

Theoretical and Numerical Studies of Dynamics in Nucleic Acids
based on Experimental NMR Data

Prashant S. Emani

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2012

Reading Committee:
Gary Drobny, Chair
Gabriele Varani
Jens H. Gundlach

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Theoretical and Numerical Studies of Dynamics in Nucleic Acids
based on Experimental NMR Data

Prashant S. Emani

Chair of the Supervisory Committee:

Professor Gary Drobny

Department of Chemistry

The collection of work presented in this thesis is directed towards building an understanding of the dynamics of nucleic acid molecules and their components using data obtained from solid state and solution NMR experiments. The focus of these studies is to develop analytical and numerical methods of elucidating motional trajectories of residues in example molecules, by simulating the impact of specific choices of models on NMR observables. Specifically, the target molecules studied were the unbound HIV-1 TAR RNA, a 29 nucleotide RNA segment, and the unbound dodecamer HhaI methyltransferase-recognition DNA. The data available for various residues in these systems include solid state line shapes, longitudinal (T_{1Z}) and quadrupolar (T_{1Q}) relaxation times, as well as solution longitudinal (T_1) and rotating frame

($T_{1\rho}$) relaxation times and Nuclear Overhauser Effects (NOEs). The four projects discussed in this thesis form a cohesive whole, with each succeeding method either building upon previous work or adding a new means of analysis: firstly, a slow exchange theory is presented where discrete-jump motional models derived using solid state NMR data can be tested against solution relaxation times, by the inclusion of both overall molecular tumbling and exchange between conformers occurring at a time scale much slower than the tumbling time scale. The time scale separation allows for a particularly simple weighted summation over the spectral density contributions from the various conformers. The second project, discussed subsequently, removes this assumption of time scale separation, and allows for any rate of exchange between conformers. Both simulation protocols use the TAR RNA molecule as the test system. Parallel work on the HhaI-recognition DNA builds a framework for testing a discrete-jump trajectory constructed using pre-existing rotamers of the molecule against solid state relaxation times. This visualization of the dynamics of a residue is then carried over to the solution domain, where the properties of computationally energy-minimized structures of TAR RNA are used to define a solution trajectory. In this last case, data available for multiple sites on the molecule are used to test the model for the trajectory, as well as to fit the rates of motion.

TABLE OF CONTENTS

List of Figures	v
List of Tables	viii
<u>Chapter I: Introduction</u>	<u>1</u>
Structure of thesis	11
<u>Chapter II: Basic NMR Theoretical Background</u>	<u>14</u>
1. Basics of nuclear magnetic resonance	14
a. The electric quadrupolar interaction	20
b. The magnetic dipolar interaction	22
c. Chemical shift interaction	23
2. Experimental observables	24
a. Relaxation times	25
b. Solid state line shape	31
<u>Chapter III: A Slow Exchange Model of Non-rigid Rotational Motion in RNA for combined Solid-state and Solution NMR studies</u>	<u>35</u>
Abstract	35
1. Introduction	35
2. Theory	39

a.NMR Relaxation in the Presence of Slow Conformational Exchange	39
3. Methods	47
a. Rates, Amplitudes and Coordinate Frames for Internal Motions in TAR RNA	47
b. Rotational Diffusion Tensors	50
4. Results	53
a. AER dependence	58
b. Dependence on direction of helical bending	61
5. Discussion	62
6. Conclusions	66
Appendix III.A: Sample Mathematica code for the slow exchange method	67
 <u>Chapter IV: Theory of Non-rigid Rotational Motion Applied to NMR Relaxation in RNA</u>	 <u>71</u>
Abstract	71
1. Introduction	72
2. Theory	76
a. General solution to single conformer diffusion equation	79
b. Solution of the conformational exchange problem	85
c. Evaluation of the correlation functions between Wigner rotation matrices	91
d. Solid-state models	94
3. Results	96
a. Comparison with Slow Exchange formalism	98
b. Rate dependence	103
c. T_1 turnover	108
d. Twist angle dependence	111

