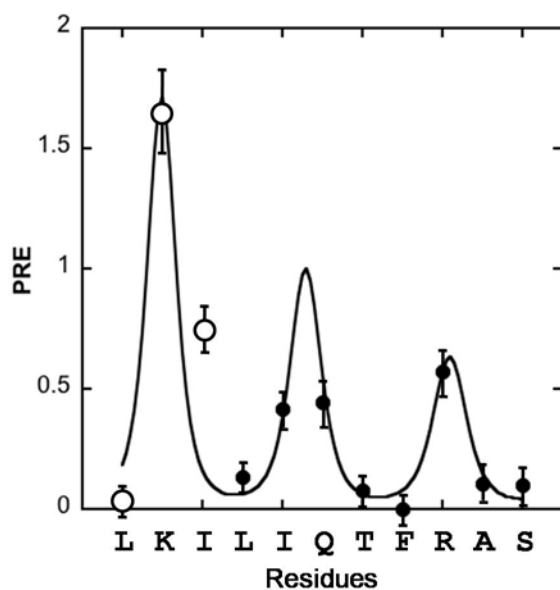
Figure 3.3 PRE analysis of Shaw2 L45 in DPC micelles. Data for residues L318 to S325 were used for fitting equation 4. Errors were derived from the nonlinear least squares fit.
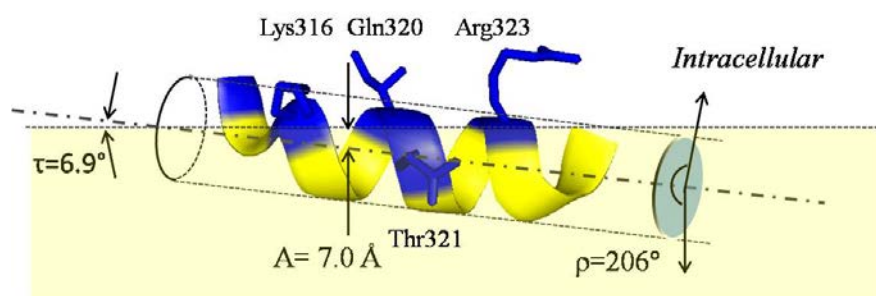
Figure 3.4 Orientation of Shaw2 L45 in micelles. The micelle is represented by the light-yellow shaded region. A, immersion depth; ρ, rotation; τ, helix tilt angle. Blue indicates polar peptide residues while yellow signifies hydrophobic amino acids.

3.2.6 1-Alkanol inhibition of Shaw2 Kv channel also requires S6 and the interaction of L45 and S6c

In addition to L45, additional elements involved in 1-alkanol binding were identified by alanine scanning to be S5 and S6, as demonstrated by the observation that mutating the second Pro in the PVP motif of S6 resulted in suppression of the 1-alkanol inhibition (*37*). This was attributed to the destabilization of the closed state. Furthermore, data from a recent study supports the presence of putative 1-alkanol and halothane binding pockets in interfaces involving the S4-S5 linker, S5 and S6 (*38*).

Despite numerous functional studies, the precise molecular interactions governing 1-alkanol binding and the mechanism of channel inhibition are not fully understood. In this study, we further investigated the participation of the S4-S5 linker and S6 C-terminus in the 1-alkanol modulation of Shaw2 channels by focusing on the structure of the S6 C-terminal peptide (S6c) in DPC micelles and the potential binding sites of 1-alkanols on both micelle bound peptides.

3.3   Materials

3.3.1   Chemicals

The Shaw2 S4-S5 linker peptide (L45, GLKILIQTFRASA) and S6 C-terminal peptides (S6c, VIVSNFAMYYSHTQ) derived from the voltage-gated *Shaw* potassium channels were purchased from Biopeptide Co., Inc. (San Diego, CA). Deuterated dodecylphosphocholine, DPC-d_{38} (D, 98%) was purchased from CDN isotopes Inc. (Quebec, Canada). 2, 2, 2-trifluoroethanol (TFE, 99.5%) was from Aldrich. TFE-d3 (D, 99.5%) and D_2O (D, 99.9%) were from Cambridge Isotope Laboratories (Andover, MA). 1-butanol (99%) was from Fisher Scientific (Fair Lawn,

NJ). 1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine (DMPC, 99%) and 1,2-Dihexanoyl-*sn*-Glycero-3-Phosphocholine (DHPC, 99%) were from Avanti polar lipid, Inc. (Alabaster, AL). Tetradecyltrimethylammonium bromide (TTAB, 99.5%) was from Sigma-Aldrich Co. (St. Louis, MO).

3.3.2 DHPC/DMPC bicelles preparation

Bicelles was prepared by mixing DMPC/DHPC (ratio: 3:1) and TTAB in 10 mM sodium phosphate buffer with pH 6.5 to a total lipid concentration of 15% w/v and TTAB of 2.4 mM. The mixture was then homogenized by 2 cooling-heating cycles (4°C - room temperature - 38 °C – room temperature - 4°C - room temperature - 38 °C) with vortexing at room temperature. The observation of clear, transparent and fluid solution at low temperature and clear, transparent but viscous solution at 38 °C, but white and milky at 25 to 30 °C.

3.4   Methods

3.4.1 CD spectroscopy

CD samples were prepared by dissolving peptides in 5 mM sodium phosphate buffer, pH 6.0, to a final concentration of 50 uM unless described otherwise. A JA-810 spectropolarimeter (Jasco, Tokyo, Japan) and 5 mm sample cuvette were used to record all CD spectra at room temperature. Each spectrum was the average of four scans. The resulting spectra were deconvoluted using CDPro (*39-40*).

3.4.2   *NMR Spectroscopy*

The micellar NMR samples were prepared by co-dissolving peptides, typically 1 mM with 30 mM DPC-d_{38} in 10 mM sodium phosphate buffer (pH 5.8, unless described otherwise) containing 10% D_2O. For D_2O experiments, the samples were lyophilized and resuspended in 100% D_2O.

All NMR spectra were collected on 500 and 600 MHz Bruker Avance systems using a 5 mm triple resonances (QXI) Z-gradient probe head or TXI cryoprobe (Bruker). For assignments and structure determination, 1D spectra were recorded using presaturation or jump-and-return pulse sequences to suppress the solvent (water) signal (*41*). 2D NMR experiments: TOCSY, NOESY, and natural abundance ^{13}C-^1H-HSQC were recorded with presaturation as appropriate and using time proportional phase increment (TPPI) for quadrature detection in F1. The mixing times were set to 44 ms for TOCSY with data matrix of 2K x 512 and 32 scans) and 75 and 400 ms for NOESY experiments with data matrix of 2K x 512 and 32 scans, respectively. NMR spectra were assigned using Sparky (*42*) following standard methods (*43*).

Diffusion measurements were carried out using Diffusion-Ordered Spectroscopy (DOSY) (*44*). A 1D setup (stegp1s1d) was run prior to DOSY experiments (stegp1s) to optimize parameters: diffusion time Δ (*d20*), the spoil gradient (*p19*) and gradient length δ (*p30*). All gradients were applied as half sine shapes. Δ (*d20*) of 200 ms and *p19* of 1.1ms (-9.8G/cm) were chosen for all 2D DOSY acquisitions, δ (p30) was set as 2.2ms and 3.5ms for peptide without and with micelles, respectively. The actual strength of the half sine shaped gradients was varied from 0.735 to 34.9G/cm. Typically, 256 scans and a data matrix of 8K x 16 were recorded. The spectra were processed using the Bruker au program (dosy2d) and data were fitted to the equation below with XWINNMR T1/T2 software:

$$I = I_0 * e^{-D(\gamma\delta G)^2(\Delta-\delta/3)} \qquad (1)$$

Where I_0 is unattenuated signal intensity, D is the diffusion coefficient, γ is the gyromagnetic ratio of the observed nucleus, in this case $\gamma(^1H) = 4.258*10^3$ Hz/G.

3.5   Results and Discussion

3.5.1   Structural features of S6c in TFE, DPC micelles and bicelles

Unlike L45, the CD spectra of S6c upon TFE titration display a two-state transition from random coil to α-helix (Figure 3.5). It also requires more TFE (~ 80%) to obtain 50% α-helical conformation than is needed for L45, signifying a lower α-helical propensity of S6c. In both DPC micelles and DMPC/DHPC bicelles, S6c adopts a partial α-helix (Figure 3.6B and Figure 3.7A). Specifically, the helical content plateaus at just 39.4% at 20 mM DPC, which is much lower than the 79.3% α-helical content of L45 at the same DPC concentration (Figure 3.6A and 3.6B). The lower α-helicity is also manifest in the peptide dynamics in micelles. This is supported by the line width and the small chemical dispersion of the peptide resonances (Figure 3.8). In addition, and in contrast to L45, no α-helical characteristics ($^3J_{NH-H\alpha}$, NOESY connectivity) were observed. Therefore, we conclude that this peptide adopts a more disordered and dynamic structure.
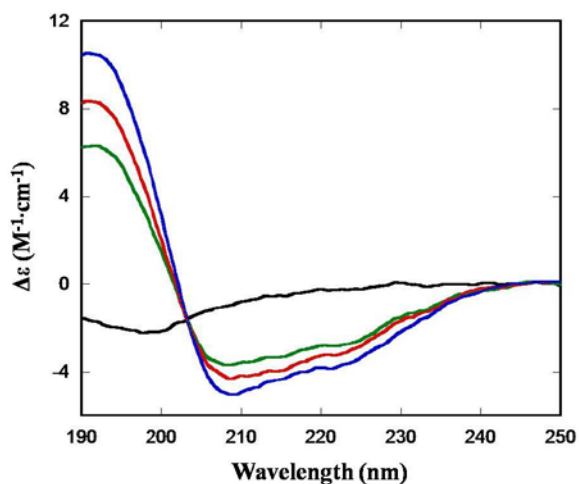
Figure 3.5 CD spectra of S6c in the presence of increasing TFE concentrations. (black: 0% TFE, green: 30%, red 50% and blue 80% TFE)
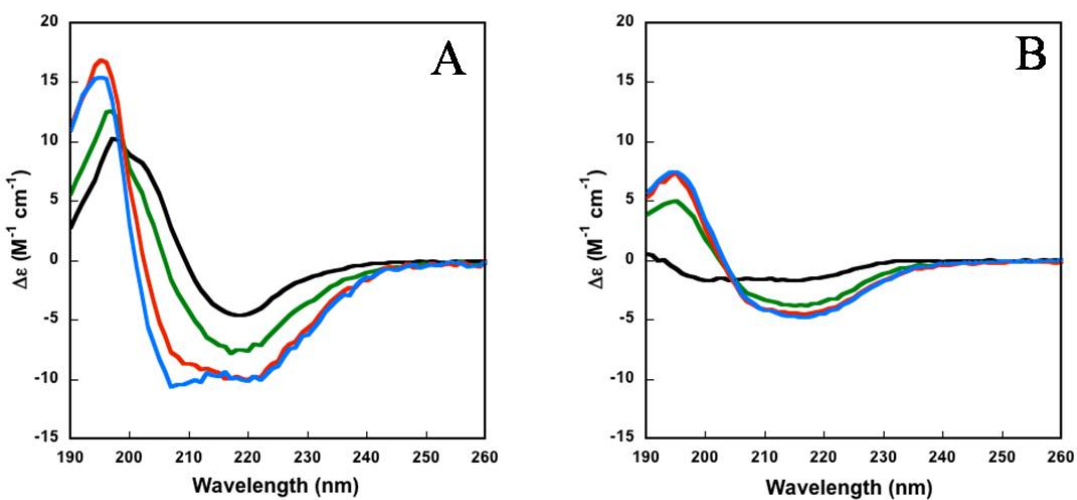
Figure 3.6 CD spectra of A): L45 and B): S6c in the presence of increasing DPC concentrations. (black: 1.5 mM, green 3.0 mM, red 20 mM and blue 30 mM DPC)
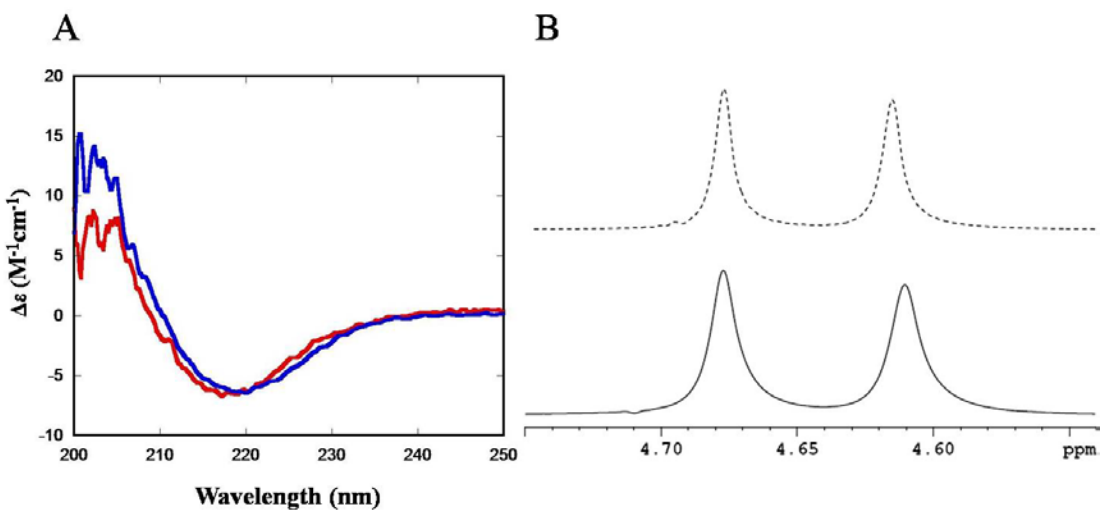
Figure 3.7 A): CD spectra of L45 (red) and S6c (blue) in the presence of 4% DMPC/DHPC (molar ratio, 3:1), 2.4 mM TTAB, 10 mM sodium phosphate (pH 6.5). B): 92.1 MHz ^{2}H NMR spectra of the 4% DMPC/DHPC bicelles sample with 15% D_2O at 303K. Solid line, freshly made bicelles; dotted line, the spectrum after 18 h. The observed splitting is indicative of stable bicelle formation.
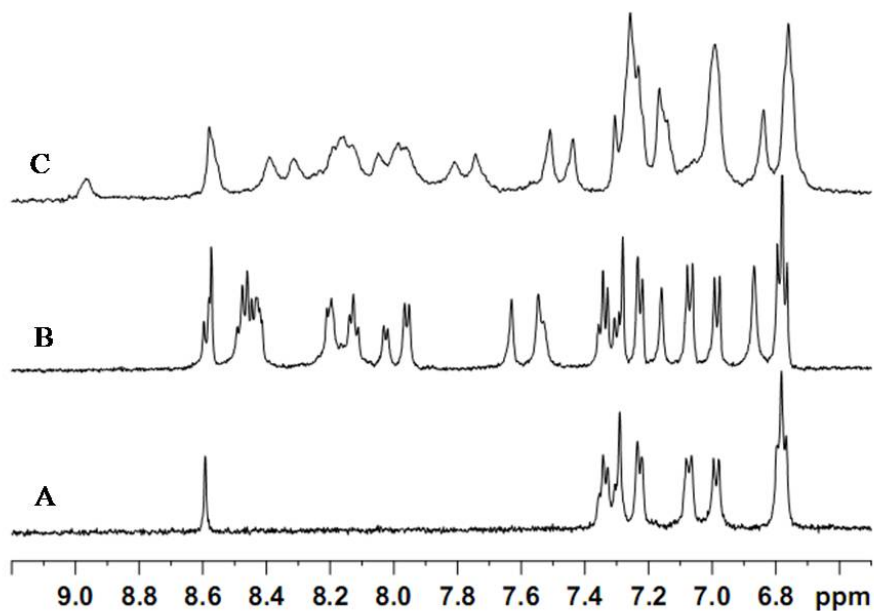
Figure 3.8 1D ^{1}H NMR spectra of the H_N region of S6c: A) in D_2O; B) in H_2O; C) with 30 mM DPC in H_2O. Sample is in 10 mM sodium phosphate, pH 4.

3.5.2   Potential alkanol binding sites on Shaw2 L45 and S6c

Upon the addition of a small amount of 1-butanol (1-5 mM) to Shaw2 L45 in micelles we observed H_N chemical shift changes (~ 0.03 ppm) for Gln320, Thr321, Phe322 and Arg323 (Figure 3.9). This appears to be specific for 1-butanol since methanol did not produce any chemical shift changes in the range tested (up to 10 mM, data not shown). A related observation is that the addition of 10% TFE to Shaw2 L45 in micelles produced an additional NOE contact between H_N of Gln320 and Thr321. This implies that Gln320 and Thr321 are sensing the presence of 1-butanol and TFE and may act as a site of interaction for such molecules. Similarly, C_α and H_α chemical shift changes induced by high TFE concentrations again map to Thr321. The local chemical shift changes are quite small and this is consistent with only minor local structural changes but would not support helix disruption or unwinding. These results are in good agreement with a recent mutagenesis study of L45 where mutating Thr321 to alanine had a negative effect on the modulation of the channel by 1-butanol (*38*).
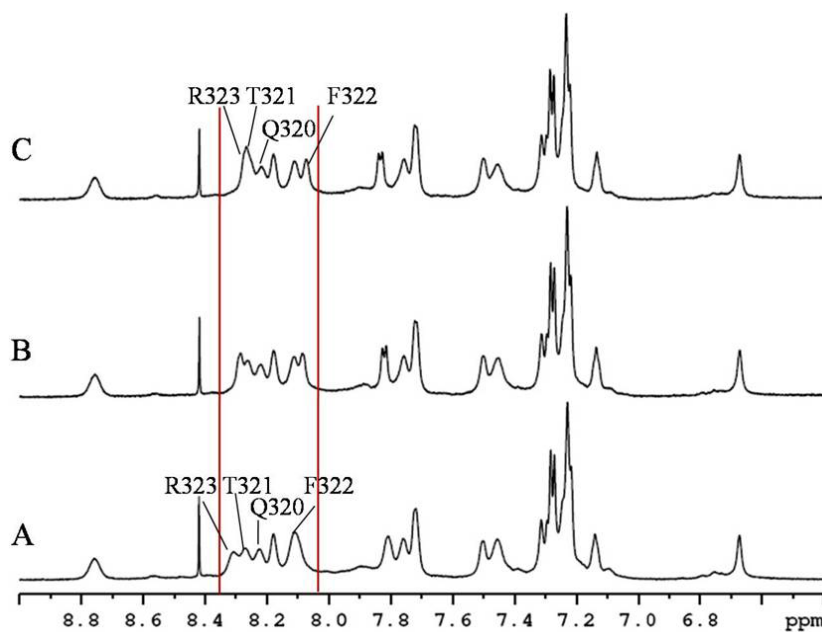
Figure 3.9 Change of the 1D NMR H_N region of L45 in micelles upon 1-butanol titration at 303 K. A, L45 in micelles. B, L45 peptide in micelles with 1 mM 1-butanol. C, L45 peptide in micelles with 5 mM 1-butanol.

1-butanol titration of S6c in DPC micelles also revealed a small chemical shift change (~ 0.02 ppm) of Thr423 and Gln424 amide protons in the presence of just 2 mM 1-butanol indicating that these residues may interact with 1-butanol while all other amide protons were insensitive (Figure 3.10).
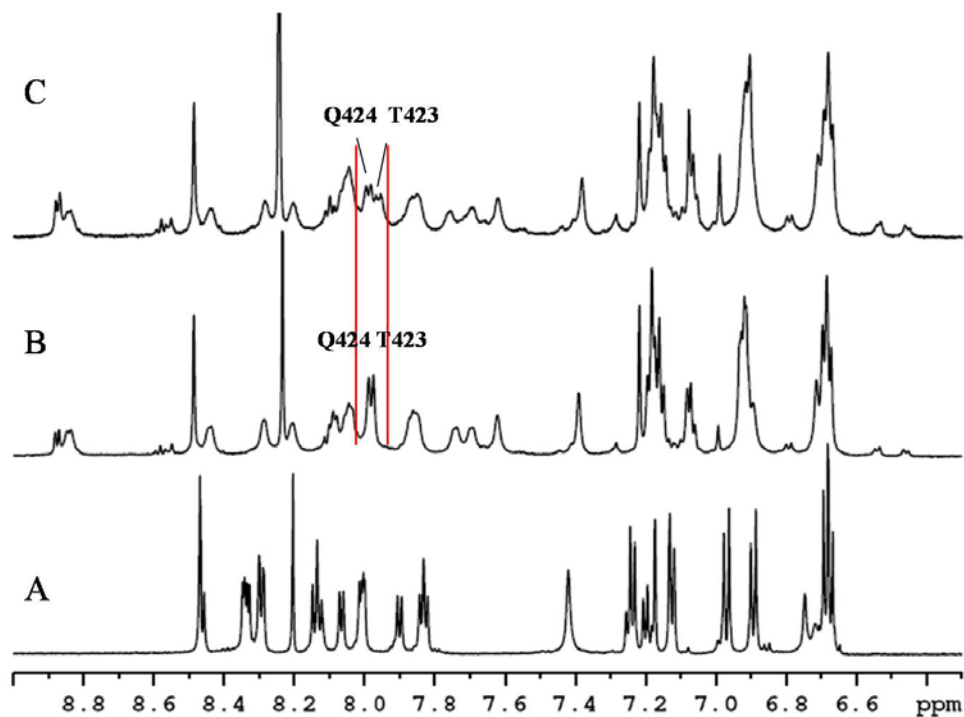
Figure 3.10 Change of the H_N region of S6c upon micelle addition and 1-butanol titration. A): S6c in 10 mM sodium phosphate buffer, pH 4. B): S6c in 30 mM DPC. C): S6c in 30 mM DPC with 2 mM 1-butanol.

3.5.3 Organization of L45, S6c and 1-butanol on micelles

Diffusion measurements using 2D DOSY can aid in exploring intermolecular interaction (*45*). The association of two binding partners is indicated by exhibiting the same diffusion and a slower diffusion compared to the free form (Figure 3.11). As shown in Table 3.1, L45 showed a 58% decrease in the diffusion constant ($15.6 *10^{-11}$ to $6.6*10^{-11}$ $m^2 s^{-1}$) in the presence of micelles. Similarly, the diffusion constant of S6c is also greatly decreased in micelles. This further demonstrates that both L45 and S6c are micelle associated. Furthermore, the lower diffusion constant measured for L45 when S6c is added suggests that both the peptides can coexist on the same micelle. The slightly lower diffusion constant of 1-butanol (65.1 to $57.3*10^{-11}$ $m^2 s^{-1}$) also

indicates its association with the micelles part of the time. This property may facilitate the interaction of 1-butanol with peptide residues that are located at the interface or inside micelles. However, the interaction of 1-butanol with the L45-S6c peptide pair could not be observed in this system. We note that the peptides may distribute unevenly among the micelles, which may complicate an interpretation.
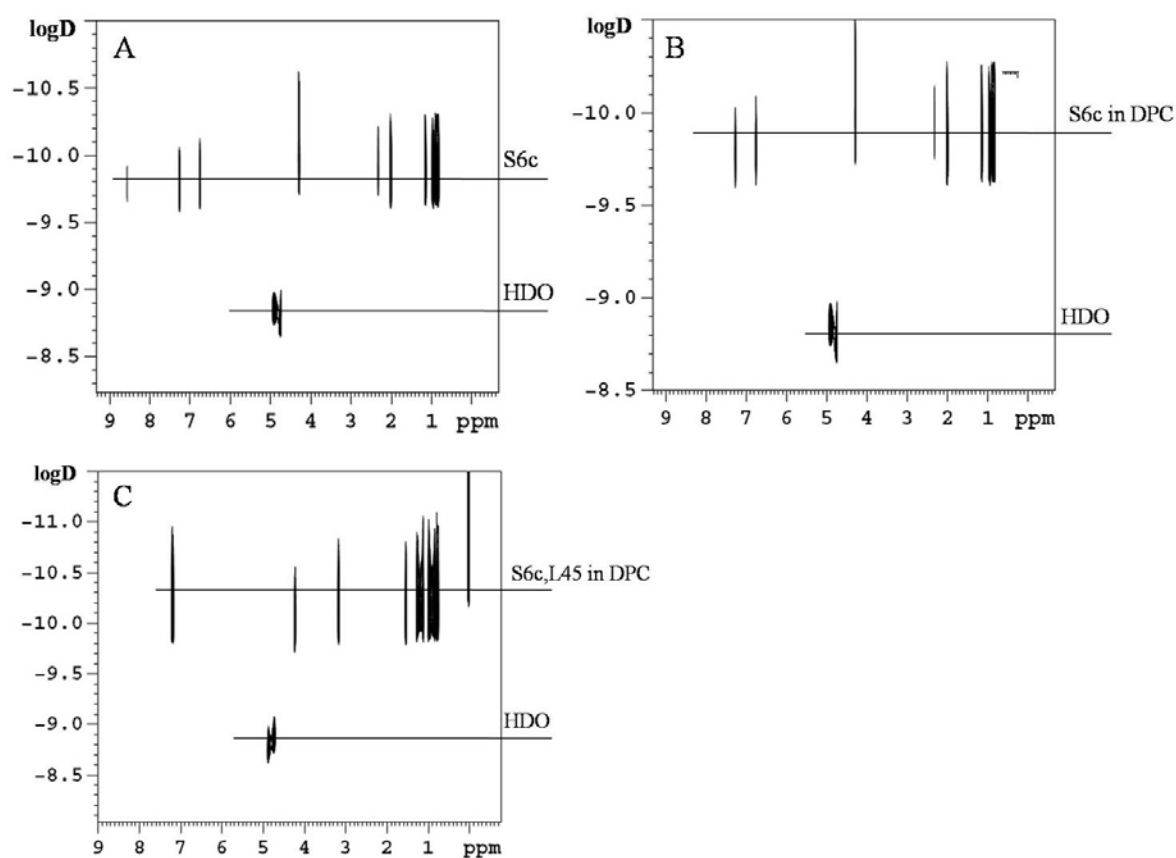
Figure 3.11 2D DOSY spectra of S6c (A), S6c in 30 mM DPC (B) and both S6c and L45 in 30 mM DPC (C) in 10 mM sodium phosphate, 99.96% D_2O (pH*=4.0).

Table 3.1 Diffusion measurements of L45, S6c, micelles and 1-butanol at 293 K. Samples were prepared in 10 mM sodium phosphate, 99.96% D_2O (pH*=4.0). The diffusion constant of HDO servers as an internal reference.

Sample	HDO(10^{-11})	1-butanol(10^{-11})	L45(10^{-11})	S6c(10^{-11})	DPC(10^{-11})
1-butanol (4 mM)	156	65.1±0.5			
DPC (30 mM)	154				7.8±0.2
1-butanol+DPC	154	57.5±0.3			7.8±0.13
L45 (~1 mM)	152		15.6±0.6		
L45+DPC	154		6.6±0.1		7.5±0.12
L45+DPC+1-butanol	156	56.8±2	6.5±0.15		7.6±0.1
S6c (~0.6 mM)	158			15.6±0.1	
S6c+DPC	158			7.6±0.05	7.9±0.25
L45+S6c+DPC+1-butanol	159	59.6±0.2	6.2±0.1	NA	7.7±0.2

3.6   Conclusions

The response of Shaw2 K^+ channels to 1-alkanols is dependent on the S4-S5 linker peptide, as the binding for 1-butanol is coupled to the α-helix propensity of the linker region (*28*). Both CD and NMR experiments established that L45 adopts an α-helical conformation in DPC micelles.

The solvent accessibility and angular orientation of Shaw2 L45 on micelles was determined using paramagnetic perturbation methods. The results substantiate that the linker peptide resides on the membrane surface and lies essentially parallel to it. Surface accessibility experiments suggest that the hydrophobic residues (Leu315, Leu318, and Phe322) are located inside the micelle, while Lys316, Gln320, and Arg323 face towards the solution. The combination of the surface location of L45, its α-helical conformation and orientation as well as the fact that it is important for channel function make it an attractive and accessible molecular target. The other important component of the alkanol response, S6c, in contrast to L45, only forms a partial α-

helix as observed by CD spectroscopy. Moreover NMR data reveal that the α-helix is not stable on the NMR time scale.

Diffusion constant measurements confirmed that L45, S6c and 1-butanol can associate with micelles, which encourages an interaction among the components. Chemical shift perturbations implicate residues Gln320, Thr321, Phe322 and Arg323 on L45 and Thr423 and Gln424 on S6c as potential 1-butanol binding candidates. This finding along with the fact that 1-butanol is capable of interacting with micelles suggests that membrane associated 1-butanol might perturb the interaction of S6c and L45.

Two distinct scenarios based on canonical models of voltage-dependent gating can explain how this perturbation stabilizes the channel's closed state to produce inhibition (*46*). In Shaker-related Kv channels, such as Shaw2, the activation gate prefers its closed conformation (*47*). At hyperpolarized voltages, the voltage sensor is in its "down" position using L45 acting as the load that keeps the activation gate in its closed state. 1-Butanol might then stabilize the "down" position of the voltage sensor by promoting an interaction between L45 and S6c. In the second scenario, the voltage sensor adopts the "up" position upon depolarization of the transmembrane voltage, relieving the activation gate from its load. Consequently, the activation gate passively snaps into its more stable open state. The voltage sensors pull L45, which acts as the "handle" that actively opens the activation gate. 1-Butanol may in this case dislodge L45 from the activation gate and, consequently, the channel remains closed (*27*). However, the available data cannot discriminate between the two models of 1-butanol action. Measurements of gating currents outside the scope of this study will be necessary to tackle this problem more directly.

The mechanism of the gating and inhibition of the Shaw2 Kv channel is complex and require the participation of all four monomers that must be embedded in a cell membrane for func-

tion. Our results shed light on the local structure and interaction of key components of the Shaw2

Kv channel, which govern the 1-alkanol sensitivity of the entire channel.

3.7   References

1.   Latorre, R., and Sáez, J. C. (1997) From Ion Channels to Cell-To-Cell Conversations, *Plenum Press, New York.*
2.   Litt, M., Kramer, P., Browne, D., Gancher, S., Brun, E. R., Root, D., Phromchotikul, T., Dubay, C. J., and Nutt, J. (1994) A gene for episodic ataxia/myokymia maps to chromosome 12p13, *Am J Hum Genet. 55*, 702-709.
3.   Gutman, G. A., Chandy, K. G., Grissmer, S., Lazdunski, M., Mckinnon, D., Pardo, L. A., Robertson, G. A., Rudy, B., Sanguinetti, M. C., Stuhmer, W., and Wang, X. (2005) International union of pharmacology. LIII.nomenclature and molecular relationships of voltage-gated potassium channels, *Pharmacol. Rev. 57*, 473-508.
4.   Ashcroft, F. M. (2006) From molecule to malady, *Nature 440*, 440-447.
5.   Shieh, C.-C., Coghlan, M., Sullivan, J. P., and Gopalakrishnan, M. (2000) Potassium channels: molecular defects, diseases, and therapeutic opportunities, *Pharmacol. Rev. 52*, 557-593.
6.   MacDonald, P. E., and Wheeler, M. B. (2003) Voltage-dependent K+ channels in pancreatic beta cells: role, regulation and potential as therapeutic targets, *Diabetologia 46*, 1046-1062.
7.   Bezanilla, F. (2000) The voltage sensor in voltage-dependent ion channels, *Physiol. Rev. 80*, 555-595.
8.   Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. (2003) X-ray structure of a voltage-dependent K+ channel, *Nature 423*, 33-41.
9.   Long, S. B., Campbell, E. B., and MacKinnon, R. (2005) Voltage sensor of Kv 1.2: structure basis of electromechanical coupling, *Science 309*, 903-908.
10. Long, S. B., Tao, X., Campbell, E. B., and MacKinnon, R. (2007) Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment, *Nature 450*, 376-382.
11. Pathak, M. M., Yarov-Yarovoy, V., Agarwal, G., Roux, B., and Barth, P. (2007) Closing in on the resting state of the shaker K+ channel, *Neuron 56*, 124-140.
12. Labro, A. J., Raes, A. L., Grottesi, A., Hoorick, D. V., Sansom, M. S. P., and Snyders, D. J. (2008) Kv channel gating requires a compatible S4-S5 linker and bottom part of S6, constrained by non-interacting residues, *J. Gen. Physiol. 132*, 667-680.
13. Han, M., and Zhang, J. Z. H. (2008) Molecular dynamic simulation of the Kv1.2 voltage-gated potassium channel in open and closed state conformations, *J. Phys. Chem. B 112*, 16966-16974.

14. Lee, S.-Y., Banerjee, A., and MacKinnon, R. (2009) Two separate interfaces between the voltage sensor and pore are required for the function of voltage-dependent K+ channels, *PLoS Biol. 7*, 0676-0686.

15. Batulan, Z., Haddad, G. A., and Blunck, R. (2010) An intersubunit interaction between S4-S5 linker and S6 is responsible for the slow off-gating component in Shaker K+ channels, *J. Biol. Chem. 285*, 14005-14019.

16. Labro, A. J., Boulet, I. R., Choveau, F. S., Mayeur, E., Bruyns, T., Loussouarn, G., Raes, A. L., and Snyders, D. J. (2011) The S4-S5 linker of KCNQ1 channels forms a structural scaffold with the S6 segment controlling gate closure, *J. Biol. Chem. 286*, 717-725.

17. Lu, Z., Klem, A. M., and Ramu, Y. (2002) Coupling between voltage sensors and activation gate in voltage-gated K+ channels, *J. Gen. Physiol. 120*, 663–676.

18. Jensen, M. Ø., Jogini, V., Borhani, D. W., Leffler, A. E., Dror, R. O., and Shaw, D. E. (2012) Mechanism of voltage gating in potassium channels, *Science Vol. 336*, 229-233.

19. Garcia, M. L., and Kaczorowski, G. J. (2005) Potassium channels as targets for therapeutic intervention, *Sci. STKE 2005*.

20. MacKinnon, R., Cohen, S. L., Kuo, A., Lee, A., and Chait, B. T. (1998) Structural conservation in prokaryotic and eukaryotic potassium channels, *Science 280*, 106-109.

21. Choi, K. L., Mossman, C., Aube, J., and Yellen, G. (1993) The internal quaternary ammonium receptor site of shaker potassium channels, *Neuron 10*, 533-541.

22. Covarrubias, M., Bhattacharji, A., Harris, T., Kaplan, B., and German, M. W. (2005) Alcohol and Anesthetic Action at the Gate of a Voltage-Dependent K+ Channel, *International Congress Series*.

23. Bett, G. C. L., and Rasmusson, R. L. (2008) Modification of K+ channel–drug interactions by ancillary subunits, *J. Physiol. 586*, 929-950.

24. Hodge, J. J. L., Choi, J. C., O'Kane, C. J., and Griffith, L. C. (2005) Shaw potassium channel genes in drosophila, *J. Neurobiol. 63*, 235-254.

25. Covarrubias, M., and Rubin, E. (1993) Ethanol selectively blocks a noninactivating K+ current expressed in Xenopus oocytes, *Proc. Natl. Acad. Sci. USA 90*, 6957-6960.

26. Harris, T., Shahidullah, M., Ellingson, J. S., and Covarrubias, M. (2000) General anesthetic action at an internal protein site involving the S4-S5 cytoplasmic loop of a neuronal K+ Channel, *J. Biol. Chem. 275*, 4928-4936.

27. Bhattacharji, A., Klett, N., Go, R. C. V., and Covarrubias, M. (2010) Inhalational anesthetics and n-alcohols share a site of action in the neuronal Shaw2 Kv channel. , *Brit. J. Pharm. 159*, 1475-1485.

28. Covarrubias, M., Vyas, T. B., Esconar, L., and Wei, A. (1995) Alcoholsinhibit a cloned potassium channel at a discrete saturable site. Insight into the moleculae basis of general anesthesia., *J. Biol. Chem. 270*, 19408-19416.

29. Shahidullah, M., Harris, T., Germann, M. W., and Covarrubias, M. (2003) Molecular features of an alcohol binding site in a neuronal potassium channel, *Biochemistry 42*, 11243-11252.

30. Bhattacharji, A., Kaplan, B., Harris, T., Qu, X., German, M. W., and Covarrubias, M. (2006) The concerted contribution of the S4-S5 linker and the S6 segment to the modulation of a Kv channel by 1-alkanols, *Mol. Pharmacol. 70*, 1542-1554.

31. Zhang, J., Qu, X., Covarrubias, M., and Germann, M. W. (2013) Insight into the modulation of Shaw2 Kv channels by general anesthetics: Structural and functional studies of S4–S5 linker and S6 C-terminal peptides in micelles by NMR, *Biochimica et Biophysica Acta 1828*, 7.

32. Guntert, P., Mumenthaler, C., and Herrmann, T. DYANA, *Version1.5. Zurich, Switzerland. 1998*.

33. Molinari, H., Esposito, G., Ragona, L., Pegna, M., Niccolai, N., Brunne, R. M., Lesk, A. M., and Zetta5, L. (1997) Probing protein structure by solvent perturbation of NMR spectra: the surface accessibility of bovine pancreatic trypsin inhibitor, *Biophys. J. 73*, 382-396.

34. Venditti, V., Bernini, A., Simone, A. D., Spiga, O., Prischi, F., and Niccolai, N. (2007) MD and NMR studies of alpha-bungarotoxin surface accessibility, *Biochem. Biophys. Res. Commun. 356*, 114-117.

35. Respondek, M., Madl, T., Gobl, C., Golser, R., and Zabgger, K. (2007) Mapping the orientation of helices in miccelle-bound peptides by paramagnetic relaxation waves, *J. Am. Chem. Soc. 129*, 5228-5234.

36. Clayton, D., Brereton, I. M., Kroon, P. A., and Smith, R. (1999) NMR studies of the low-density lipoprotein receptor-binding peptide of apolipoprotein E bound to dodecylphosphocholine micelles., *Protein Sci. 8*, 1797-1805.

37. Harris, T., Graber, A. R., and Covarrubias, M. (2003) Allosteric modulation of a neuronal Kv channel by 1-alkanols is linked to a key residue in the activation gate, *Am. J. Physiol. Cell. Physiol. 285*, C788-796.

38. Barber, A. F., Liang, Q., Amaral, C., Treptow, W., and Covarrubias, M. (2011) Molecular Mapping of General Anesthetic Sites in a voltage-Gated Ion Channel, *Biophys. J. 101*, 10.

39. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem. 287*, 243-250.

40. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Anal. Biochem. 287*, 252-260.

41. Plateau, P., and Gueron, M. (1982) Exchangeable proton NMR without base-line distorsion, using new strong-pulse sequences, *J. Am. Chem. Soc. 104*, 7310-7311.

42. Goddard, T. D., and Kneller, D. G. SPARKY 3, *University of California, San Francisco*.

43. Wuthrich, K. NMR of proteins and nucleic Acids, *A Wiley-Interscience Publication, first ed. New York.1986*.

44. Kerssebaum, R. DOSY and diffusion by NMR, *XWinNMR 3.1/3.5 Version 1.03, Rheinstetten, Germany.*

45. Lucas, L. H., and Larive, C. K. (2004) Measuring ligand-protein binding using NMR diffusion experiments, *Concepts Magn. Reson. Part A 20A*, 24-41.

46. Vardanyan, V., and Pongs, O. (2012) Coupling of voltage-sensors to the channel pore: a comparative view, *Front. Pharmacol 3*, 145-154.