4. Discussion	113
5. Results based on corrected equations	119
6. Conclusions	124
Appendix IV.A: Sample Mathematica code for the general rate theory	124
Appendix IV.B: Generating modified atomic-coordinate files to simulate changes in helical parameters	124
<u>Chapter V: Studying experimental structure-based trajectories for base-flipping in HhaI methyltransferase recognition DNA</u>	<u>134</u>
Abstract	134
1. Introduction	134
2. Theory	139
3. Methods	144
a. Reconstruction of base and accounting for the contributions of multiple labels	144
b. Evaluation of frame transformations	145
c. Automated calculation of relaxation times	147
4. Results	150
5. Discussion	155
6. Conclusions	155
Appendix V.A: Sample Mathematica code for best-fit parameter search for the backbone sites	155
<u>Chapter VI: Constructing an RNA motional trajectory through energy-minimized structures</u>	<u>165</u>
Abstract	165
1. Introduction	165
2. Methods	169

3. Results	176
4. Discussion	183
5. Conclusions	183
Appendix VI.A: Mathematica code for the jump rate parameter search algorithm using the energy-minimized structures	183
<u>Bibliography</u>	<u>210</u>

List of Figures

Figure Number	Page
III.1	HIV-1 TAR RNA secondary and tertiary structure. 39
III.2	Solid state NMR-derived motional models for U38 48
III.3	Artificially modified structures of TAR RNA with bend angles of 0° to 50° applied to the upper helix and different directions of helical bending (“90 degree series” and “30 degree series”).....54
III.4	Diffusion tensor eigenvalues and bond angles of the “90 degree series” and “30 degree series”56
III.5	Relaxation times vs. bend angles for various value of the AER for the “90 degree series” 57
III.6	Relaxation times vs. bend angles for various values of the AER for a two-site exchange between the “30 degree series” and 1ANR-1 59
III.7	Relaxation times vs. bend angles for a two-site exchange between structures with various directions of helical bending (“90 degree series”, “30 degree series” and “150 degree series”) and 1ANR-1..... 61
IV.1	HIV-1 TAR RNA secondary and tertiary structure (different rendition) 74
IV.2	Artificially modified structures with bend angles of 0° to 50° applied to the upper helix and two different twist angles applied to the upper helix (“Tw15 series” and “Tw-15 series”) 96
IV.3	Comparison of relaxation times calculated using the slow exchange and general rate of exchange methods..... 100
IV.4	Relaxation times vs. bend angles for various rates of two-site exchange ranging from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$ between the “90 degree series” and 1ANR-1 102
IV.5	Relaxation times vs. bend angles for various rates of two-site exchange ranging from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$ between the “30 degree series” and 1ANR-1 102
IV.6	Relaxation times vs. bend angles for various rates of two-site exchange ranging from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$ between the “Tw15 series” and 1ANR-1 105
IV.7	Relaxation times vs. bend angles for various rates of two-site exchange ranging from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$ between the “Tw-15 series” and 1ANR-1..... 105