47. Yifrach, O., and MacKinnon, R. (2002) Energetics of pore opening in a voltage-gated K+ channel, Cell 111, 231-239.

3.8   Supplementary Materials

Table S3.2 Dyana angle constraints generated by HABAS

Residues		Angle	Value(rad)		Residues		Angle	Value(rad)	
2	LEU	PHI	-65.0	-25.0	7	GLN	PHI	-75.0	-35.0
2	LEU	CHI1	15.0	345.0	7	GLN	CHI1	-165.0	-25.0
2	LEU	CHI1	-155.0	145.0	7	GLN	CHI1	-55.0	295.0
2	LEU	PSI	-285.0	-5.0	7	GLN	CHI2	25.0	335.0
3	LYS	PHI	25.0	315.0	7	GLN	CHI2	-115.0	115.0
3	LYS	PHI	-195.0	95.0	7	GLN	PSI	-85.0	-15.0
3	LYS	PHI	-85.0	205.0	8	THR	PHI	-325.0	-25.0
3	LYS	CHI1	-215.0	-15.0	8	THR	PHI	-195.0	85.0
3	LYS	CHI2	25.0	335.0	8	THR	CHI1	-145.0	-15.0
3	LYS	PSI	-115.0	115.0	8	THR	CHI21	25.0	325.0
4	ILE	PHI	35.0	325.0	8	THR	CHI21	-95.0	215.0
4	ILE	PHI	-155.0	45.0	8	THR	PSI	-195.0	-25.0
4	ILE	CHI1	-155.0	-15.0	8	THR	PSI	-115.0	235.0
4	ILE	CHI21	15.0	345.0	9	PHE	PHI	25.0	305.0
4	ILE	PSI	-155.0	125.0	9	PHE	PHI	-175.0	85.0
5	LEU	PHI	-75.0	-35.0	9	PHE	PHI	-95.0	215.0
5	LEU	CHI1	-335.0	-15.0	9	PHE	PSI	-105.0	105.0
5	LEU	CHI1	-285.0	35.0	10	ARG+	PHI	-75.0	-45.0
5	LEU	CHI1	-145.0	145.0	10	ARG+	CHI1	-195.0	-85.0
5	LEU	CHI1	-35.0	285.0	10	ARG+	CHI2	25.0	335.0
5	LEU	CHI2	35.0	325.0	10	ARG+	PSI	-105.0	-15.0
5	LEU	CHI2	-305.0	45.0	11	ALA	PHI	25.0	305.0
5	LEU	CHI2	-115.0	225.0	11	ALA	PHI	-185.0	95.0
5	LEU	PSI	-105.0	5.0	11	ALA	PHI	-85.0	215.0
6	ILE	PHI	-125.0	-35.0	11	ALA	PSI	-135.0	135.0
6	ILE	CHI1	-95.0	-25.0	12	SER	PHI	25.0	315.0
6	ILE	CHI21	85.0	225.0	12	SER	PHI	-215.0	95.0
6	ILE	CHI21	-185.0	125.0	12	SER	PSI	-25.0	125.0
6	ILE	PSI	-95.0	45.0	13	ALA	PHI	25.0	305.0
					13	ALA	PHI	-185.0	95.0
					13	ALA	PHI	-85.0	215.0

A

Residues	G	L	K	I	L	I	Q	T	F	R	A	S	A
$T_1(2)/T_1(0)$	0.38	0.96	0.33	0.45	0.84	0.62	0.63	0.87	1.00	0.52	0.82	0.88	1.07

B
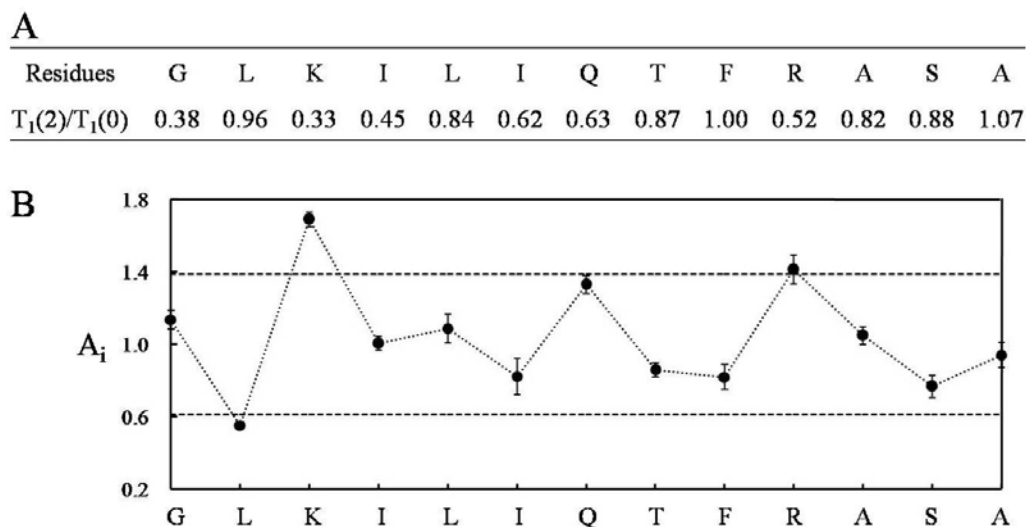
Figure S3.1 A): T_1 ratio of L45 H_α protons. $T_1(0)$ and $T_1(2)$ are T_1 of L45 H_α protons in absence or presence of 2mM of Gd-DTPA-BMA. B): Paramagnetic attenuation A_i of L45 H_α protons versus residue number.
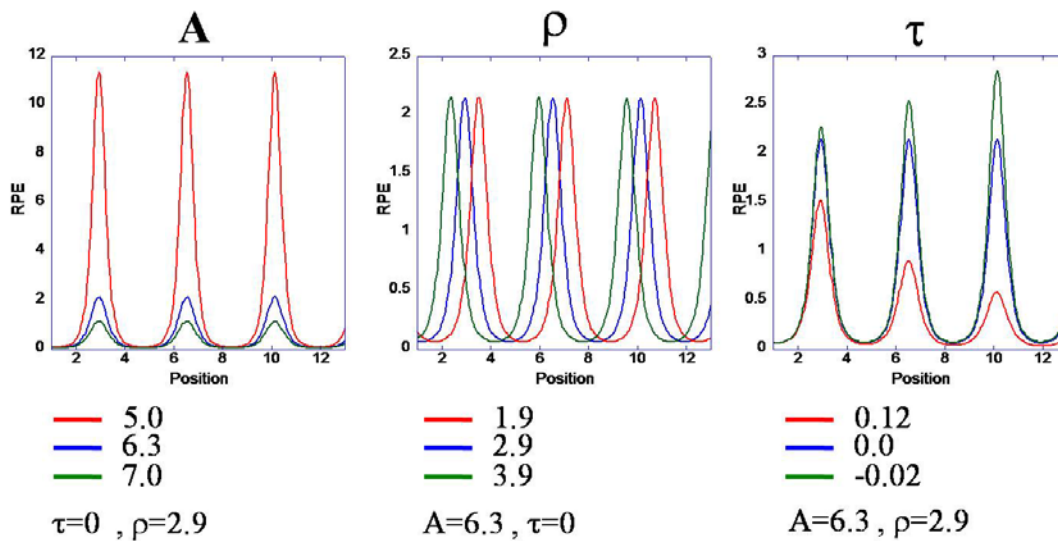
Figure S3.2: simulated PRE as a function of A, τ and ρ using a model α-helical system. All τ and ρ are in unit of rad.

4   INTERACTIONS AND STRUCTURAL STUDIES OF TIAR PROTEINS AND WEST NILE VIRAL RNA

4.1   Preface

This project was initiated in Dr. Brinton's lab in the Department of Biology. The expression and initial purification of the proteins were carried out in her lab.

4.2   Introduction

4.2.1   Protein-RNA interactions and their role in biological systems

Protein-RNA interactions (PRIs) play diverse roles in regulating many cellular events. They are essential in the structure of the spliceosome and ribosome, and important in alternative splicing, RNA editing, polyadenylation, mRNA export, mRNA stabilization, mRNA localization and mRNA translation (*1-2*). In addition, protein–RNA interactions are also very important for viral function where host proteins were found in the replicase complexes of several RNA viruses (*3-6*).

4.2.2   Protein-RNA recognition modes

RNA-binding proteins exhibit highly specific recognition of their RNA targets by recognizing their sequences and/or structures. Sudies of protein-RNA interaction have lagged compared to protein-DNA mainly because of the diversity in RNA structure and the more complex protein-RNA recognition patterns. Based on information obtained from recently available struc-

tures of RNA-protein complexes (*7-9*), RNAs exhibit various conformations when they bind to proteins (*10*). Generally, the irregular secondary structure elements of RNA, such as hairpin loops and bulges where bases are more exposed are the usual binding targets for proteins (*11-14*). This indicates a very different recognition strategy from the accessibility-controlled model of protein-DNA binding. In Draper's review (*15*), protein-RNA recognition modes were categorized into two groups: In group A, peptides and proteins place a secondary structure element into RNA helix grooves, recognizing both the specific sequence of RNA bases and the shape or dimensions of the RNA groove. In group B, the β-sheet surface of the proteins creates a binding pocket that reads single-stranded RNA in a sequence-specific mode. It also has been argued that protein-RNA binding is accompanied by a co-folding process, where both the protein and RNA experience conformational changes to form tight binding (*16*).

4.2.3   RNA binding protein motifs

Structural studies have revealed several protein domains exhibited by RNA binding proteins, such as: the RNA Recognition Motif (RRM, also known as RBD or RNP domain), K Homology (KH) domain, Zinc finger (mainly C-x8-C-x5-C-x3-H type), double-stranded RNA binding domain (DS-RBD), RGG box, DEAD/DEAH box, Pumilio/FBF (PUF) domain, Piwi/Argonaute/Zwille (PAZ) domain and Sm domain, etc. RBPs have one or multiple copies of the same RNA binding domain, or they may have multiple distinct domains (*17*). Several RNA binding domains are suggestive of the molecular function of the protein; for example, the DEAD/DEAH box is indicative of RNA helicase activity, whereas the PAZ domain suggests the protein targets short single-stranded RNA such as RNAi or microRNAs (miRNA). However, other domains only predict RNA binding, such as RRMs.

4.2.4 Intermolecular forces involved in protein-RNA binding and interface residue propensity

Intermolecular forces involved in protein-RNA recognition are weighted slightly different from those in protein-DNA recognition. Proteins prefer to bind to the wide major groove of DNA, where positively charged amino acid side chains take part in electrostatic interactions with the DNA phosphate backbone. However, in protein-RNA binding, van der Waals interactions and hydrophobic contacts dominate the binding interactions. Comparable numbers of contacts are made by the proteins to the backbone and to the bases (*16, 18*). A higher percentage of hydrogen bonds between the protein main chain and the RNA is another trait in protein-RNA interactions, which facilitates close contacts (*19*). In the protein-RNA interface, in addition to positively charged amino acid residues, aromatic residues are frequently found (*19-20*).

4.2.5 Methods used in detecting and studying protein-RNA interaction

4.2.5.1 General methods

(A) Spectroscopic methods:

This group of techniques takes advantage of the fact that the formation of a complex may perturb the response to electromagnetic radiation, or respond to the input radiation by emitting its own characteristic signals. The methods include ultraviolet (UV), fluorescence, circular dichrosim (CD) and nuclear magnetic resonance (NMR) spectrophotometry. For each type of spectroscopy, the changes in the spectroscopic signals can be used to determine the relative concentrations of free and bound proteins; therefore, the affinity or K_d of an interaction can be determined. In general, the interpretations of the spectra fall into two categories according to the relative time scales of both the spectroscopy and the binding event. In a case where the spectroscopic tech-

nique is rapid relative to the rate of exchange between free and bound species, the free and bound species can be resolved in the spectra and the intensity of the signal represents the concentration of each species. UV, fluorescence and CD usually fall in this category. In the second category, the spectroscopic techniques are slower than the association/dissociation event and the observed signal in spectra will be the population-weighted average of the two species and changes in the signals are proportional to the ratio of complex to free species. This is often the case for NMR.

(B) Calorimetry:

This technique measures the heat generated or absorbed when ligands bind to a receptor. More specifically, a receptor is titrated with a ligand in a calorimeter and the heat associated with the binding at each titration point is measured. The binding isotherms can be generated by plotting the heat against free ligand concentration. The heat produced by dilution of the receptor or by non specific binding must be considered and subtracted.

Other general binding detection methods include gel chromatography-based and affinity chromatography-based methods.

4.2.5.2 Specific methods for detection of protein-RNA interactions

(A) Gel Mobility Shift Assay:

This technique is used to detect protein-RNA interactions through changes in RNA migration upon binding to protein during gel electrophoresis. Traditionally, RNA probes are radioactively labeled, or incorporate fluorescent or chemiluminescent groups. Specificity of the binding can be determined through a competition reaction where excess unlabeled RNA is incubated in the binding reaction as a competitor, resulting in a decrease in the shifted signal if the labeled

and unlabeled RNA sequences compete for binding to the same protein. This permits determination of the dissociation constant by examining the bound fraction as a function of protein concentration. However, the technique may tend to be limited to fairly tight or slow binding, where the $t_{1/2off}$ is also on the order of seconds or longer. Alternatively, the protein-RNA complex may be cross-linked prior to a denaturing gel analysis.

(B) RNA Pull-down Assay:

The RNA pull-down assay utilizes high affinity tags, such as biotin or azido-phosphine. RNA probes are biotinylated and then allowed to bind to a protein in a cell lysate. The complex is purified using agarose or magnetic beads with an antibody against the protein of interest. The RNA is then detected by Northern blotting or through real time-PCR analysis and the proteins can be detected by Western blotting or mass spectrometry.

(C) Oligonucleotide-Targeted RNase H Protection Assays:

RNase protection assays can be used to map protein binding sites on the entire target RNA sequence. In this assay, a DNA probe is designed to hybridize to a specific site on the RNA, then RNase H is incubated with the protein-RNA reaction. If a protein is bound to the RNA at the target sequence, it will block the RNA-DNA hybridization and prevent the cleavage by RNase H, therefore indicating a site of interaction between protein and RNA (*21-22*).

4.2.6   *West Nile virus RNA and its interaction with cellular proteins TIA-1 and TIAR*

4.2.6.1 West Nile virus and its genome

West Nile virus belongs to the genus *Flavivirus* in the family *Flaviviridae*. It is a mosquito-borne virus that was first identified in the West Nile subregion in the East African nation of Uganda in 1937. The most common symptom in infected humans is fever; about 3% of the cases are severe, resulting in neurological disease and even death. However, the virus has now spread globally and is considered a serious threat to public health. This brings urgency for the development of specific antiviral treatments and vaccines against WNV infection for humans (*23*).

The genome of WNV is a single-stranded, positive sense RNA with an approximate length of 11 kb. The genome contains a single, long open reading frame (ORF) of 10.3 kb flanked by 5' and 3' noncoding region (NCR) (*24*). The 5' end of the genome ORF encodes the structural proteins while the 3' end encodes the nonstructural proteins that serve multiple functions during the virus life cycle. Unlike cellular mRNA, the WNV genome lacks a 3' polyadenylate tail (*25*).

4.2.6.2 WNV replication cycle and the roles of genome RNA and its complementary negative strand

The WNV replication cycle starts with viron attachment to and entry into the host cell. The genomic RNA is then released into the cytoplasm by fusion of the viral and cell endosome membranes and translated into a single polyprotein which is sequentially cleaved to generate the mature viral proteins. The genomic RNA also serves as a template for synthesis of complementary minus strands; these minus-strand RNAs in turn serve as templates for the production of more positive genomic RNAs. It has been reported that WNV RNA synthesis is semiconserva-

tive and asymmetric, with plus strands accumulating in about a tenfold excess over minus strands (*26*). When a sufficient pool of endoplasmic reticulum membrane associated structural proteins is available, the nascent viron RNA will self-assemble into new virial particles and that are transported in vesicles to the cell plasma membrane where the virons are released by fusion of the vesicles and plasma membranes (*27*).

During the replication cycle, the nascent genome RNAs function as templates for translation, transcription and as substrates for viron formation; while the negative strands are only found in the virus replication complexes serving as templates for genome RNA synthesis.

4.2.6.3 Structural conservation of the terminal regions of the WNV genome

As predicted by chemical digestion, stable stem loop (SL) structures are located at both the 3' and 5' termini of the WNV genome RNA (*28-29*) (Figure 4.1). These secondary structures are conserved at the termini of different flaviviruses genome even though the majority of the sequences are not well conserved (*29-30*). The existence of SL structures was subsequently predicted in the complementary region (3' termini) of the negative strand, and this was supported by CD spectroscopic, thermal melting curve and RNase probing studies (*31*). These structures are functionally important since mutation of the nucleotides forming the terminal SL structure has a negative effect on the efficiency of virus replication and deletion of the SL has been shown to be lethal for flavivirus infectious clones (*32*).

4.2.6.4 Cellular protein TIA-1 and TIAR

Cellular protein TIA-1 (T cell restricted intracellular antigen-1) and TIAR (TIA-1 related protein) are evolutionarily conserved RNA binding proteins of the RNA recognition motif

(RRM)/ribonucleoprotein (RNP) family (*33-35*). They are found in the cytoplasm and nucleus of most cells and tissues, and shuttle between these two compartments (*36*). They are both multi-functional RNA-binding proteins. In the nucleus, they have been shown to regulate transcription by binding to single stranded T-rich DNA and alternative pre-mRNA splicing by binding to U rich RNA targets and promoting the recognition of atypical 5' splice sites (*37-43*). In the cyto-plasm, they are involved in silencing translation when cells are under stress by binding to the polyA of target mRNAs and sequestering the bound mRNAs into cytoplasmic stress granules (*44-46*). They have been also reported to regulate Fas-mediated apoptosis (*47*) and to provide critical functions during embryonic development (*48*). Recently, they have been found to arrest translation at the initiation step by binding to 5'-oligopyrimidine elements under stress (*49*).

TIA-1 and TIAR both possess three RRMs composed of 90 to 100 amino acids in the N terminus (*33, 50*), that form the RNA binding domain. The C-terminus is an auxiliary domain rich in glutamine and structurally related to prion proteins (*36, 51*). TIA-1 and TIAR share more than 90% amino acid identity in their RRM regions with most of the differences found in the C-terminus. A conserved peptide octamer (K/R-G-F/Y-G/A-F/Y-V/I/L-X-F/Y) and a hexamer (I/V/L-F/Y-I/V/L-X-N-L), referred to as the (ribonucleoprotein) RNP1 and RNP2, respectively, are located in each RRM (*52*). In addition, each RNP contains three conserved aromatic residues, which accommodate two bound nucleotides via ring stacking interactions and specific hydrogen bonding networks. It was argued that these interactions are the determinants for the RNA binding specificity (*53-54*).

Structures of individual RRMs have revealed that the RRMs of TIA-1 and TIAR consist of four anti-parallel β-strands packed against two α-helices with the canonical βαββαβ topology (*55-56*) (PDB: TIAR RRM1, 2CQI; RRM2, 2DH7; RRM3,1X4G). The surface of the four β-

strands has been shown to recognize single stranded RNA, specifically short stretches of uridine (*55, 57*). Among the three RRMs, the second domain, RRM2 has been shown to be necessary and sufficient for tight RNA binding, but RRM2 together with RRM1 and RRM3 shows increased affinity for U-rich or AU-rich RNA (*57-58*). The inter-domain linker, the loops between secondary structure components and the C-terminal extension also contribute to RNA-binding specificity and affinity (*55, 59-63*). Recently, TIAR was reported to also bind to C-rich sequences (*64*).

4.2.6.5 TIA-1 and TIAR interact with the 3' stem loop of WNV complementary negtive strand RNA

TIAR and TIA-1 were reported to colocalize with the viral replication complexes in flavivirus infected cells (*65*). Specifically, they bind to the stem loop region at the 3' terminus of WNV complementary negative-strand (WNV 3' (-) SL) (*66*), which serves as a site for initiating positive strand synthesis during RNA replication. This result together with the observation that WNV growth was less efficient in murine TIAR knockout cell line suggests the interaction of TIAR and TIA-1 with the WNV 3'(-) SL facilitates viral replication (*66*).

Data obtained from a binding study of the truncated proteins and WNV 3'(-) SL RNA have also shown that RRM2 domain of TIA-1 and TIAR is responsible for the major interaction to WNV 3' (-) SL with high affinity (*66*). This is in agreement with previous findings showing the interaction of TIAR and TIA-1 with AU-rich sequence (*57-58*). By truncation and mutation of the WNV 3' (-) SL, the binding sites of RRM2 on the WNV 3' (-) SL has been mapped to two UUAAU sequences located on two adjacent single strand loops (*67*), loop 1 and 2 as shown in Figure 4.2. Deletion or mutation of loop 3 (L3) had neglectable effect on binding affinity. This target sequence differs from the ARE class I sequences which consist of one to three copies of

scattered AUUUA motifs or the class II ARE sequences consisting of at least two overlapping UUAUUUA(U/A)(U/A) nonamers (*68-69*). However, the viral sequence is thought to be capable of outcompeting the cellular RNAs which are most abundant in host cells.

In this chapter, we focus on the interaction of the TIAR protein and its RRM2 with WNV 3' (-) SL (WNV SL75) and different truncated versions of WNV SL75, as well as oligoU sequences of different lengths. By applying a combination of calorimetry, spectroscopy, computational and other techniques, the biophysical and structural properties of the binding partners were explored. The findings provide insights into the basis of the specific cellular protein–viral RNA binding interaction and preliminary data for the development of strategies on how to interfere with virus replication.

4.3 Materials and Methods

4.3.1 Protein expression and purification

In this study, all of the recombinant proteins (hTIAR RRM2 and hTIAR wild type) were fused to N-terminal Glutathione-S-transferase (GST) tags and expressed in *Escherichia coli* strain Rosetta 2(DE3) pLysS cells (Novagen). Cells were incubated at 37 °C until an optical density of 0.6-0.7 at 600 nm was reached. Protein expression was induced by addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). After overnight post-induction at 20°C, cells were harvested by centrifugation. Cell pellets were lysed upon thawing at room temperature with phosphate buffered saline (PBS) buffer containing CelLytic Express (Sigma-Aldrich Co. LLC) and 1 mM dithiothreitol (DTT) and 1 tablet/ml complete EDTA-free protease inhibitor cocktail (Roche). The lysate was clarified by centrifuge at 4°C, 10,000 × g. The resulting supernatant was purified with Glutathione Sepharose 4B (GE Healthcare) matrix following the manufacturer's

protocol. The GST tag was then removed by addition of the site-specific protease PreScission protease (GE Healthcare) to the sepharose bound fusion protein at a concentration of 80 units/ml. After overnight incubation, the tag-free protein was collected in the flow-through and multiple elutions. Gel filtration purification using a Superdex 75 column (Phamacia Biotech) was then carried out to ensure the purity of the protein and buffer-exchange to a typical sample buffer containing 20 mM sodium phosphate buffer (pH 7.0) and 100 mM KCl was done.

4.3.2   RNA sample preparation

The RNA sequences rU_5 (5'-UUUUU-3'), rC_5 (5'-CCCCC-3'), rA_5 (5'-AAAAA-3') used in this study were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). The rU_{12} (5'-UUUUUUUUUUUU-3'), rU_{15} (5'-UUUUUUUUUUUUUUU-3'), rU_{20} (5'-UUUUUUUUUUUUUUUUUUUU-3') and 5' fluorescein labeled rU_{20} (F-rU_{20}) as well as truncated WNV RNA sequences, SL40 (5'-GGCUAAUUGUUGUUAAUCCUCGCGGAGACGCCAGCUCGCC-3') (Figure 4.2B) and SL20 (5'-UUAAUUGUUGUUAAUCCUCA-3') were purchased from Dharmacon, Thermo Scientific, (Pittsburgh, PA) as the 2-hydroxyl protected form. The RNA samples were deprotected in house following the manufacturer's protocol and desalted afterwards using a HiTrap^{TM} desalting column (GE Healthcare). The dsDNA template for SL75 (5'-CAGCUCGCACCGUGUUAAUUGUUGUUAAUCCUCACAAACACUACU-AAGUUUGUCAGCUCACACAGGCGAACUACU-3') (Figure 4.2A) was amplified by PCR using pWNV75NCR as the template and primers 5'-[T7]-CAGCTCGCACCGTGTTAATTGTTG-3' and 5'-AGTAGTTCGCCTGTGTGAGC-3'. T7 polymerase was used to *in vitro* synthesize SL75 from the PCR template. The SL75 RNA transcript

was isolated from the reaction mixture using a cation exchange column, Mono Q (GE Healthcare). The purity of all the longer RNA sequences was confirmed by urea polyacrylamide gel electrophoresis.

4.3.3 Electrophoresis

4.3.3.1 Tris nondenaturing acrylamide gels

Tris nondenaturing 14% acrylamide gels were prepared using $4 \times$ Tris gel buffer containing 200 mM Tris.Cl , pH 7.1 and 30% acrylamide/Bis (29:1) solution. The protein was solubilized in 25 mM Tris·Cl, pH 7.0 and 75 mM NaCl with 5% sucrose from which 20 μg was loaded for analysis. For protein/RNA complex samples, short RNA oligos were mixed with the protein in a 1:1 molar ratio. Cytochrome (10 μg) dissolved in the same sample buffer was used as a marker. The gel was prerun at 4 °C with a voltage of 30 V/cm for 30 min. Then, a voltage of 5 V/cm was applied to the gel with the electrode reversed at the power supply to drive positively charged protein migration to the cathode. After electrophoresis, the proteins in the gel were stained using GelCode Blue stain reagent (Thermo Scientific).

4.3.3.2 SDS polyacrylamide gels

To analyze RRM2 and its multimers, 13.5 % SDS polyacrylamide gels (SDS PAG) were made from 30% acrylamide/Bis (29:1) solution following the standard protocol.

4.3.3.3 Electrophoretic mobility shift assay

Nondenaturing polyacrylamide gels (8%-12%) made from 40% acrylamide/Bis (19:1) solution were used for gel shift assays. Typically, 50 pmol of RNA was incubated with 50 to 150 pmol of RRM2 in 10 μl of sample buffer (20 mM sodium phosphate buffer, pH 7.0 and 100 mM KCl) on ice for 10 min prior to electrophoresis. A voltage of 10 V/cm was applied and the gels run for ~3 h at 4 °C. After electrophoresis, RNA on the gel was stained with SYBR Green II and visualized using Typhoon9400 (GE Healthcare).

4.3.4   Protein crosslinking

The protein crosslinker dithiodis[succinimidypropionate] (DSP) was dissolved in dry DMSO at a concentration of 50 mM. Proteins (50 μM) were prepared with or without WNV SL20 RNA in a reaction of 50 μl with sample buffer. Then, 50 fold excess of the crosslinker was added to the protein or the protein-RNA mixture. The reaction was incubated for 2 h on ice. The mixture was loaded directly onto 13.5% SDS PAG and subjected to electrophoresis to separate the RRM2 monomer and multimers.

4.3.5   CD spectroscopy

TIAR RRM2 was dialyzed against 10 mM sodium phosphate buffer (pH 7.0) with 50 mM Na_2SO_4 and the concentration was adjusted to 50 μM. A 2 mm cuvette was used with a JA-810 spectropolarimeter (Jasco, Tokyo, Japan) for recording all CD spectra at room temperature. Each spectrum was recorded from 260 to 180 nm at a scan rate of 10 nm/min for a total of eight scans. The resulting spectra were deconvolved using the online server, DichroWeb (*70*).

4.3.6 Steady-state fluorescence anisotropy measurement

Fluorescence anisotropy experiments were carried out on a Perkin Elmer LS55 fluorescence spectrometer (Waltham, MA) using an excitation and emission wavelength of 495 nm and 521 nm, slit widths of 10 mm and 3 mm, respectively. F-rU$_{20}$ (100-150 nM) was prepared in 600 µl of sample buffer or low salt buffer (4 mM sodium phosphate, 20 mM KCl with pH 7.0) using a 5×5 mm fluorimeter quartz cuvette. Anisotropy of F-rU$_{20}$ was measured after each addition of 0.5-2 µl protein sample. The data were analyzed with Graphpad Prism.

4.3.7 Isothermal titration calorimetry (ITC) experiments

All the ITC experiments were carried out on a VP-ITC microcalorimeter (MicroCal, LLC, Northampton, MA) at 25 °C with reference power of 18 µcal/sec and a stirring speed of 351 rpm. Prior to each experiment, both the RNA and the protein samples were dialyzed overnight against the same buffer (20 mM sodium phosphate, pH 7.0, 100 mM KCl) and adjusted to the desired concentration. Samples were degassed before loading into the sample cell or the autopipette. The titration was performed by adding 8-10 µl aliquots of the RNA to protein every 300 s for a total of 25 injections. The normalized heat signals were calculated using the bundled Origin software and the data was fitted using a one site binding model.

4.3.8 NMR experiments

All NMR spectra were recorded on a Bruker 600MHz instrument equipped with a 5 mm triple resonances (QXI) Z-gradient probe head. Pre-saturation was used in 1D ^{1}H and 2D NOESY spectra to suppress the solvent (H$_2$O) signal. The 2D NOESY was collected with a data matrix of $2k \times 512$, 32 scans and mixing time of 250 ms.

4.3.9   Electrostatic potential map calculation

The electrostatic potential maps of TIAR RRM1, 2 and 3 were calculated using the online server PDB2PQR (*71*) to generate a PQR file from the protein PDB file. The PQR file was then calculated with an APBS package complied in the VMD program to assess the molecular electrical potential surfaces. The resulting files were visualized with VMD (*72*).

4.3.10   Molecular docking

The interaction of TIAR RRM2 with short RNA sequences, rU_5, rA_5 and rC_5 were evaluated using AutoDockVina (The Scripps Research Institute) (*73*). The RNA molecular models were built and minimized with molecular mechanics force fields in Spartan (Wavefunction, Inc.). PDB files were also generated in Spartan. The protein and the RNA PDB files were then prepared in AutoDockTools (*73*) to generate pdbqt files for molecular docking. The final results were visualized using AutoDockTools.
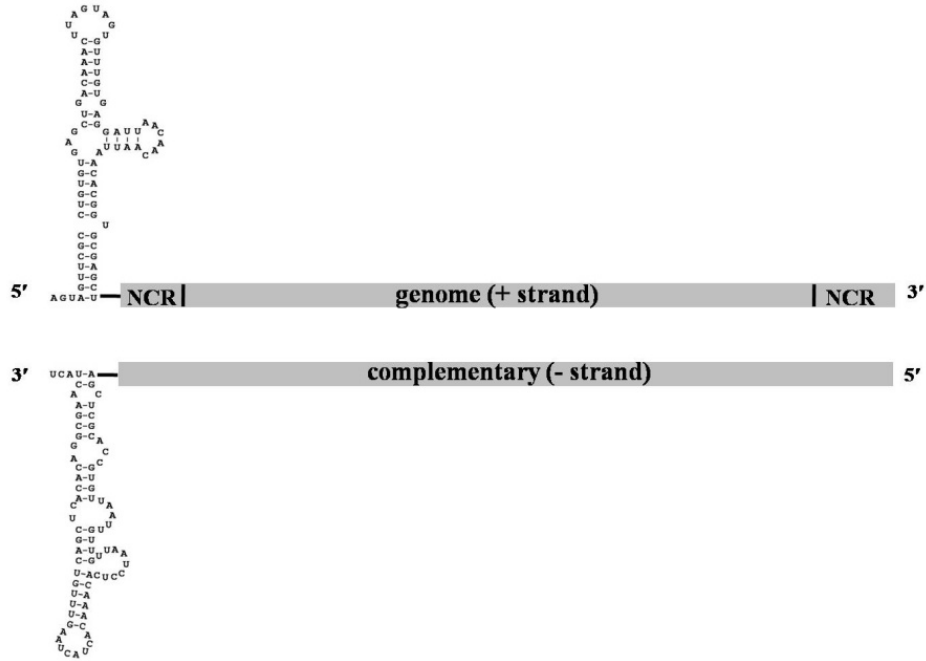
Figure 4.1 WNV genome and the complementary negative strand with predicted secondary structure at the NCR of the positive strand 5' terminus and negative strand 3' terminus.
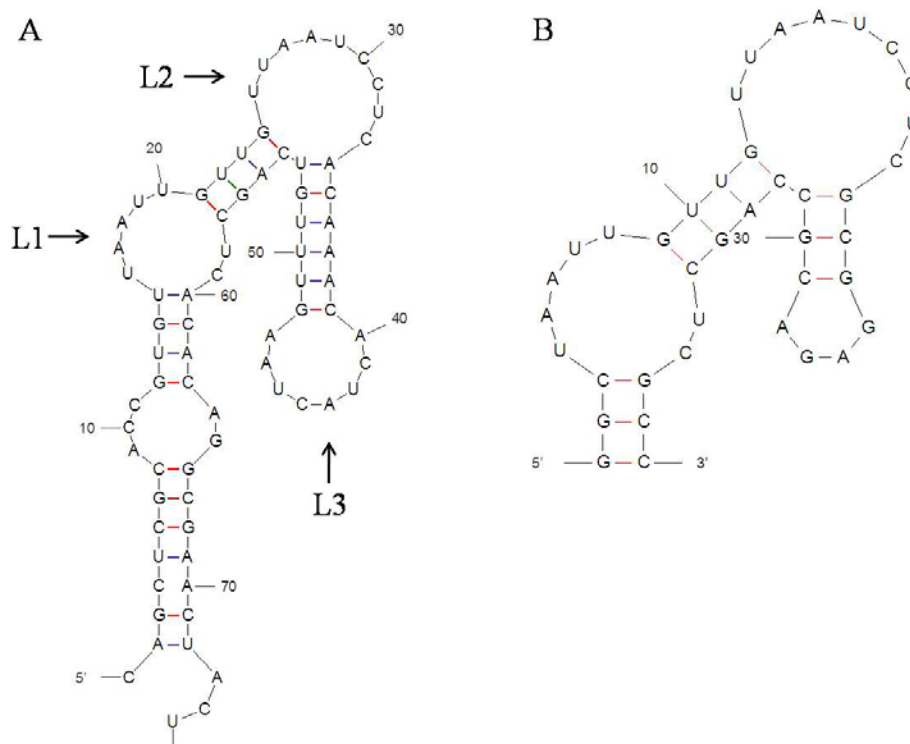
Figure 4.2 SL75 and SL40 structure generated using Mfold (*74*). Loop1, 2 and 3 on SL75 are indicated with L1, 2, 3.

4.4   Results and Discussion

4.4.1   Predicting the interaction of RRM2 with short RNA substrate from molecular docking

Molecular docking predicts the preferred orientation of ligands when bound to a receptor by searching for a minimum energy state. We performed molecular docking on RRM2 and various short RNA sequences to probe the potential fits of different RNAs to RRM2 and evaluate the differences in their contact pattern. As shown in Figure 4.3, the RRM2-rU$_5$ complex with the lowest binding energy shows that the rU$_5$ binds at the second β strand (β2). The RNA was stretched to fit the accessible surface of the protein. This search also predicted other binding models where rU$_5$ binds the loops regions (Figure 4.3 C and D). The binding sites of rA$_5$ on

RRM2 were found to be similar to rU$_5$ (Figure 4.4A). In the predicted model for RRM2-rC$_5$, rC$_5$

lies on the β$_4$ strand in addition to loop binding (Figure 4.4B). In all docking experiments, the

phosphodiesters and the glycosidic bonds are fully rotatable, allowing the RNA to mold to the

protein surface. However, the simulation provides limited information on the binding preference

of RRM2 as judged from the calculated binding energy. Clearly, experiments need to be con-

ducted to analyze this interaction.
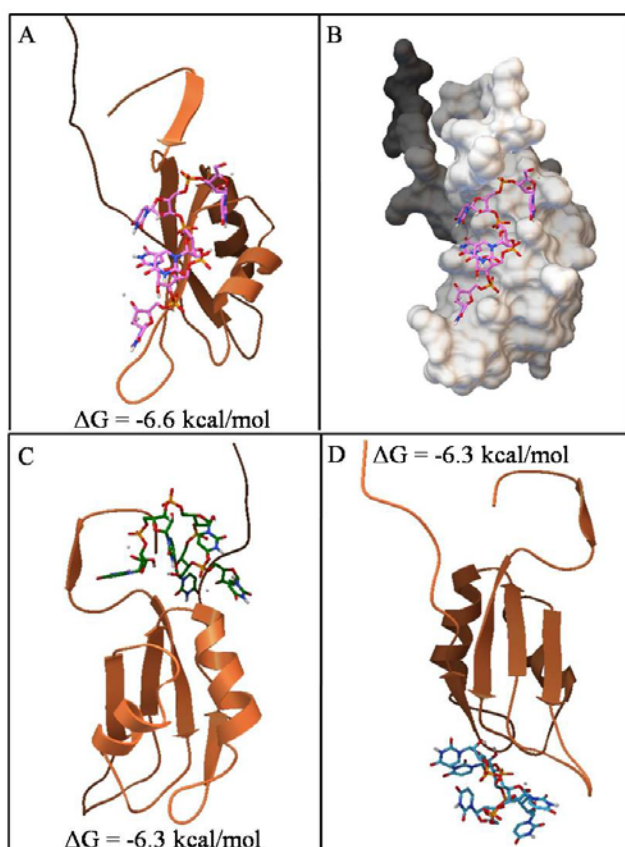
Figure 4.3 The calculated models of TIAR RRM2 interacting with rU$_5$, (A), (C) and (D) with different binding patterns; (B) the accessible surface representation of the model in (A). Binding free energies are indicated.
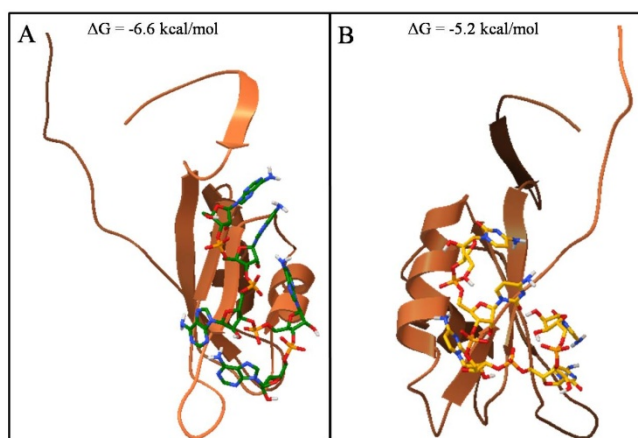
Figure 4.4 Representative models of TIAR RRM2 interacting with rA$_5$, (A) and rC$_5$, (B).
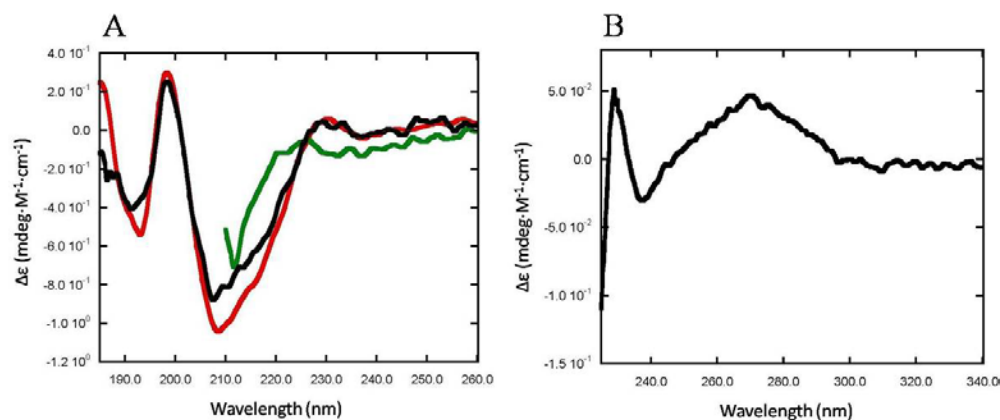
Figure 4.5 A): CD spectra of TIAR RRM2 in the far UV region; red, initial spectrum; green, spectrum of the sample containing 6 M guanidine HCl and spectrum of the refolded sample, black; B): near UV CD spectra. Protein sample is in 10 mM sodium phosphate buffer (pH 7.0) with 50 mM Na$_2$SO$_4$.