IV.8	T_1 vs. bend angles for rates of two-site exchange of $1 \times 10^{10} \text{ s}^{-1}$ and $5 \times 10^9 \text{ s}^{-1}$ between 1ANR-1 and four sets of structures (“90 degree series”, “30 degree series”, “Tw15 series” and “Tw-15 series”) 108
IV.9	T_2 vs. bend angles for rates of two-site exchange of $1 \times 10^{10} \text{ s}^{-1}$ and $5 \times 10^9 \text{ s}^{-1}$ between 1ANR-1 and four sets of structures (“90 degree series”, “30 degree series”, “Tw15 series” and “Tw-15 series”) 109
IV.10	T_1 vs. bend angles for a rate of two-site exchange of $1.38 \times 10^6 \text{ s}^{-1}$ between 1ANR-1 and structures with an upper helical twist ranging from -60° to $+60^\circ$... 110
IV.11	T_2 vs. bend angles for a rate of two-site exchange of $1.38 \times 10^6 \text{ s}^{-1}$ between 1ANR-1 and structures with an upper helical twist ranging from -60° to $+60^\circ$... 111
IV.12	Relaxation times vs. bend angles for various rates of two-site exchange ranging from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$ between the “Tw15 series” and 1ANR-1, based on corrected equations (includes comparison with previous graphs) 119
IV.13	Relaxation times vs. bend angles for rates of two-site exchange of $1 \times 10^{10} \text{ s}^{-1}$ and $5 \times 10^9 \text{ s}^{-1}$ between the “Tw15 series” and 1ANR-1, based on corrected equations (includes comparison with previous graphs) 121
IV.14	Relaxation times vs. bend angles for a rate of exchange of $1.38 \times 10^6 \text{ s}^{-1}$ between 1ANR-1 and structures with a positive upper helical twist ranging from 0° to $+60^\circ$, based on corrected equations (includes comparison with previous graphs) 122
IV.15	Relaxation times vs. bend angles for a rate of exchange of $1.38 \times 10^6 \text{ s}^{-1}$ between 1ANR-1 and structures with a negative upper helical twist ranging from -60° to 0° , based on corrected equations (includes comparison with previous graphs) 123
V.1	The three structures used in the simulation of the NMR data for the HhaI DNA . 137
V.2	Flowchart of the Metropolis-Hastings algorithm used to find the best-fit parameters for the solid state NMR relaxation time data 149
V.3	Histograms of the parameter values obtained from the MCMC algorithm for the backbone sites 152
V.4	Histograms of the parameter values obtained from the MCMC algorithm for the furanose sites 153
V.5	Histograms of the parameter values obtained from the MCMC algorithm for the base sites 154

VI.1	Relation between the Euler angles determined from the atomic coordinates and the associated domain motions for the HIV-1 TAR RNA	172
VI.2	α and β Euler angle distributions of upper helical vectors relative to the lower helix for the energy minimized structures of HIV-1 TAR RNA	177
VI.3	γ Euler angle distributions of U38 C6-H6 vectors relative to the lower helix for the energy minimized structures of HIV-1 TAR RNA	178
VI.4	Distributions of the bulge uridine (U23 and U25) C6-H6 vectors relative to a frame fixed to the lower helix.....	179

List of Tables

Table Number	Page
IV.1	Eigenvalues and eigenvectors of the single conformer rotational diffusion operator in terms of coefficients of Wigner rotation matrices 83
V.1	Best-fit dynamic parameters for the three labeled sites on the target cytosine residue of the HhaI DNA, as well as the $T_{1\rho}$ values calculated for these parameter choices and the corresponding experimental relaxation times 150
VI.1	Relaxation times and NOEs from fit to the solution relaxation data for the uridine bases of the HIV-1 TAR RNA, for the case where the local librations and glycosidic rotations are not included..... 180
VI.2	Relaxation times and NOEs from attempted fit to the solution relaxation data for the uridine bases of the HIV-1 TAR RNA, for the case where the local librations and glycosidic rotations are included..... 181

Acknowledgements

Bear with me. I have a lot of people to thank. My graduate research owes a great deal of debt to my advisor Professor Gary Drobny, whose guidance and insight, as delivered through several engaging conversations over the years, have directed me towards greater productivity. I especially appreciate his enthusiasm with respect to new ideas and his willingness to let me take the initiative to pursue them without micromanagement. I would also like to thank Professor Gabriele Varani, who served in many ways as a second advisor with a much-valued penchant for un-sugar-coated advice. My graduate student mentor, Professor Jens Gundlach, was a well-respected advisor at one time, and has continued to be a wonderful guide throughout my academic life. I would like to thank the rest of my committee members as well: Professor Marcel den Nijs, with whom I have had many fruitful conversations; Professor Paul Wiggins, who was always a pleasure to converse with; and Professor Walter Ruzzo, the Graduate Student Representative on my committee, and a very responsive one at that.