4.4.2   The folding status of hTIAR RRM2

The wild type TIAR protein contains three RRM domains. We first focused on the binding properties of RRM2, this single domain was fused to an N terminal GST tag, which was removed during purification. To ensure proper folding of the protein, the purified protein was

characterized by CD spectroscopy. The existence of α-helix and β-strand components is indicated by the negative peak in the range of 208-220 nm and the positive peak around 195 nm in far UV spectra, respectively. The negative peak in the range of 190-200 nm indicates β turns and random coil (Figure 4.5A). The protein was denatured by 6 M Guanidine HCl and refolded by dialysis against the sample buffer. The CD spectra of the refolded protein showed the same profile as the initial protein sample. Deconvolving the CD spectra using the online server DichroWeb reveals about 10% α-helix, 40% β-strand and 20% turns. These results together with the positive peak in the near UV region are indicative of proper folding of the protein (Figure 4.5B). In addition, RRM2 NMR resonances are well dispersed in 1D ^{1}H spectrum, specially three sets of peaks in the range of 5-6 ppm characterizing the β strand components. The 2D NOESY spectra also showed connection of H_N to H_N and H_N to side chain Hs (Figure 4.6). These data further confirmed the folding of the protein.
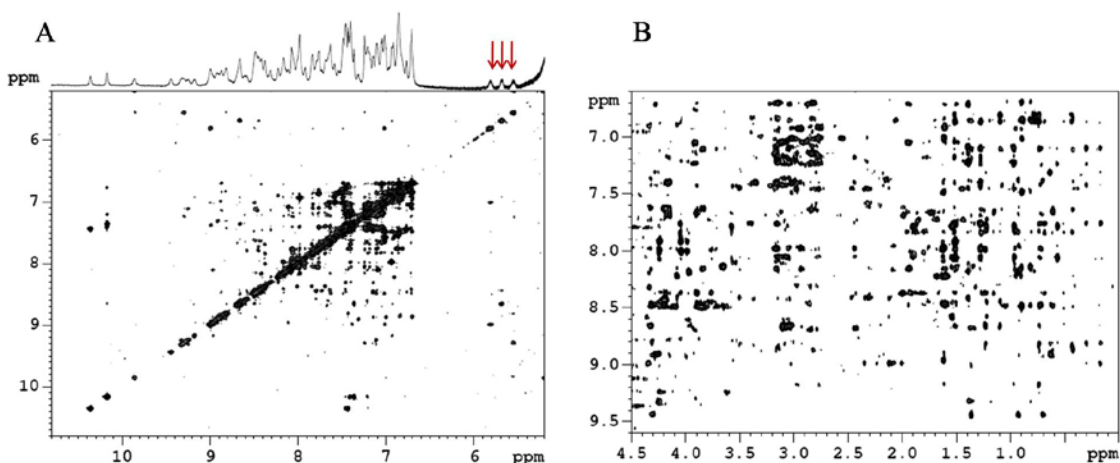
Figure 4.6 2D NOESY spectrum of hTIAR RRM2 with 1D projection (A) and (B) side chain to H_N region. Red arrows indicate the resonances of residues in β sheet conformation. Protein is in 20 mM sodium phosphate buffer (pH 7.0) containing 100 mM KCl, 1 mM DSS.

4.4.3   The binding preferences of RRM2 to short RNA substrates

The initial screening for the binding preferences of TIAR RRM2 to short RNA sequences was performed utilizing a Tris nondenaturing acrylamide gel (Figure 4.7). In this system, the migration of the protein is driven by the protein intrinsic positive charges in the applied buffer and gel condition. The disappearance of the protein band in the RRM2-rU$_{20}$ pair sample indicates the mobility pattern changes upon binding, specifically due to the neutralization of the protein charges by the negatively charged RNA molecule. The protein band did not show any changes in other lanes containing RRM2 with rU$_5$, rA$_5$ and rC$_5$. This indicates either weak interaction or no interaction. TIA-1a RRM2 also showed the same mobility pattern suggesting they share similar binding preference for the tested RNA substrates.
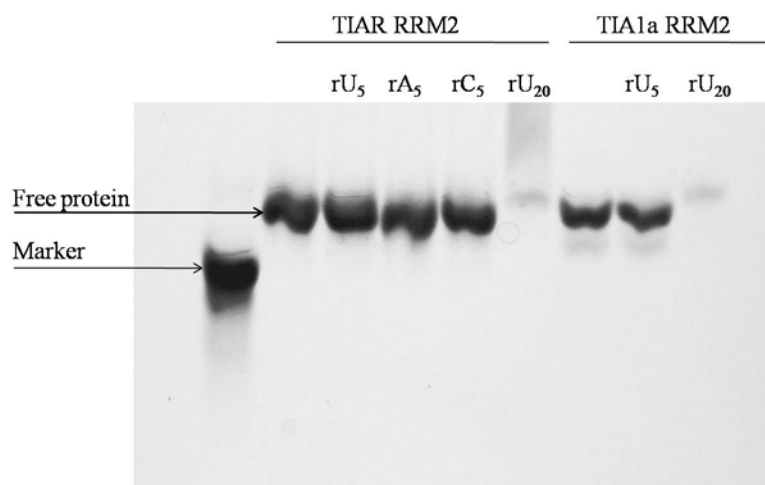
Figure 4.7 Tris nondenaturing 14% PAGE of TIAR and TIA-1a RRM2 with rU_5, rA_5, rC_5 and rU_{20}.

The application of nondenaturing gels for studying molecular association is limited by the fact that the formation of the complex is accompanied by the charge neutralization, so that the complex may not enter the gel. This can complicate the interpretation of the experimental data. The problem can be avoided using ITC, because the calorimetric measurements of ITC can be done in solution and so do not require separation of the free and bound molecules.

ITC experiments of TIAR RRM2 with various short RNA sequences showed that the protein binds specifically to rU_5, has weak interaction with rC_5 but does not bind to rA_5 (Figure 4.8). These results differentiate the affinity of RRM2 towards different RNA sequences which were not evident from the gel experiment. The TIAR RRM2-rU_5 binding event is exothermic as shown by the heat released upon addition of rU_5. Analysis of the titration plot resulted in a dissociate constant K_d of 8.7 μM and n of 0.87, which suggests a 1:1 binding stoichiometry. The results are in agreement with a previous report (*55*).
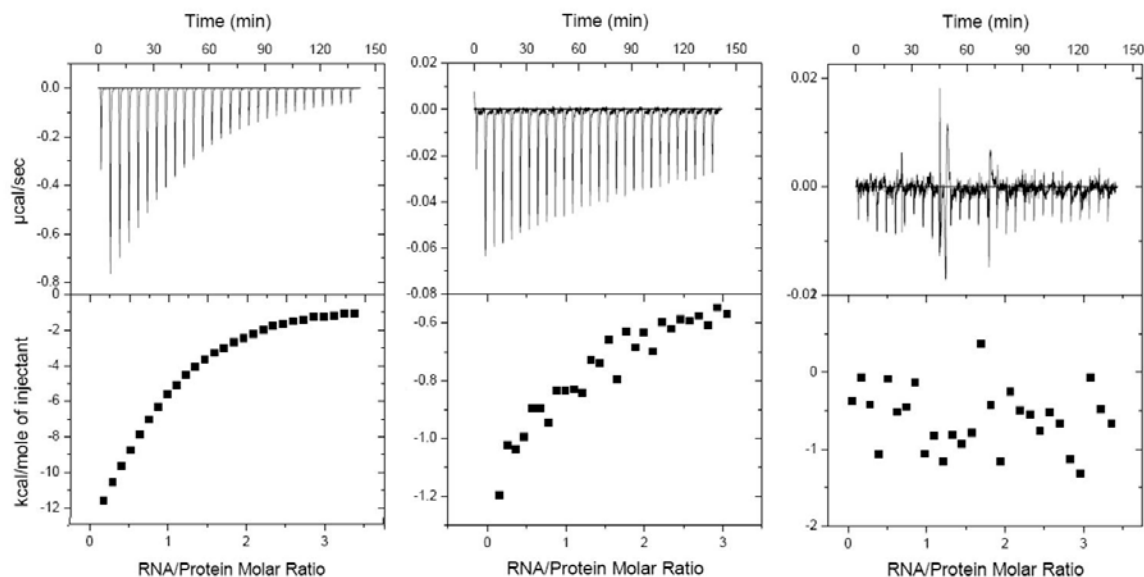
Figure 4.8 ITC isotherm from experiment for rU_5 (left), rC_5 (middle) and rA_5 (right) titrated into TIAR RRM2.

4.4.4 The length of oligoU sequence determines the binding ratio

The affinity of RRM2 and rU_{20} is much higher based on the ITC isotherm (Figure 4.9A). A K_d of 250 nM and more interestingly three identical binding sites on rU_{20} were observed. This indicates each RRM2 domain requires 6-7 nucleotides as a binding site. To verify this finding, we performed ITC experiments on oligoU RNA with difference lengths. The isotherm of rU_{15} titrated into RRM2 yielded an n of 0.52 ± 0.01, corresponding to a binding ratio of 2:1 (RRM2: rU_{15}) (Figure 4.9B). For the titration of RRM2 with rU_{12} (Figure 4.9C), fitting the titration with a one binding site mode produced an n of 0.73 ± 0.02, suggesting an average of 1.3 binding sites was provided by the 12 nucleotides. Two scenarios are proposed to explain this result. In the first model, one or two RRM2s can bind to the RNA sequence dependent of the position of the contact. In the second model, one RRM2 binds to the RNA sequence and a second RRM2 only par-

tially binds to the RNA with a lower affinity. However, both models need to be rationalized by more experimental data.

The β sheet surface of a single RRM has been reported previously to bind to variable numbers of nucleotides dependent on the sequence and accessibility of the RNA segment (*75-79*). Analysis of the complex of nucleolin RRM12 with a target stem-loop RNA showed that RRM1 contacts three and RRM2 contacts two nucleotides on each side of the loop. While in the complex of the spliceosomal U2B" protein associated with its cognate RNA, eight nucleotides are involved in the binding. In our study, the results from titration of RRM2 with rU_{20}, rU_{15} and rU_{12} showed that the oligoU sequences allow the binding of two or three RRM2s with each RRM2 requiring 5-7 nucleotides. These results inspired us to investigate the binding stoichiometry of RRM2 and its target viral RNA sequence, WNV 3' (-) SL RNA.
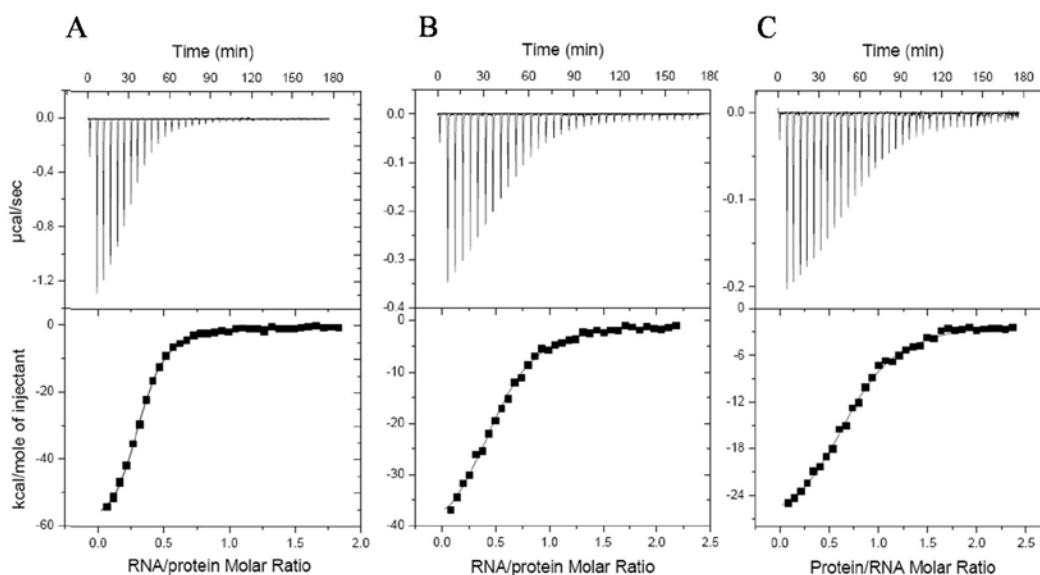
Figure 4.9 Isotherm from ITC experiment for rU_{20} (A), rU_{15} (B) and rU_{12} (C) titrated into TIAR RRM2. Titration was fitted using a one binding site mode (bottom panel).

In this study, in addition to the full length RNA target SL75, a 40 bases sequence (SL40) was designed to include the two loop region sequences and maintain the loop conformation. A 20 base single stranded RNA substrate containing the two loop sequences was also selected (Figure 4.2).

The ITC experiment on RRM2_SL20 (Figure 4.10) yielded a K_d of 410 nM which indicates a slightly lower affinity compared to the RRM2-rU_{20} pair. This result agrees with the notion that TIAR has a higher affinity for U-rich than AU-rich sequences (80). The binding site n is 0.3 per protein molecule, equivalent to the result in RRM2- rU_{20}. This again suggests that the 20 RNA bases can provide three binding sites. Alternatively, RRM2 may trimerize via protein-protein contacts when bound to an RNA sequence of 20 bases.

To verify if RRM2 can exist as a trimer when bound to rU_{20} or SL20, the protein was chemically crosslinked in the presence of SL20 prior to SDS PAGE analysis. As shown in Figure 4.11, RRM2 showed evidence of dimer formation when no RNA in present, but a trimer and even higher order oligomers are detected when bound to SL20. A gel shift assay was also applied to SL20 incubated with increasing amounts of RRM2. The migration of SL20 was altered upon adding RRM2, and a thinner band was observed. At a 3-fold excess of RRM2, the binding was near completion. This result supports the hypothesis that three RRM2s bind to a single molecule of SL20.

Figure 4.10 Isotherm from ITC experiment for SL20 titrated by RRM2, Titration was fitted using a one set of binding site mode (bottom panel).
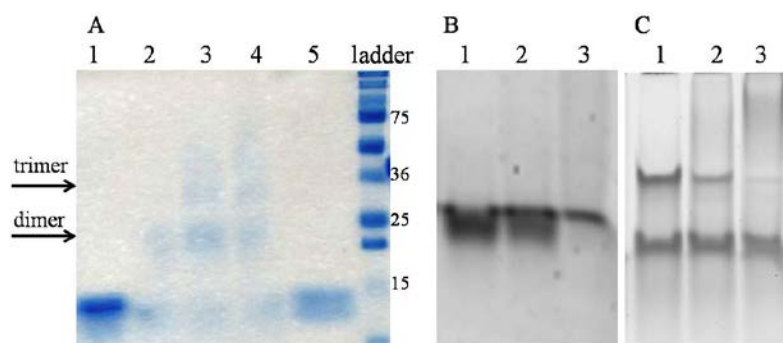
Figure 4.11 (A) A 13.5% SDS PAGE analysis of RRM2 crosslinked by DSP: Lane 1, RRM2 without crosslinking; Lane 2, RRM with crosslinking; Lane 3, Crosslinked RRM2 with equal molar of SL20; Lane 4, Crosslinked RRM2 with SL20 in molar ratio of 3 to 1. SYBR green stained 12% nondenaturing acrylamide gel for WNV SL20, (B) and SL40, (C) with increasing amount of TIAR RRM2. Lane 1: free SL20 (B1) or SL40 (C1), SL20 (B2) and SL40 (C2) with equal molar of RRM2; SL20 (B3) and SL40 (C3) with 3-fold excess of RRM2 (B3).

4.4.5 *The binding of TIAR RRM2 to SL40 and SL75*

The titration of RRM2 with SL40 surprisingly showed no significant heat release, which suggests there is no binding between these two partners. A gel shift assay was conducted to verify this result. The SL40 exhibited two bands in its free form. The band at the higher position corresponds to a dimer. When titrating with increasing amounts of RRM2, the bottom band stayed unshifted, but the top band disappeared and a smeared band appeared at a higher position (Figure 4.11C). This result suggested the dimer is capable of binding RRM2 while the monomer does not bind to RRM2. To confirm this, the sample was denatured and refolded with heat-chill cycle prior to a native acrylamide gel. This resulted in the disappearance of the dimer band. Reevaluating the SL40 sequence with M-fold (*74*), four possible conformations were predicted with similar free energies (Figure 4.12), where three of the predicted conformations have no exposed loops with a UUAAU sequence. Therefore, this sequence was not further utilized.

The interaction of RRM2-WNV SL75 probed by ITC titration displayed an apparent endothermic profile (Figure 4.13B). This is likely due to the heat absorption of RNA refolding upon binding and dilution (Figure 4.13B), which renders the overall reaction to become entropy driven. The fitting of the final net heat associated with this protein-RNA interaction produced binding sites n of 0.33, consistent with that of the RRM2-rU$_{20}$ or RRM2-SL20 binding pairs. This result combined with the fact that RRM is a single stranded RNA binding protein, partially supports the hypothesis that the active binding sites on this large sequence are the loop regions that can spatially hold three RRMs.
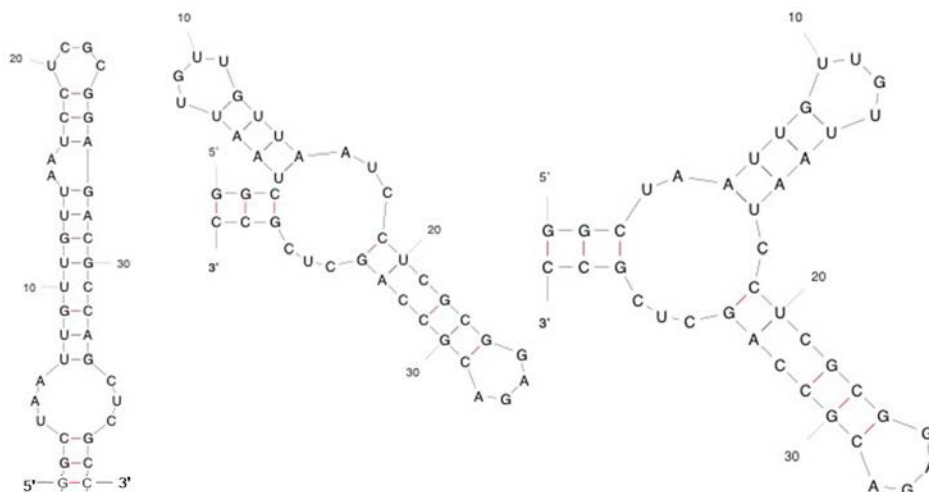
Figure 4.12 Other possible 2D structures of SL40 generated using Mfold (*74*).
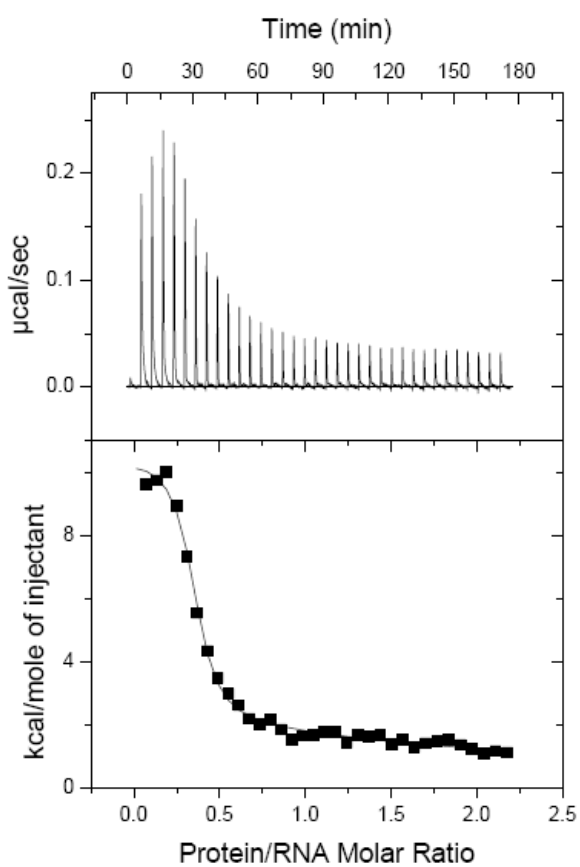
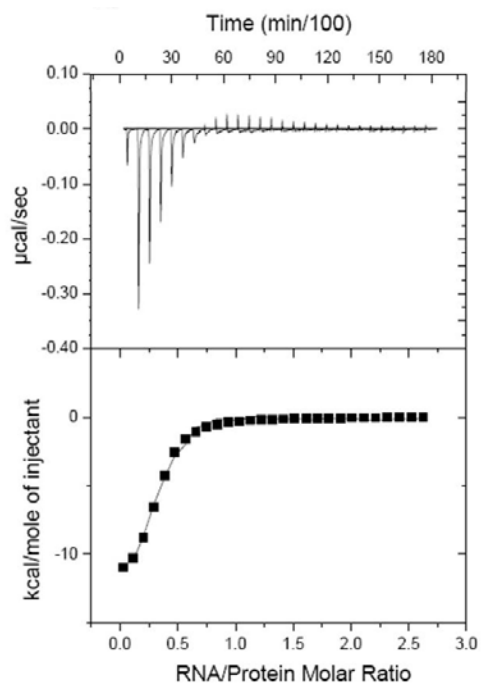**Figure 4.13** Isotherm from an ITC experiment for SL75 titrated with RRM2. The titration was fitted by a one set of binding site mode (bottom panel).

With the exception of RRM2-SL75, the binding interactions are all enthalpy driven. The RRM2-rU$_{20}$ complex exhibits the largest enthalpy of -5.4 $\times$ 10^4 kcal/mol and also shows the greatest entropy penalty of -150 kcal/mol·K. The large negative entropy reflects the loss of conformational freedom. SL75 experiences large conformational change upon binding to RRM2, which requires large heat input; however, the complex formation is entropy favored (Table 4.1).

Table 4.1 Thermodynamic parameters from ITC experiments for RRM2 with all the RNA substrates.

system	K$_d$ (nM)	n	Ratio (RRM2:RNA)	ΔH (kcal/mol)	ΔS (cal/mol·K)
RRM2_rU$_5$	8700±150	0.89±0.01	1:1	-24.0±0.3	-59.3
RRM2_ rU$_{12}$	478±34	0.75±0.02	1.3: 1	-32.7±0.9	-80.7
RRM2_rU$_{15}$	341±40	0.52±0.01	2:1	-45.0±1.6	-121
RRM2_rU$_{20}$	250±12	0.34±0.003	3:1	-53.6±0.7	-150
RRM2_SL$_{20}$	410±51	0.30±0.01	3:1	-12.7±0.5	-13.4
RRM2_SL$_{75}$	45±25	0.31±0.02	3:1	6.1±1.2	54.3

4.4.6   *RRM2-RNA binding followed with fluorescence anisotropy*

RRMs have been observed to be cooperative when interacting with target RNA in proteins containing multiple copies of RRMs. Our system focuses on a single RRM2 that contributes most to the binding affinity. However, multiple identical binding sites are present in the longer RNA sequences. Fluorescence experiments were carried out to further investigate the binding interaction.

Fluorescence anisotropy (r) is a measurement of fluorescence depolarization caused by rotational diffusion of the fluorophore during the excited life time. Briefly, when the fluorophore

is irradiated with plane-polarized light, only molecules that are properly aligned can be excited to emit fluorescence. If, during the fluorescence life time, the molecule rotates, the emitted radiation will be depolarized to a degree that depends on the extent of rotation. Therefore a decrease in the rotational diffusion results in an increase in anisotropy. In practice, anisotropy is calculated by measuring the vertical and horizontal fluorescence emission components (I_{VV} and I_{VH}) when the sample is excited with vertically polarized light based on the following equation:

$$r = \frac{I_{vv} - GI_{VH}}{I_{vv} - 2GI_{VH}}$$

Where G is the G factor, $G = I_{HV}/I_{HH}$. I_{VV} is the light intensity with the excitation and emission polarizers mounted vertically. I_{HH} is the intensity with the excitation and emission polarizers mounted horizontally. I_{HV} when a horizontal excitation polarizer and a vertical emission polarizer are used; I_{VH} is the intensity when a vertical excitation polarizer and horizontal emission polarizer are used.

Fluorescence anisotropy has been widely used in studying molecular associations, where the measured anisotropy represents a weighted average anisotropy of the free and bound fluorophore. When the relative quantum yields of the free and bound fluorophores are identical, the fraction of bound fluorophore α can be calculated using the equation:

$$\alpha = \frac{A - A_f}{A_b - A_f}$$

When the yields are not identical, α can be corrected as follows:

$$\frac{A - A_f}{(A - A_f) + (F_b/F_c)(A_b - A)}$$

Where A is the observed anisotropy; A_f is the free and A_b is the bound fluorophore; F_b and F_c are the fluorescence intensities of the bound and free fluorophore, respectively (*81*).

We measured the fluorescence anisotropy under high and low salt conditions. Under both conditions, the addition of RRM2 to labeled rU_{20} produces an increase in anisotropy. Under low salt conditions, the appearance of the titration curve indicates cooperative binding (*82*). In addition, half saturation appears at lower RRM2 concentrations indicating a higher binding affinity compared to the binding under high salt conditions. Applying Hill analysis to both titrations, we obtained a Hill coefficient h of 0.75±0.2 for the titration in high salt buffer but a h of 2.1±0.2 for the titration under low salt conditions. These results suggest the binding is cooperative.
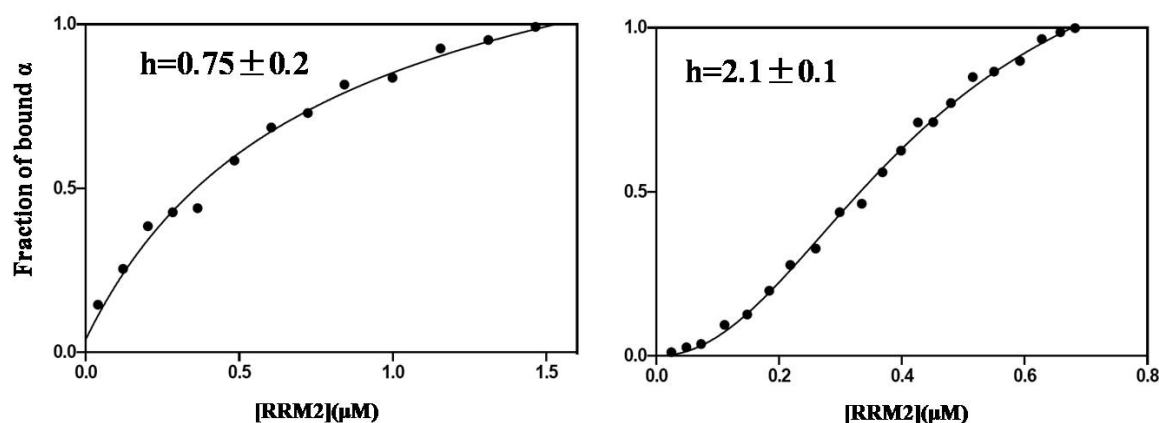
Figure 4.14 Fluorescence anisotropy titration of F-rU_{20} with TIAR RRM2 under typical buffer conditions (A) and low salt buffer conditions (B). The calculated fraction of bound F-rU_{20} was plotted against the total RRM2 concentration and the data were fitted using specific binding with hill plot (bottom panel) in Graphpad Prism.

4.4.7 *Electrostatic potential properties of the three RRMs of TIAR*

In a protein that contains multiple RRMs, the RRMs can form a larger platform or cleft during the recognition to their target DNA or RNA sequence. In TIAR, the three RRMs share

86

similar structures, but contribute differently when binding to U rich or AU-rich RNA sequences, with RRM2 possessing the highest affinity (*57*). This suggests that RRM2 has unique properties. Bauer's *ab initio* model based on SAXS data suggested a "V" shape of RRM123 (*83*). However, the precise arrangement of the three RRMs is not clear because the structure of the wild type protein in not known. Sequence alignment reveals that these RRMs have only about 40% sequence similarity. Electrostatic potential calculations were carried out to examine the charge distribution on each RRM to understand their different contributions to RNA binding. In addition, this analysis may also provide insights about protein-protein contacts.
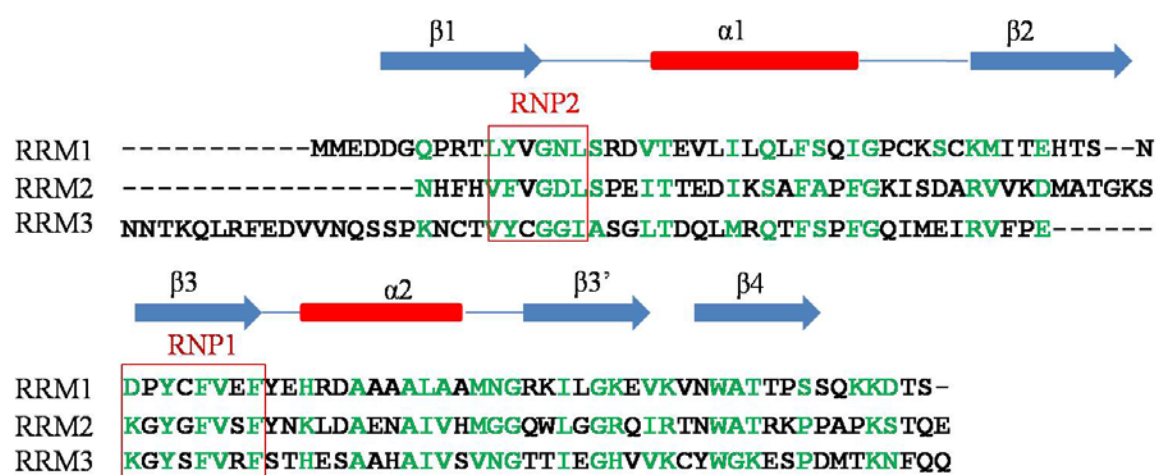
Figure 4.15 Sequence alignments of RRM1, RRM2 and RRM3. Identical, strong similar and weak similar residues are in green. Secondary structure elements are depicted with blue arrows (β strand) and red cylinders (α helix). Red boxes highlight the two known conserved peptide sequences.

The potential maps show that positive and negative charges are almost equally distributed on the α-helical surfaces and the loops that connect secondary structure elements of all three RRMs (Figure 4. 15). However, the charge distribution on the β sheet surface differs with the

RRM2 surface being more positive than those of RRM1 and RRM3. This result suggests electrostatic forces contribute significantly to RRM2-RNA recognition and emphasize the major role of RRM2 in RNA bindings. The result also suggests that in high salt conditions, protein-protein interactions can be facilitated. This allows multiple RRM2s to bind simultaneously to identical binding sites on longer oligoU RNAs as observed in ITC study.
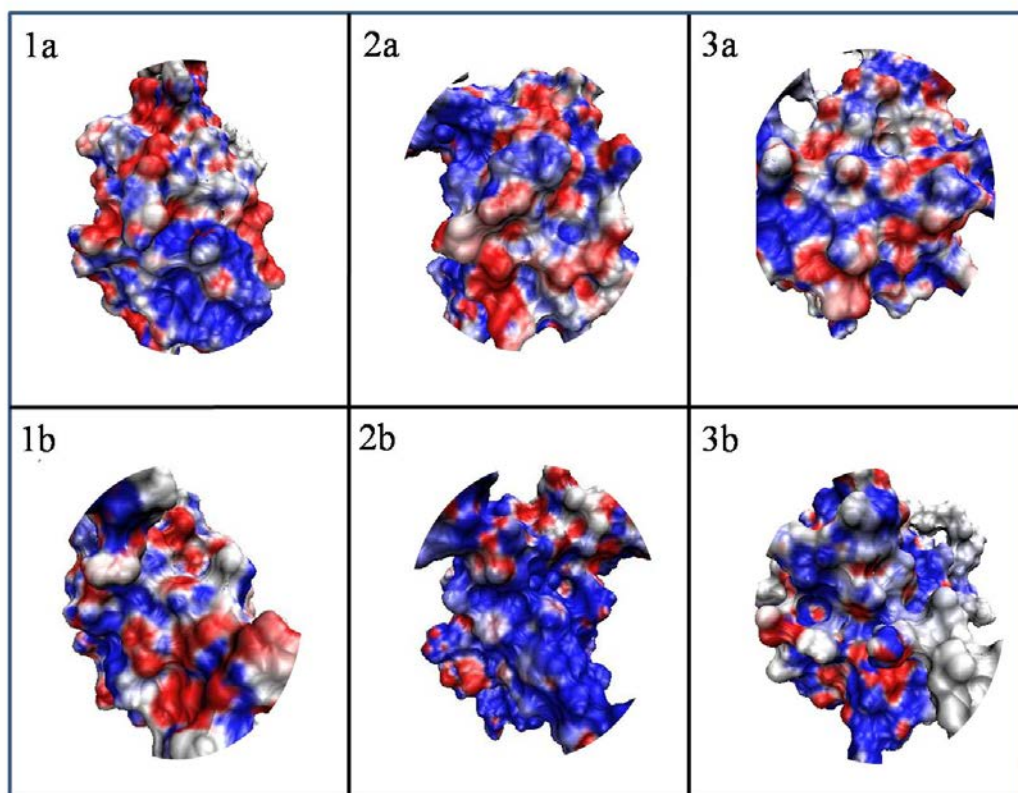
Figure 4.16 Electrostatic potential maps of TIAR RRM1 (1a and 1b); RRM2 (2a and 2b) and RRM3 (3a and 3b). a and b indicate the surface of the α helix and the β sheet of each RRM, respectively. Only the cores of the surfaces are shown.

In addition, repulsion may exist between the three RRMs in the wild type hTIAR due to the relatively positive charges on their β sheets. As a result of this, the three domains might face in opposite directions which would place RRM2 toward the target RNA. To test this hypothesis, investigation of the wild type protein must be conducted.

4.5   Conclusion

Applying a combination of computational and experimental techniques including electro-phoresis, calorimetry and spectroscopy, we evaluated the binding interaction of hTIAR RRM2 to various RNA sequences.

RRM2 was shown to bind U rich sequences with higher affinity than AU rich sequences based on comparising the binding to oligoU and to a single stranded AU rich RNA target. The longer oligoU allowed simultaneous binding of multiple RRM2s. The results of the studies with rU_5, rU_{12}, rU_{15} and rU_{20} demonstrated that a single RRM recognizes 5-7 nucleotides.

Fluorescence anisotropy experiments conducted in buffers with different salt concentra-tions revealed cooperation between multiple RRM2s when bound to oligoU.

Electrostatic potential maps revealed that the β sheet surface of RRM2 is highly positive-ly charged. This may account for its dominant role in RNA binding. A hypothesis on the ar-rangement of three RRM domains based on electrostatics was also proposed.

4.6   Future Work

The binding isotherm of wild type hTIAR and various RNA sequences must be investi-gated using ITC to explore the biologically relevant binding affinity, stoichiometry and thermo-dynamic parameters.

To map the binding sites on SL20, NMR spectroscopy can be applied to ^{15}N or/and ^{13}C labeled RRM2. Titration of labeled RRM2 with unlabeled SL20 is expected to verify the binding stoichiometry.

A model of the arrangement of the RRMs of TIAR is needed to rationalize the data.

The binding of RRM2 to SL75 will be evaluated by molecular docking using Haddock, a program suitable for protein-RNA interactions. A predicted tertiary structure of the RNA will be used.

4.7   References

1.   Moras, D. (1992) Aminoacyl-tRNA synthetases, *Curr.Opin.Struct.Biol. 2*, 138-142.
2.   Siomi, H., and Dreyfusst, G. (1997) RNA-binding proteins as regulators of gene expression, *Current Opinion in Genetics & Development 7*, 9.
3.   Barton, D. J., Black, E. P., and Flanegan, J. B. (1995) Complete replication of poliovirus in vitro: preinitiation RNA replication complexes require soluble cellular factors for the synthesis of VPg-linked RNA, *J. Virol. 69*, 5516–5527.
4.   Barton, D. J., Sawicki, S. G., and Sawicki, D. L. (1991) Solubilization and immunoprecipitation of alphavirus replication complexes, *J. Virol. 65*, 1496–1506.
5.   Hayes, R. J., and Buck, K. W. (1990) Complete replication of a eukaryotic virus RNA in vitro by a purified RNA-dependent RNA polymerase, *Cell 63*, 363–368.
6.   Landers, T. A., Blumenthal, T., and Weber, K. (1974) Function and structure in ribonucleic acid phage Qb ribonucleic acid replicase., *J. Biol. Chem. 249*, 5801–5808.
7.   Tan, J., and Frankel, A. D. (1995) Structural variety of arginine-rich RNA-binding peptides, *Proc. Natl. Acad. Sci. U S A 92*, 5282-5286.
8.   De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3-RNA recognition element, *Science 279*, 384-388.
9.   Qubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92A resonlution of the RNA-binding domain of the U1A splicsomal protein complexed with an RNA hairpin *Nature 372*, 432-438.
10.   Nagai, K. (1992) RNA-protein interactions, *Current Opinion in Structural Biology 2*, 131-137.
11.   Scherly, D., Boelens, W., Dathan, D. A., Van Venrooij, W. J., and Mattaj, I. W. (1990) Major Determinants of the Specificity of Interaction between Small Nuclear Ribonucleoproteins U1A and U2B" and Their Cognate RNAs, *Nature 345*, 502-506.
12.   Witherell, G. W., Gott, J. M., and C., U. O. (1991) Specific Interaction Between RNA Phage Coat Proteins and RNA., *Prog Nucleic Acid Res Biol 40*, 185-220.
13.   Scherly, D., Dathan, N. A., Boelens, W., van Venrooij, W. J., and Mattaj, I. W. (1990) The U2B" RNP motif as a site of protein-protein interaction., *EMBO J. 9*, 3675-3681.

14.   Query, C. C., Bentley, R. C., and Keene, J. D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein, *Cell 57*, 89-101.

15.   Draper, D. E. (1999) Themes in RNA-protein recognition, *J. Mol. Biol. 293*, 255-270.

16.   Morozova, N., Allers, J., Myers, J., and Shamoo, Y. (2006) Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures, *Bioinformatics 22*, 2746-2752.

17.   Stefl, R., Skrisovska, L., and Allain, F. H.-T. (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle, *EMBO Reports 6* 33-38.

18.   McDonald, I. K., and Thornton, J. M. (1994) Satisfying hydrogen-bonding potential in proteins *J. Mol. Biol. 238*, 777-793.

19.   Allers, J., and Shamoo, Y. (2001) Structure-based analysis of Protein-RNA interactions using the program ENTANGLE, *J. Mol. Biol. 311*, 75-86.

20.   Lejeune, D. (2005) Protien-nuclei acid recognition:statistical analysis of atomic interactions and influence of DNA structure, *Stucture. Proteins 61*, 258-271.

21.   Krämer, A., Keller, W., Appel, B., and Lührmann, R. (1984) The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors., *Cell 38*, 299–307.

22.   Günzl, A., and Bindereif, A. (1999) Oligonucleotide-Targeted RNase H Protection Analysis of RNA-Protein Complexes *Methods Mol Biol: RNA-Protein Interaction Protocols 118*, 93-103.

23.   Filette, M. D., Ulbert, S., Diamond, M. S., and Sanders, N. N. (2012) Recent progress in West Nile virus diagnosis and vaccination, *Vet Res. 43*, 16.

24.   Lanciotti, R. S., Roehrig, J. T., Deubel, V., Smith, J., Parker, M., and al, e. (1999) Origin of theWest Nile virus responsible for the outbreak of encephalitis in the Northeastern United States, *Science 286*, 2333-2337.

25.   Wengler, G., and Gross, H. J. (1978) Studies on virus-specific nucleic acids synthesized in vertebrate and mosquito cells infected with flaviviruses, *Virology 89*, 423–437.

26.   Cleaves, G. R., Ryan, T. E., and Schlesinger, R. W. (1981) Identification and characterization of type 2 dengue virus replicative intermediate and replicative form RNAs., *Virology 111*, 73–83.