I have had the good fortune of working with great colleagues. Greg Olsen and Dorothy Echodu were wonderful mentors, co-workers and, above all, good friends. Jason Ash, Nicholas Breen, Michael Groves, Wei Huang, Kun Li, Moise Ndao, Kari Pederson, Matt Powers, Adrienne Roehrich and Ariel Zane have been fantastic collaborators and great people to be around.

I can rightfully brag that I have made some amazing choices when it comes to my friends. The extended family I have developed here in Seattle, and those who have continued to be in my life from before moving here have made the quality of life incomparably better and I cannot really repay them all for it. Except for Tessa. She would be happy just to have a ball

kicked around for an hour in the backyard. She has been a happiness-inducing part of my family in these years.

I do not say this often enough to my family, but none of what I did would be possible without my wonderful mother, Nagalakshmi Yenamandra, and cricket-obsessed brother, Gautam Emani, my late grandfather, Suryanarayana Yenamandra and grandmother, Lalitha Yenamandra, and all my close relatives who I continue to grow up with. I wanted to take this opportunity to thank my uncle, Dr. Sury Yenamandra, who is an amazing inspiration for continuing to fight cancer while still being a support structure unto himself for many, many people.

Chapter I

Introduction

The staggering diversity in the functioning of biological molecules arises ultimately as a consequence of their varied composition from smaller functional units, as well as their interaction with their external environments. Proteins occur as combinations of amino acids of varying length, while nucleic acids are built from constituent nucleotide units, and they could find themselves in solution, in a crystalline state, in proximity of organic or inorganic surfaces, in association with other biological molecules or in external fields, amongst a myriad of other scenarios. Other biomolecules may exploit interactions with any number of other molecules or atoms to exhibit entirely idiosyncratic behaviour. It is understandable, therefore, that, while individual cases will obviously hold interest to researchers, there is also a push towards developing a unified understanding of biological molecule function. To that effect, a vocabulary has developed around structure identification: α -helices and β sheets in proteins, A-form, B-form or Z-form helices in nucleic acids, single-stranded loops, and so on. However, biomolecules for the most part do not occur as static, unchanging systems, but as dynamic respondents to a potentially complicated energy landscape. A complete description therefore necessitates a characterization of structure at any moment as well as the time-evolution of the structures, both of which determine how the molecule will perform its function or, just as significantly, fail to perform its function.

It is this relevance of dynamics to function that is exciting to researchers seeking solutions to medical problems at the molecular level. If it is possible to identify the primary motional modes and the associated degrees of freedom that are intrinsic to a biological system (molecule plus a particular environment), then we can seek to enhance or inhibit the functionally

relevant changes in the conformation of the system. Consider riboswitches, for example, which are non-coding messenger RNAs (mRNAs) that regulate the expressions of genes by the binding of specific metabolites, occurring in prokaryotic cells and to a lesser extent in eukaryotic cells. The aptamer, or ligand-binding region, binds the metabolite and this in turn drives changes in the structure of the attached expression platform, which impacts gene expression. It is the mechanism of the folding and unfolding of the molecule in response to the increase or decrease of metabolites in the cellular environment that is therefore of great importance in understanding the regulation of purines¹⁻³, glycine⁴, S-adenosyl-methionine (SAM)⁵ and other compounds. Identifying the sites along the structure that are primarily responsible for changes in the conformation can allow one control over the switching behaviour of these molecules.

To further underscore the importance of determining the dynamic characteristics of molecules, let us consider the two biological systems used as test cases for the development of the analytical methods in this manuscript: (1) the HIV-1 transactivation response (TAR) RNA and (2) the HhaI DNA recognition sequence.