27.   Brinton, M. A. (2002) The Molecular Biology of West Nile Virus: A New Invader of the Western Hemisphere, *Annu. Rev. Microbiol. 56*, 371-402.

28.   Brinton, M. A., Fernandez, A. V., and Dispoto, J. H. (1986) The 30-nucleotides of flavivirus genomic RNA form a conserved secondary structure, *virology 153*, 113–121.

29.   Brinton MA, and JH., D. (1988) Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA, *Virology 162*, 10.

30.   Thurner C, Witwer C, and Hofacker IL. (2004) Conserved RNA secondary structures in Flaviviridae genomes. , *J Gen Virol 85*, 12.

31. Shi, P.-Y., Li, W., and Brinton, M. A. (1996) Cell proteins bind specifically to West Nile virus minus-strand 30 stem-loop RNA., *J. Virol. 70*, 6278–6287.

32. Cahour, A., Pletnev, A., Vazeille-Falcoz, M., Rosen, L., and C.-J, L. (1995) Growth-restricted dengue virus mutants containing deletions in the 59 noncoding region of the RNA genome., *Virology 207*, 68-76.

33. Kawakami, A., Tian, q., duan, X., Streuli, M., Schlossman, S. F., and Anderson, P. (1992) Identification and functional characterizationof a TIA-1-related nucleolysin, *Proc. Natl. Acad. Sci. USA 89*, 8681-8685.

34. Brand, S., and Bourbon, H. M. (1993) The developmentally-regulated Drosophila gene rox8 encodes an RRM-type RNA binding protein structurally related to human TIA-1-type nucleolysins, *Nucleic Acids Res. 21*, 3699-3704.

35. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., and Cooper, J. e. a. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans, *Nature 368*, 32-38.

36. Zhang, T., Delestienne, N., Huez, G., Kruys, V., and Gueydan, C. (2005) Identification of the sequence determinants mediating the nucleo-cytoplasmic shuttling of TIAR and TIA-1 RNA-binding proteins, *J Cell Sci. 118*, 5453-5463.

37. Zhu, H., Hasman, R. A., Young, K. M., Kedersha, N. L., and Lou, H. (2003) U1 snRNP-dependent function of TIAR in the regulation of alternative RNA processing of the human calcitonin/CGRP pre-mRNA, *Mol Cell Biol 23*, 5959-5971.

38. Förch, P., and Valcárcel, J. (2001) Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1, *Apoptosis 6*, 463-468.

39. Aznarez, I., Barash, Y. S., O, He, D., Zielenski, J., Tsui, L. C., Parkinson, J., Frey, B. J., Rommens, J. M., and Blencowe, B. J. (2008) A systematic analysis of intronic sequences downstream of 5' splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation, *Genome Res. 18*, 1247-1258.

40. Förch, P., Puig, O., Kedersha, N., Martínez, C., Granneman, S., Séraphin, B., Anderson, P., and Valcárcel, J. (2000) The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing, *Mol Cell 6*, 1089-1098.

41. Förch, P., Puig, O., Martínez, C., Séraphin, B., and Valcárcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites, *EMBO J. 21*, 6882-6892.

42. Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stévenin, J., and Del Gatto-Konczak, F. (2001) TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs, *J Biol Chem 276*, 40638-40646.

43. Suswam, E. A., Li, Y. Y., Mahtani, H., and King, P. H. (2005) Novel DNA-binding properties of the RNA-binding protein TIAR, *Nucleic Acids Res 33*, 4507-4518.

44. Mazan-Mamczarz, K., Lal, A., Martindale, J. L., Kawai, T., and Gorospe, M. (2006) Translational repression by RNAbinding protein TIAR, *Mol Cell Biol 26*, 2716-2727.

45. López de Silanes, I., Galbán, S., Martindale, J. L., Yang, X., Mazan-Mamczarz, K., Indig, F. E., Falco, G., Zhan, M., and Gorospe, M. (2005) Identification and functional outcome of mRNAs associated with RNA-binding protein TIA-1, *Mol Cell Biol. 25*, 9520-9531.

46. Kedersha, N. L., Gupta, M., Li, W., Miller, I., and Anderson, P. (1999) RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules, *J Cell Biol 147*, 1431-1442.

47. Li, W., Simarro, M., Kedersha, N., and Anderson, P. (2004) FAST Is a Survival Protein That Senses Mitochondrial Stress and Modulates TIA-1-Regulated Changes in Protein Expression, *Mol. Cell. Biol. 24*, 10718-10732.

48. Beck, A. R. P., Miller, I. J., Anderson, P., and Streuli, M. (1998) RNA-binding protein TIAR is essential for primordial germ cell development, *Proc. Natl. Acad. Sci. USA 95*, 2331–2336.

49. Damgaard, C. K., and Lykke-Andersen, J. (2011) Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR, *Genes Dev. 25*, 2057-2068.

50. Tian, Q., Streuli, M., Saito, H., Schlossman, S. F., and Anderson, P. (1991) A polyadenylate binding protein localized to the granules of cytolytic lymphocytes induces DNA fragmentation in target cells., *Cell 67*, 629-639.

51. Prusiner, S. B. (1989) Scrapie prions, *Annu. Rev. Microbiol. 43*, 345-374.

52. Burd, C. G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins, *Science 265*, 615-621.

53. Cléry, A., Blatter, M., and Allain, F. H. (2008) RNA recognition motifs: boring? Not quite, *Curr Opin Struct Biol 18*, 290-298.

54. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2006) Sequencespecific binding of single-stranded RNA: is there a code for recognition?, *Nucleic Acids Res 34*, 4943-4959.

55. Kuwasako, K., Takahashi, M., Tochio, N., Abe, C., Tsuda, K., Inoue, M., Terada, T., Shirouzu, M., Kobayashi, N., Kigawa, T., Taguchi, S., Tanaka, A., Hayashizaki, Y., Güntert, P., Muto, Y., and Yokoyama, S. (2008) Solution structure of the second RNA recognition motif (RRM) domain of murine T cell intracellular antigen-1 (TIA-1) and its RNA recognition mode, *Biochemistry 47*, 6437-6450.

56. Kumar, A. O., Swenson, M. C., Benning, M. M., and Kielkopf, C. L. (2008) Structure of the central RNA recognition motif of human TIA-1 at 1.95A resolution, *Biochem Biophys Res Commun. 367*, 813-819.

57. Dember, L. M., Kim, N. D., Liu, K.-Q., and Anderson, P. (1996) Individual RNA Recognition Motifs of TIA-1 and TIAR Have Different RNA Binding Specificities, *J. Biol. Chem. 271*, 2783-2788.

58. Kim, H. S., Headey, S. J., Yoga, Y. M., Scanlon, M. J., Gorospe, M., Wilce, M. C., and Wilce, J. A. (2013) Distinct binding properties of TIAR RRMs and linker region., *RNA Biol. 10*, 1-11.

59.   Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature 372*, 432–438.

60.   Dominguez, C., Fisette, J. F., Chabot, B., and Allain, F. H. (2010) Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs, *Nat Struct Mol Biol 17*, 853-861.

61.   Fialcowitz-White, E. J., Brewer, B. Y., Ballin, J. D., Willis, C. D., Toth, E. A., and Wilson, G. M. (2007) Specific protein domains mediate cooperative assembly of HuR oligomers on AU-rich mRNA-destabilizing sequences., *J Biol Chem 282*, 20948-20959.

62.   Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J. 272*, 2118-2131.

63.   Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein, *Nature 398*, 579–585.

64.   Kim, H. S., Kuwano, Y., Zhan, M., Pullmann, R. J., Mazan-Mamczarz, K., Li, H., Kedersha, N., Anderson, P., Wilce, M. C., Gorospe, M., and Wilce, J. A. (2007) Elucidation of a C-rich signature motif in target mRNAs of RNA-binding protein TIAR, *Mol Cell Biol. 27*, 6806-6817.

65.   Emara, M. M., and Brinton, M. A. (2007) Interaction of TIA-1/TIAR with West Nile and dengue virus products in infected cells interferes with stress granule formation and processing body assembly, *Proc Natl Acad Sci U S A 104*, 9041-9046.

66.   Li, W., Li, Y., Kedersha, N., Anderson, P., Emara, M., Swidered, K. M., Moreno, G. K., and Brinton, M. N. (2002) Cell Proteins TIA-1 and TIAR Interact with the 3 Stem-Loop of the West Nile Virus Complementary Minus-Strand RNA and Facilitate Virus replication, *JOURNAL OF VIROLOGY 76*, 11989-12000.

67.   Emara, M. M., Liu, H., Davis, W. G., and Brinton, M. A. (2008) Mutation of mapped TIA-1/TIAR binding sites in the 3' terminal stem-loop of West Nile virus minus-strand RNA in an infectious clone negatively affects genomic RNA amplification, *J Virol. 82*, 10657-10670.

68.   Barreau, C., Paillard, L., and Osborne, H. B. (2005) AU-rich elements and associated factors: are there unifying principles?, *Nucleic Acids Res. 33*, 7138–7150.

69.   Chen, C. Y., Xu, N., and Shyu, A. B. (1995) mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translatio, *Mol Cell Biol. 15*, 5777-5788.

70.   Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res 32*, W668-673.

71. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations, *Nucleic Acids Res 32*, W665-W667.

72. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - Visual Molecular Dynamics, *J. Molec. Graphics 14*, 33-38.

73. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility, *J Comput Chem. 30*, 2785-2791.

74. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res. 31*, 3406-3415.

75. Calero, G., Wilson, K. F., Ly, T., Rios-Steiner, J. L., Clardy, J. C., and Cerione, R. A. (2002) Structural basis of m7GpppG binding to the nuclear cap-binding protein complex, *Nat Struct Biol. 9*, 912-917.

76. Mazza, C., Segref, A., Mattaj, I. W., and Cusack, S. (2002) Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex, *EMBO J. 21*, 5548-5557.

77. Allain, F. H., Bouvet, P., Dieckmann, T., and Feigon, J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin, *EMBO J. 19*, 6870-6881.

78. Price, S. R., Evans, P. R., and Nagai, K. (1998) Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA, *Nature 394*, 645-650.

79. Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A., and Feigon, J. (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target, *J Mol Biol. 337*, 799-816.

80. Kim, H. S., Wilce, M. C., Yoga, Y. M., Pendini, N. R., Gunzburg, M. J., Cowieson, N. P., Wilson, G. M., Williams, B. R., Gorospe, M., and Wilce, J. A. (2011) Different modes of interaction by TIAR and HuR with target RNA and DNA., *Nucleic Acids Res. 39*, 1117-1130.

81. Malencik, D. A., and Anderson, S. R. (1988) Associstion of Melittin with the Isolated Myosin Light Chain, *J Am Chem Soc 27*, 1941-1949.

82. Kowalczykowski, S. C., Paul, L. S., Lonberg, N., Newport, J. W., McSwiggen, J. A., and von Hippel, P. H. (1986) Cooperative and noncooperative binding of protein ligands to nucleic acid lattices: experimental approaches to the determination of thermodynamic parameters, *Biochemistry 25*, 1226-1240.

83. Bauer, W. J., Heath, J., Jenkins, J. L., and Kielkopf, C. L. (2012) Three RNA recognition motifs participate in RNA recognition and structural organization by the pro-apoptotic factor TIA-1, *J Mol Biol 415*, 727-740.

5   APPLICATION OF 2D SELECTIVE TOCSY IN ASSIGNMENT OF ADENINE H2 AND H8 RESONANCES IN OLIGONUCLEOTIDES USING HOMONUCLEAR LONG-RANGE COUPLING CONSTANTS

5.1   Preface

The DFT shielding and spin-spin coupling calculations on 2-endo deoxyadenosine were carried out by Alexander Spring. This work was published in *J. AM. CHEM. SOC.* (2009) *131*: 5380–5381.

5.2   Introduction

NMR structural studies of nucleic acids require a nearly complete and unambiguous assignment of the proton resonances. This includes the adenine H2 protons, which, especially in B-type DNA, are generally far from other protons, rendering their assignment cumbersome. In both A- and B-type nucleic acid helices, the H2 protons are located in the minor grooves, where they serve as useful monitors for binding in or near adenines in the DNA minor groove (*1*).

The conventional assignment method, which relies on Nuclear-Overhauser effects (NOEs) recorded in D$_2$O solution, generally does not allow complete assignment of the H2 resonances. In practice, these resonances are often assigned from their NOEs to the imino protons of A·U or A·T base pairs (*2-3*). This requires performing experiments in H$_2$O and the presence of detectable imino protons whose exchange with the solvent is sufficiently slow. For adenine residues in unpaired regions, other strategies are needed. Oligonucleotides that are ^{13}C-labeled allow magnetization transfer between the H2 and H8 protons via common couplings to intervening ^{13}C nu-

clei (*4-5*). However, this approach is hampered by rapidly relaxing carbons. Subsequently, triple resonance experiments using [^{13}C, ^{15}N]-labeled oligonucleotides were developed to correlate the H2 and H8 protons (H → ^{15}N → ^{13}C → H) (*6-7*). Such schemes feature increased sensitivity but require labeled oligonucleotides and use of H_2O as the solvent. The development of higher-sensitivity NMR probes and the use of low-artifact gradient-enhanced experiments enabled the assignments of H2 and H8 protons at natural abundance using Heteronuclear Multiple Bond Correlation (HMBC) experiments in D_2O (*8*). These experiments rely on a common heteronuclear long-range coupling of both H2 and H8 to C4. However, the inherently low sensitivity requires very long acquisition times and/or high sample concentrations.

The adenine H2 resonances are readily identified in D_2O solution, since they appear as narrow singlets with long longitudinal (T_1) and transverse (T_2) relaxation times due to their isolated location. The long T_1 is often a nuisance because it limits the repetition rate in many NMR experiments. However, the long T_2 permits the use of pulse sequences with longer durations, which aid in the detection of even small couplings.

Long-range six-bond homonuclear coupling constants have been reported previously for polycyclic aromatic systems (*9*). This encouraged us to explore whether there is an adenine H2-H8 homonuclear coupling that could be utilized to track H2 using H8.

5.3   Materials and Methods

5.3.1   NMR sample preparation

Adenosine 5' monophosphate (AMP) was prepared as 50 mM and 200 mM in 10 mM sodium phosphate buffer containing 0.5 mM EDTA in 99.996% D_2O, at pH* 6.6, 298 K for selective TOCSY and COSYLRQF experiments and for the XLOC experiment. A 0.5 mM DNA

decamer duplex with the sequence 5'-GTTCGAGACG-3', 5'-CGTCTCGAAC-3' and a 0.75 mM DNA hairpin with sequence 5'-ACTGCTACCCCTAGCAGT-3' were dissolved in the above mentioned NMR buffer containing 100 mM NaCl at pH* 6.6 and 50 mM NaCl at pH* of 6.84, respectively.

5.3.2 *NMR experiments*

NMR experiments were obtained on Bruker Avance 600 MHz (5 mm QXI probe) and 500 MHz (5 mm TBI, 5 mm TXI cryo probe) instruments. T_1 and T_2 relaxation times were determined using inversion recovery (*10-11*) and CPMG (*12*) experiments, respectively. Selective TOCSY experiments were acquired using a 4.5 - 6 ms E-burp excitation pulse centered in the aromatic region. Long-range COSY and XLOC experiments were recorded as described previously (*8, 13-14*).

5.3.3 *NMR simulations*

Simulation were carried out with NMR-SIM 3.1 (Bruker) for the AMP spin system using H2 = 8.17 ppm, T_1 = 6.0 s, T_2 = 1.6 s; H8 = 8.54 ppm, T_1 = 2.1 s, T_2 = 1.1 s.

5.3.4 *DFT shielding and spin-spin calculations*

2-endo deoxyadenosine was optimized with the HF/6-31G in Gaussian03 package (*15*). NMR shielding and spin-spin calculations were obtained using the B3LYP and 6-31G** basis set with the polarizable continuum model and the UA0 cavity for water effects (*16-17*).

5.4   Results and Discussion

5.4.1   Quantum-mechanical calculations of the 2-endo deoxyadenosine H2-H8 coupling

Small coupling constants, particularly when they are on the order of the line widths, are difficult to measure (*18-20*). Inspection of adenosine NMR spectra does not provide any obvious support for long-range H2-H8 couplings. Therefore, quantum-mechanical calculations were initially performed to predict the magnitude of the small long-range coupling, which was estimated to be -0.18 Hz (*15-17*). However, these calculations also forecasted several other nonvanishing couplings of H2 (to H3', H5', and H5") (Figure 5.1). In view of these results, experimental verification was clearly needed.
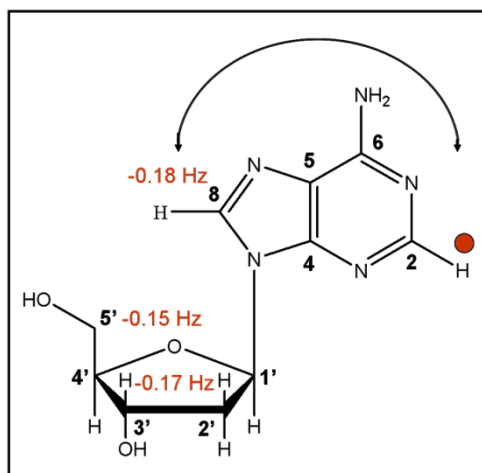

Figure 5.1 Coupling constants of 2-endo deoxyadenosine H2 to H8, H3' and H5' were calculated using B3LYP and 6-31G** basis set in Gaussian03. Coupling constant values are indicated in red.

5.4.2 Assessment of the amplitude of the AMP H2-H8 long-range coupling constant using NMR experiments and NMR sim simulation

The H2-H8 coupling was assessed using a long-range optimized COSY experiment on adenosine 5' monophosphate (AMP) (*21*). The coupling became apparent for long-range evolution delays larger than 0.10 s. We also noted a long-range COSY correlation for the H8 and H1' protons (Figure 5.2). The maximum cross-peak intensity was obtained when a delay of 0.75 s was applied, from which a coupling constant around 0.6 Hz is predicted using $J = 1/(2\tau)$.