The HIV-1 TAR RNA is a 59-nucleotide segment that binds the viral regulatory protein Tat, which in turn acts to enhance transcription elongation. This binding process is essential to the explosive replication of HIV-1. The Tat-TAR complex is formed in a region of the RNA in proximity to the apical loop where the Tat binding site surrounds the single-stranded, tri-nucleotide (UCU) bulge.^{6,7} A 29 nucleotide TAR segment that includes the relevant helical stems, connecting bulge and apical loop of the RNA has been used in NMR studies as a representative of the binding-capable region.^{6,8} Hereafter, the term TAR will refer to this reduced 29-nucleotide construct. The bulge region interlinking the two helical stems in TAR has been shown to be the primary binding site for the Tat protein.^{7,9} The single stranded bulge exhibits

considerable flexibility¹⁰⁻¹³ and allows for the two helical regions to adopt a wide range of relative orientations.^{10,14} The system is believed to undergo a process of “conformational capture” where the free TAR RNA exchanges between multiple conformers, one or more of which is amenable to Tat binding.¹⁵ Given the importance of the bulge region, the search is on for Tat mimetics or other molecules that will show a high affinity for TAR RNA and adversely impact Tat-TAR complex formation.^{16,17} Using this as a paradigmatic example, the hope is that an evaluation of the dynamic behaviour of residues in a sequence can inspire the development of drugs with specific binding affinities.

The second example is that of the recognition DNA sequence, [d(G₁A₂T₃A₄G₅C₆G₇C₈T₉A₁₀T₁₁C₁₂)]₂, for the HhaI methyltransferase. The underlined residues form the conserved recognition site for the methyltransferase, with the C₆ being methylated in the process. In this system, the DNA sequence in question serves as the specific binding site of the methyltransferase protein which catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to the C5 atom of the target cytosine on the sequence. This methylation process is known to protect bacterial DNA from cleavage by restriction endonucleases, and is thought to play a role in eukaryotic transcription, X chromosome inactivation, DNA repair^{18,19} and potentially in cancerous cells²⁰. This essential process in cells proceeds by means of the binding of the methyltransferase at the recognition site and the subsequent “base-flipping” of the target cytosine into the active site of the protein.¹⁸ It is of interest to us to understand whether this flipping process is active or passive, i.e. whether the cytosine is actively flipped out by the protein, or whether the protein opportunistically captures a naturally dynamic residue, similar to the TAR RNA conformational capture idea. This has been investigated extensively^{21,22} with

ultimate hope of being able to prevent hypo- or hyper-methylation, if there is a proven medical need to do so.

Among the methods developed for such detailed observation, nuclear magnetic resonance (NMR) experiments have emerged at the forefront of techniques available for discerning the structure and motion of molecules and their components. Exploiting the sensitivity of the spins of individual nuclei to their local environment, NMR experiments are able to probe intra- and inter-molecular interactions at the atomic level through a variety of sample conditions (such as in solution and in the solid state with varying degrees of hydration) and a massive array of radiofrequency pulse-based spin manipulation techniques (spectral line shape studies, relaxation time analyses, nuclear Overhauser effect (NOE) measurements, among many more). Moreover, depending upon the question under investigation, the sample may be artificially modified by replacing nuclei with magnetically sensitive (i.e. non-zero spin) isotopes at particular sites in the molecule or the sample may be studied by means of its natural-abundance isotopes. Each such experiment, with some combination of sample conditions, labeling and pulse technique, can yield specific information about the system at hand, such as the relative proximity of functional groups, or the rates and amplitudes of motions of a section of the molecule. Indeed, different methods can often complement each other if they are sensitive to different aspects of structure and motion.

The atomic-level resolution of these techniques implies that the details of the types and strengths of the interactions of these spins with their environments need to be understood. The interactions considered in NMR are all electromagnetic in nature, and can include dipolar spin-spin (magnetic), quadrupolar electric field gradient-based and indirect, electron polarization-dependent contributions. Every choice of spin label comes with its own characteristic interaction

set, with the relative contributions of each type being determined by the neighbouring environment. Moreover, the specific tensorial nature of an interaction Hamiltonian plays a vital role, as it manifests itself in the unique orientation-dependence of the electromagnetic field experienced by the spin. Knowledge of these interactions and their tensor properties are essential to the development of new methods and to evaluations of the sensitivities of various techniques to motional rates. This information is incorporated into an understanding of the system that also includes due consideration of the sample conditions.

Solution NMR studies, which unsurprisingly refer to the entire spectrum of studies on molecules in solution, are widely used on account of most biological molecules being in solution *in vivo*. Some examples of solution experiments are longitudinal and rotating frame relaxation time measurements, NOE studies and residual dipolar coupling (RDC) or residual chemical shift anisotropy (RCSA) measurements. For dynamics studies, the motional rates accessible to these methods vary both on the measured quantity and on the choice of label. For reference, the following are the regimes of sensitivity for the experiments that are most important for the work to follow: longitudinal relaxation times are affected most by motions with characteristic time scales in the nanosecond to picosecond range, the rotating frame relaxation time can help discern motion from the picosecond scale down to the scale of overall diffusion (as explained below), while the NOEs probe motions on the nanosecond to picosecond scale. The important consideration in such studies is that the molecule undergoes overall diffusive rotation that may or may not be hindered by other molecules or surfaces. This rotation induces averaging over those local tensorial interactions of atoms at a time scale determined by the hydrodynamic properties of the molecule and the viscosity of the surrounding liquid. This means that for the case of unhindered overall rotation (i.e. a uniform sampling of all orientations) any experiment with rate

sensitivity in a regime slower than the tumbling rate of the molecule will see an interaction tensor pre-averaged to zero, resulting in no signal. This implies that there is a cutoff point in sensitivity to the rates of motions in these experiments depending on the system being considered. One exception to this is the relaxation dispersion experiment which depends on the scalar chemical shift of the nuclei and is therefore not affected by isotropic rotation. This method is sensitive to motions slower than the microsecond time scale.

Solid state NMR experiments are intended to probe molecules that are prevented from overall diffusive rotation by being in a no-solvent or limited-solvent environment. This includes samples with molecules arrayed in a periodic crystalline state and those in an amorphous “powder” state, as well as lyophilized (no solvent) samples and those with varying degrees of solvent added. The change in sample conditions from solution to solid state makes a significant difference, as any tensorial interactions will retain their orientation dependence. It is this absence of the averaging effects of overall molecular tumbling that allows solid state NMR methods to discern a range of motional rates inaccessible to solution techniques. For instance, the crucial nanosecond to microsecond time scale window, where many conformational changes take place, is visible to solid state deuterium NMR line shape measurements (deuterium is the primary choice of labels for most of the solid state experiments reported on in this manuscript) on account of the retention of the anisotropy of interactions. Thus, by arresting molecular tumbling, we can hope to capture motions on a scale complementary to those in a solution setting. This is the primary motivation of much of the work presented in this manuscript: we hope to build a bridge between two well-studied, yet largely separate domains of experimental investigation, solid state NMR and solution NMR, by combining the analyses of the dynamics of biological molecules into a single model-based framework. In other words, we hope to develop a model of

motion based on data from one set of sample conditions, and check that the model is consistent with the data from the other set. This is the main focus of the work presented in Chapters III, IV and VI.

A natural question that may be asked is: are the motions themselves the same in the solid state and in solution? There is no *a priori* reason to believe that the same natural modes of the system occur when the sample is in a limited solvent environment, given that solvent-solute interactions play a very important role, especially in nucleic acids. It is not possible to answer this question for an arbitrary sample considering that the exact nature of aggregation in the solid state is hard to discern for amorphous arrangements, but one can look at the continuum from solid state to solution and study the behaviour as a function of increasing solvent content. In addition, the local motions of residues in the sample, which are less impacted by intermolecular interactions with relatively distant partners, can for the most part be assumed to be the same in both situations. This is an important question that is necessarily addressed with every new system and will be explored in the manuscript.