NMR-SIM simulations of this experiment taking into account the relaxation were carried out with a coupling constant of 0.3 Hz and 0.5 Hz. The maximum cross-peak intensity appeared at a 0.75 s long-range delay when a 0.5 Hz coupling was simulated. This was consistent with NMR long-range optimized COSY result that indicated a H2-H8 coupling of ~0.5 Hz (Figure 5.3). Sørensen and co-workers introduced an innovative pulse program named XLOC that allows a direct measurement of small homonuclear couplings (*10, 13*). This approach disentangles small homonuclear J_{HH} couplings in the F2 dimension by the large one-bond J_{CH} coupling in F1. Using XLOC on the AMP sample yielded a coupling constant of 0.5 ± 0.2 Hz, in agreement with our previous results (Figure 5.4).
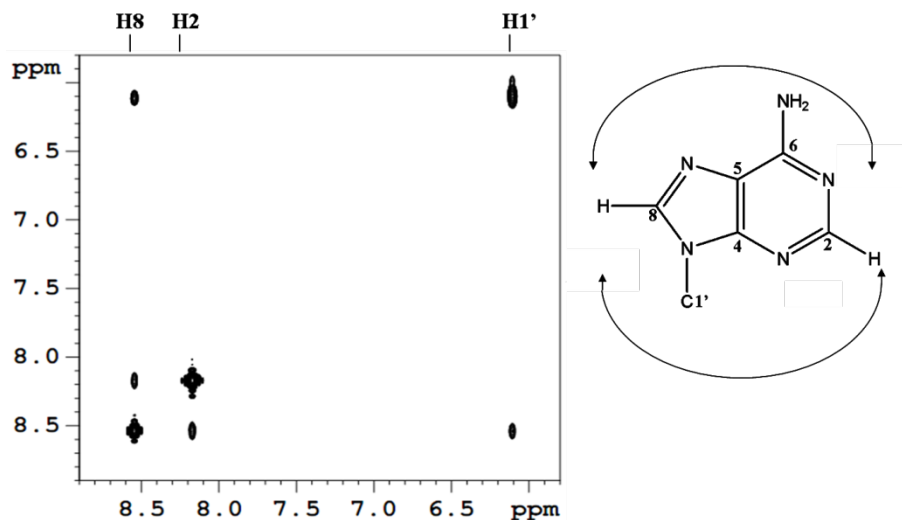
Figure 5.2 The 500 MHz long-range optimized COSY (cosylrqf) spectrum of a 50 mM 5' AMP sample recorded at 298 K with long-range evolution delay of 300 ms and a 10 s relaxation delay. A 1024 × 128 data point matrix was acquired in a spectral window of 3.0 × 3.0 ppm using 8 scans.
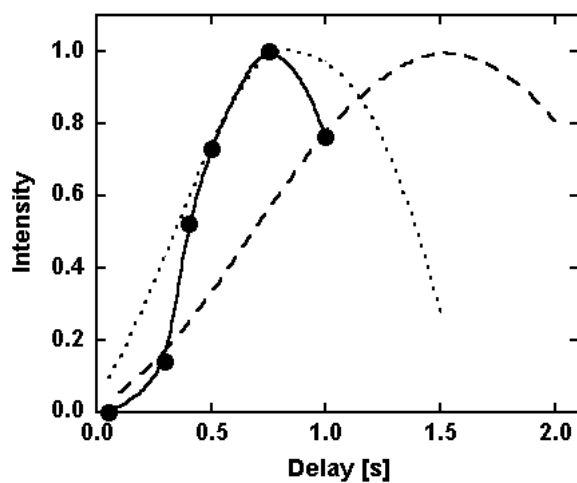

Figure 5.3 Dependence of the cross-peak intensity of 5'-AMP on the long-range evolution delay in the NMR cosylrqf experiments (solid line with dots). NMR sim simulation on 5'-AMP using pulse program cosylrqf for a coupling constant of 0.5 Hz (dotted line) and simulation for a coupling constant of 0.3 Hz (dashed line).
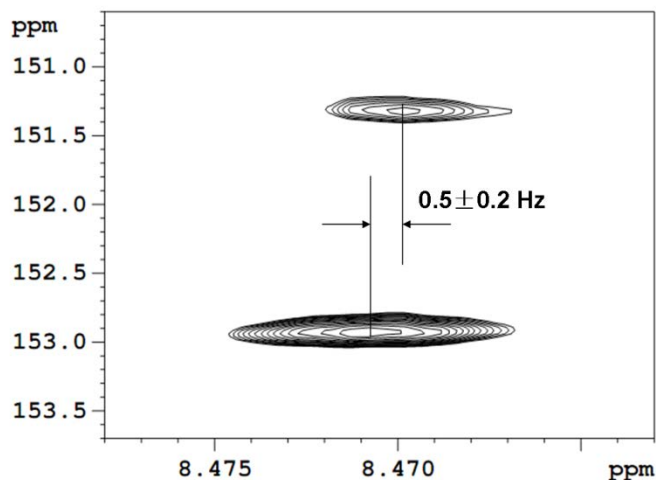
Figure 5.4 The 500 MHz expansion of the XLOC spectrum for the 200 mM AMP sample. The spectrum was recorded at 298 K using a first order low-pass J filter with a 1.5 s relaxation delay and an excitation delay of 0.343 s [10]. A 2K × 512 matrix was acquired in a spectral window of 4 × 90 ppm with 16 scans per increment. For processing, the window functions were cosine in T_1 and sine in T_2.

5.4.3   Application of the H2-H8 long-range coupling in DNA oligos

In addition to the apparent cross-peak of 5' AMP H2-H8 coupling in long-range optimized COSY experiment, a H2-H8 cross-peak in the total correlation spectroscopy (TOCSY) experiment also appeared at long mixing times (0.20 to 1.25 s). Such long durations are possible because of the slow relaxation behavior of the A H2 proton. This also permits the use of a selective TOCSY experiment that focuses the base region to probe the correlation of H2-H8 of A in DNA oligos. The selective TOCSY incorporating an E-Burp pulse for excitation of the base region limits the appearance of unwanted signals and permits the use of large receiver gains. For a 10-mer DNA duplex at mixing times larger than 0.25 s, a clear correlation of all of the H2 and H8 protons was observed (Figure 5.5). The cross-peak intensity remained nearly constant for mixing times between 0.40 and 1.0 s (Figure 5. 6).

An unambiguous assignment was also obtained for a 0.75 mM 18-mer DNA hairpin in just 45 min using a mixing time of 0.30 s (Figure 5.7).
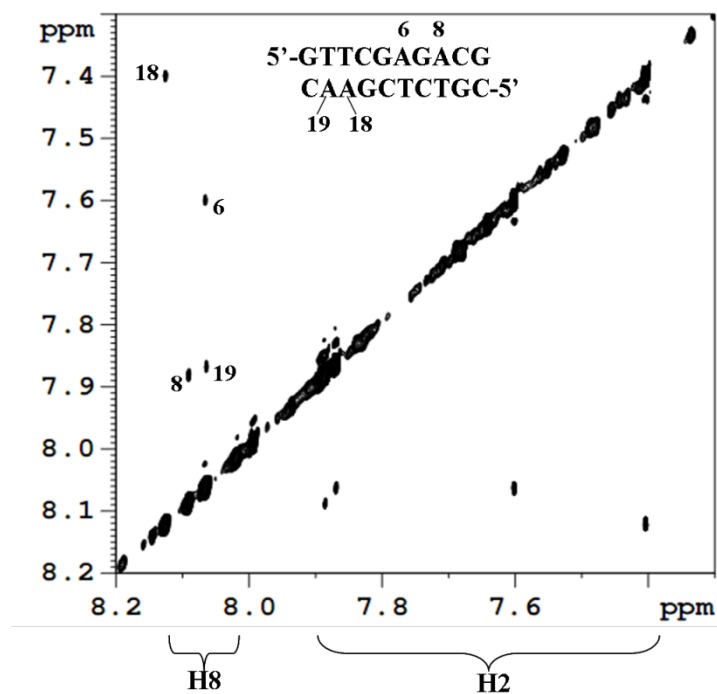

Figure 5.5 The 600 MHz selective TOCSY spectrum of a 0.5 mM decamer DNA duplex sample, recorded at 298K with a 5 s relaxation delay and a mixing time of 500 ms. The carrier was set at 7.6 ppm and an E-Burp pulse of 4.5 ms was used for excitation. A 1024 × 100 data matrix was acquired in a spectral window of 2.4×2.4 ppm with 64 scans per increment resulting in an experiment time of 5 h.
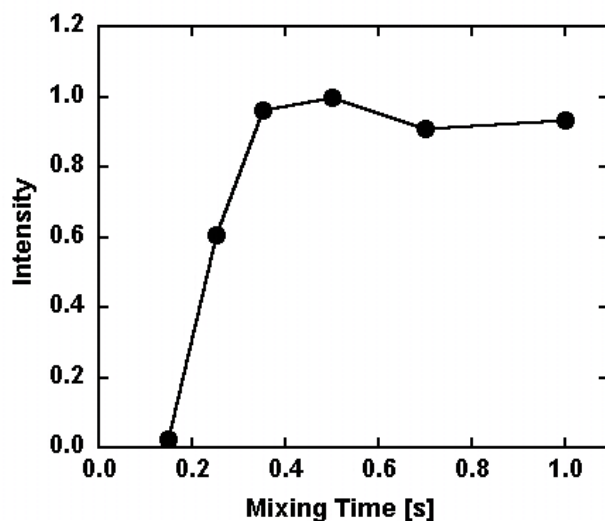
Figure 5.6 Dependence of the cross peak intensity on TOCSY mixing time for the A6 of the decamer DNA duplex.

Depending on sequence and structural peculiarities, H2 protons can be assigned from regular NOESY experiments, particularly for RNA (*22*). However, the relevant cross-peaks are often missing, weak, ambiguous, or appear in overlapped regions (Figure 5.8), especially in non-helical segments. Although not observed for the DNA decamer and 18-mer hairpin shown previously, a rotational NOE (ROE) artifact could potentially also arise if H2 is close to another proton. However, a direct ROE has the opposite sign compared to a TOCSY cross-peak, and a through-space correlation is also readily recognized from a companion NOESY experiment. However, the through-bonds correlations are unambiguous.

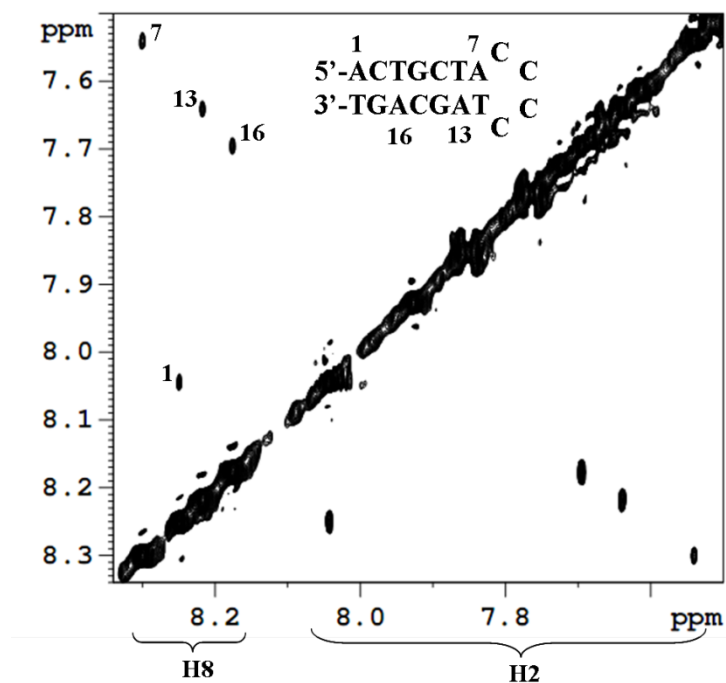**Figure 5.7** The 600 MHz selective TOCSY spectrum of a DNA hairpin (0.75 mM), recorded at 298 K with relaxation time of 3.5 s and mixing time of 300 ms. The carrier was set at 7.6 ppm and an E-Burp pulse of 6.0 ms was used for excitation The spectrum was required using a 3.0 × 3.0 ppm spectral window. A 1024 × 128 data point matrix was acquired using 4 scans per increment resulting in an experiment time of 45 minutes.
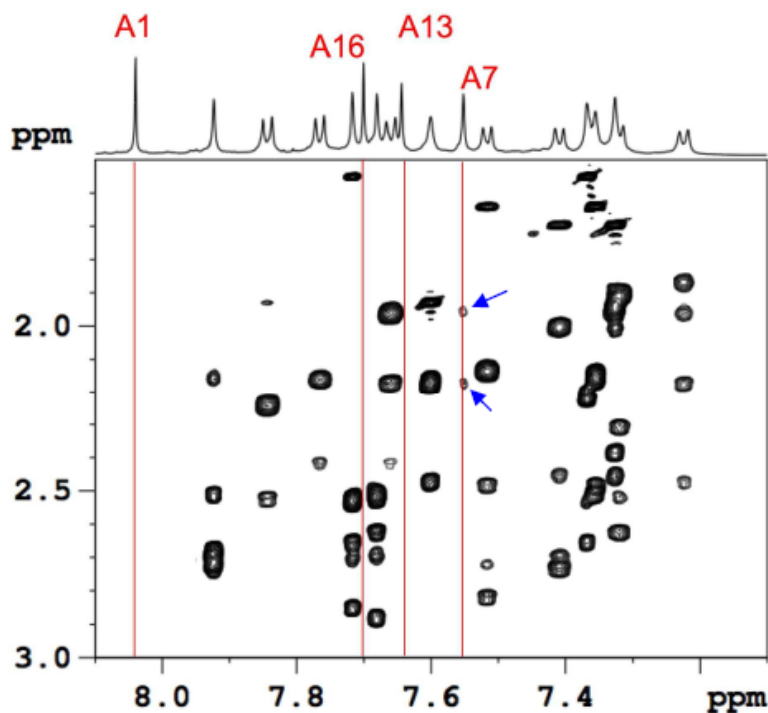
Figure 5.8 The 2′, 2″-Base region of a 600 MHz NOESY spectrum of the 0.75 mM DNA hairpin recorded at 298 K (99.99% D_2O,10 mM sodium phosphate buffer, 0.1 mM EDTA 50 mM NaCl, pH*6.84) with a mixing time of 200 ms and a relaxation delay of 8 s. Only the adenine 7 H2 protons next to the harpin loop show weak NOEs to H2', H2" protons of hairpin loop residues (marked with arrows).

5.5   Conclusion

A long-range homonuclear coupling of adenine H2-H8 proton was observed in long-range optimized COSY and TOCSY experiments with long duration elements. The amplitude of the coupling constant was estimated by NMR sim simulation and experimentally measured by an XLOC pulse program, yielding a value of 0.5 +/- 0.2 Hz.

Using a selective TOCSY experiment with an E-Burp shape pulse to selectively excite the base region, we evaluated the application of the 6/7 bond homonuclear coupling to identify H2 by the H2-H8 correlation in different DNA sequences. The successful use of the optimized

selective TOCSY to explore H2-H8 coupling identified a rapid and unambiguous method for assigning A H2 resonances in unlabeled oligonucleotides.

5.6   References

1.   Halle, B., and Denisov, P. V. (*1998*) Water and Monovalent Ions in the Minor Groove of B-DNA Oligonucleotides as Seen by NMR, *Biopolymers (Nucleic Acid Sciences) 48*, 210-233.
2.   Varani, G., and Tinoco, I. (1991) RNA structure and NMR spectroscopy, *Jr. Q. Rev. Biophys 24*, 479-532.
3.   Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, *John Wiley & Sons: New York.*
4.   Legault, P., Farmer, T. B., Mueller, L., and Pardi, A. (1994) Through-Bond Correlation of Adenine Protons in a [13]C- Labeled Ribozyme, *J. Am. Chem. Soc. 116*, 2203-2204.
5.   Marino, P. J., Prestegard, H. J., and Crothers, M. D. (1994) Correlation of Adenine H2/H8 Resonances in Uniformly [13]C Labeled RNAs by 2D HCCH-TOCSY: A New Tool for [1]H Assignment, *J. Am. Chem. Soc. 116*, 2205-2206.
6.   Simorre, J., Zimmermann, R. G., Mueller, L., and Pardi, A. (1996) Triple-Resonance Experiments for Assignment of Adenine Base Resonances in ^{13}C/^{15}N-Labeled RNA, *J. Am. Chem. Soc. 118*, 5316-5317.
7.   Simon, B., Zanier, K., and Sattler, M. (2001) A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly ^{13}C-labeled RNA molecules, *Journal of Biomolecular NMR 20*, 173-176.
8.   van Dongen, M. J. P., Wijmenga, S. S., Eritja, R., Azorin, F., and Hilbers, C. W. (1996) Through-bond correlation of Adenine H2 and H8 protons in unlabeled DNA fragments by HMBC spectroscopy, *J. Biomol. NMR 8*, 207-212.
9.   Cho, B. P., and Harvey, R. G. (1987) Complete 'H and 13C NMR Assignment of Polycyclic Aromatic Fluoranthenes by Long-Range Optimized Homo-and Heteronuclear Correlation Spectroscopy *J. Org. Chem. 52*, 5679–5684.
10.   Meissner, A., and Sørensen, O. W. (2001) Measurement of J(H,H) and long-range J(X,H) coupling constants in small molecules. Broadband XLOC and J-HMBC, *Magn. Reson. Chem. 39*, 49-52.
11.   Volt, L. R., Waugh, S. J., Klein, P. M., and Phelps, E. D. (1968) Measurement of Spin Relaxation in Complex Systems, *J. Chem. Phys. 48*, 3831.
12.   Kingsley, B. P. (1999) *Concepts Magn. Reson. Part A 11*, 243-276.
13.   Thøgersen, H., and Sørensen, O. W. (1994) Improved resolution in nonselective XLOC experiments for measurement of 1H-1H coupling constants, *J. Magn. Reson. Series A 110*, 118-120.

14. Meiboom, S., and Gill, D. (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times *Rev. Sci. Instrum. 29*, 688.

15. Frisch, M. J., Trucks, G. W., and Schlegel., H. B. (2004) Gaussian 03, Revision C.02, *Gaussian, Inc., Wallingford CT*.

16. Cammi, R., Mennucci, B., and Tomasi, J. (**1999**) Second-Order Møller−Plesset Analytical Derivatives for the Polarizable Continuum Model Using the Relaxed Density Approach, *J. Phys. Chem. A 103*.

17. Cammi, R., Mennucci, B., and Tomasi, J. (2000) Fast Evaluation of Geometries and Properties of Excited Molecules in Solution: A Tamm-Dancoff Model with Application to 4-Dimethylaminobenzonitrile, *J. Phys. Chem. A 104*, 5631-5637.

18. Mahi, L., and Duplan, J. C. (1997) Measurements of Long-Range Homonuclear Coupling Constants, *Magn. Reson. Chem. 35*, 379-383.

19. Meissner, A., Schulte-Herbruggen, T., and Sørensen, O. W. (1998) Relaxation Artifacts and Their Suppression in Multidimensional E.COSY-type NMR Experiments for Measurement of J Coupling Constants in 13C- or 15N-Labeled Proteins, *J. Am. Chem. Soc 120*, 7989-7990.

20. Bertini, I., Jimenez, B., Piccioli, M., and Poggi, L. (2005) Asymmetry in 13C−13C COSY Spectra Provides Information on Ligand Geometry in Paramagnetic Proteins, *J. Am. Chem. Soc. 127*, 12216-12217.

21. Bax, A., and Freeman, R. (1981) Investigation of Complex Networks of Spin—Spin Coupling by two dimensional NMR, *J. Magn. Reson. Series A 44*, 542-551.

22. Heus, H. A., and Pardi, A. (1991) Novel proton NMR assignment procedure for RNA duplexes, *J. Am. Chem. Soc. 113*, 4360-4361.

APPENDIX

Abbreviations

A_i: paramagnetic attenuation

AMP: Adenosine 5' monophosphate

CD: circular dichrosim

COSY: correlation spectroscopy

COSYLRQF: Magnitude-mode Long-Range optimized 2D COSY

DHPC: 1, 2-dihexanoyl-*sn*-Glycero-3-Phosphocholine

DMPC: 1, 2-dimyristoyl-*sn*-Glycero-3-Phosphocholine

DMSO:Dimethyl sulfoxide

DOSY: diffusion-ordered Spectroscopy

DPC: deuterated dodecylphosphocholine

DSC: differential scanning calorimetry

DSP: dithiodis[succinimidypropionate]

DTT: dithiothreitol

EXSY: Exchange Spectroscopy

Gd-DTPA-BMA: gadolinium-diethylenetriaminepentaacetic acid bismethylamide

GST: Glutathione-S-transferase

HMBC: Heteronuclear Multiple Bond Correlation

HSQC: heteronuclear single-quantum correlation spectroscopy

IPTG: isopropyl β-D-thiogalactopyranoside

 ITC: isothermal titration calorimetry

ITC: Isothermal titration calorimetry

L45: peptide derived from the S4-S5 linker of Shaw2 Kv

NCR: noncoding region

NMA: N-methylacetamide

NMR: Nuclear magnetic resonance

NOE: Nuclear-Overhauser effect

NOESY: Nuclear Overhauser effect spectroscopy

ORF: open reading frame

PBS: phosphate buffered saline

PDB: protein data bank

PPIs: Protein-protein interactions

PRE: paramagnetic relaxation enhancement

PRIs: protein–RNA interactions

RBD: RNA-binding domain

RBPs: RNA binding proteins

RNP: ribonucleoprotein domain

RRM: RNA Recognition Motif

S6c: peptide derived from the S6 C-terminus of Shaw2 Kv

SDS PAGE: SDS polyacrylamide gels electrophoresis

SL: stem loop

SPR: surface plasmon resonance

TFE: 2, 2, 2-trifluoroethanol

TOCSY: total Correlation Spectroscopy

TTAB: tetradecyltrimethylammonium bromide

UV: ultraviolet

VSD: voltage sensing domain

XLOC: X Nucleus for Long-range Coupling Experiment