More specifically, the work presented herein will primarily consider the solid state NMR deuterium longitudinal and quadrupolar relaxation times—with time scale sensitivities in the nanosecond range; the deuterium line shape, i.e. the spectra of the spins in the presence of the magnetic field—whose sensitivity extends over the nanosecond to millisecond range; the solution NMR ^{13}C longitudinal and rotating frame relaxation rates—covering the picosecond to diffusion time scale; and the solution NOEs—sensitive to motions on the nanosecond to picosecond time scale. These observables will be described in greater detail in the next chapter.

Once this data is available in sufficient quantity, however, what remains is the significant challenge of relating the observables to the underlying physical situation. One of the more

common techniques is to ascribe a fixed number of motions to the residue in question, with some physical motivation as to this choice. This can be done in a semi-analytic, model-dependent way, with a number of free parameters, and with the motion of the bond being confined either within a continuous potential energy function of relevant variables²²⁻²⁹ or to a set of discrete orientations with different relative probabilities.^{10,30-36} In both these cases, the choice of model is often made based on physically reasonable expectations of the behaviour of the system under consideration. For example, Meirovitch et al²⁸ use a restoring potential to describe the behaviour of the peptide vibrations around an equilibrium orientation, in addition to overall molecular rotation. Echodu et al²⁹ similarly consider the motions of the furanose ring in the recognition site cytosines of the HhaI recognition DNA as being determined by a harmonic potential whose minimum is the C2'-endo conformation. The two-site jump internal motions of the helical uridine base in TAR RNA explored in Olsen et al¹⁰ are confined in a plane defined by the plane of the base, as the proximities of adjacent base-pairs prevent any significant motion in a perpendicular direction. Once this choice is made, the experimental data is fit by varying parameters such as the curvature and barriers of the potential energy function, or the rates and amplitudes of a jump between discrete sites. One of the limitations of such a technique is that assumptions must be made beforehand as to the degrees of freedom available to the system. As mentioned above these assumptions can receive support from physical arguments or even previous experimental observations, but by their nature do not allow for the unexpected. If the model is complicated by the addition of further degrees of freedom and consequently more free parameters, the simulations runs the risk of being under-constrained by the data available. A so-called Model-Free (MF) approach^{37,38} was proposed whereby the motions of a sample in solution were assumed to consist of overall rotational diffusion and just one additional internal motion at a

faster rate, resulting in a significantly simplified analysis. This attempted to address the concern of an overload of parameters, and together with its extended version³⁹ found a place among the most widely used of simulation techniques (see Barbato et al⁴⁰, King et al⁴¹ and Bardaro et al¹¹ for a randomly selected and miniscule fraction of the available examples). While the MF and extended-MF approaches do allow a characterization of two internal rates and average orientations, they do not provide a picture of the directions of motion, nor do they leave room for a more complex model. Once again, the key is to strike a balance between maintaining a level of physical realism in the models and keeping the number of free parameters at a sufficiently low number relative to the number of available data points. Apropos the requirement of building realistic models, molecular dynamics (MD) simulations and de novo structure prediction programs have made significant progress recently. Both involve the establishment of accurate potential energy functions for the molecule and treatments of the solvent, and seek to predict molecular structure and dynamics armed with an understanding of atomic-level interactions. Some common MD software packages available include AMBER⁴²⁻⁴⁵ and CHARMM^{46,47} as well as other programs, which have contributed significantly to simulations of proteins (see for example, Lindorff-Larsen et al⁴⁸) and to interesting applications in nucleic acids (for instance, the work of Horton et al⁴⁹ and Riccardi et al⁵⁰). The work of Riccardi et al⁵⁰ deserves special mention as some of the work presented in this manuscript bears similarities to the method used by the authors of that paper. The authors utilize the AMBER all-atom force field to run a series of trajectories for the 36-nucleotide SL1 segment of the Ψ site of HIV-1, and then carry out a dihedral angle-based principal component analysis (PCA) on the resulting structures in order to discern clusters. These clusters were interconnected by a transition matrix to allow for a calculation of the time evolution of the system. In the research considered in Chapter VI, a

similar PCA was attempted to discern patterns in a set of energy-minimized structures, but in the end was determined to be unable to find significant clustering. Instead, other methods were used to suggest a molecular trajectory, but the same philosophy of seeking clusterings remained a driving force. MD calculations involve solving the equations of motion for all atoms, given a set of initial conditions and a properly vetted force-field. On the other hand, the field of de novo structure prediction is mainly based on energy minimization, where some means of iteratively altering the conformation is used until the energy minimum is found. The system does not evolve according to a force-field equation as it does in the case of MD. The Rosetta suite of programs⁵¹ has been developed by David Baker and colleagues to predict the structure of proteins (for example, Bradley and Baker⁵²), and more recently, nucleic acids as well,^{53,54} using a variety of fragment assembly and coordinate refinement techniques. The software has also been expanded to apply experimental chemical shift, RDC and NOE constraints for protein structure determination.⁵⁵ While the structure prediction methodology does not inherently provide dynamic information, the possibility of utilizing a series of energy-minimized conformations of the same molecule as discrete sites in a diffusive exchange process (similar to the transition between clusters in Riccardi et al⁵⁰) is explored in Chapter VI. This analysis is based on structures generated by the fragment assembly of RNA with full-atom refinement (FARFAR) program⁵⁴ and attempts to use the assumption of ergodicity in the following manner: if a molecule adopts a set of conformations dispersed throughout a chosen coordinate space, then a time-average over a sufficiently long time can be represented by an ensemble-average over the conformation set. It is hoped that the variation of the initiation conditions of the energy minimization process can provide an ensemble that is representative of the full dynamic range of

the molecule. The particular method considered in Chapter VI combines the computational and phenomenological modeling techniques in order to simulate solution relaxation times.

As a personal observation, one aspect of NMR simulations that makes them unique and interesting (I dare say, fun even) is the fact that the spin systems under observation find themselves in an environment determined simultaneously by quantum mechanical interactions and macroscopic, diffusive forces. In thermodynamic terms, the specific interactions of spins with the reservoir, i.e. the specific means by which the motion induced by solvent bombardment, natural frequency responses of the molecule and intermolecular interactions translate into a modulation of the orientation of the interaction tensor, are the essential determining factors of how a given observable probes the system. Thus, any simulation necessitates at least a basic understanding of both realms of influence.

Structure of thesis

The work presented herein largely consists of a combination of analytic and numerical methods designed to bridge multiple experimental data sets, such as solid state relaxation times and line shapes, and solid state-derived models and solution relaxation times. This would allow the knitting together of a complementary patchwork of rate and sample condition sensitivities to provide a more definitive model-testing platform. The final goal is, naturally, to be able to build a clearer picture of the dynamic behaviour of biological molecules at all time scales and in multiple physical settings.

Chapter II provides a basic NMR redux, including the theoretical ingredients that go into calculations of line shapes, relaxation times and RDCs. This gives the necessary background to understand the rate sensitivities of the different experimental techniques as well as the quantum

mechanical interactions that are of importance for a certain choice of spin label. Chapters III through VI describe different methods developed for studying the dynamics of nucleic acids. Chapter III elaborates on the so-called Slow Exchange model, where motional models derived from solid state NMR data are tested against solution relaxation times by treating the discrete sites in a jump process as being rigid conformations that exchange infinitely slowly. If the conformers transition amongst themselves at a rate much slower than the rate of diffusion of any single conformer this method produces a reasonable approximation to the actual physical situation. Then, given an infinitely long exchange process, each state essentially behaves as if independent of the others. The solution relaxation rates can be then described as the weighted sum of contributions from independent populations of conformers, and each individual relaxation rate contribution in turn is calculated using the solution to the rotational diffusion equation for that conformer. Chapter IV removes this assumption of slow exchange and incorporates the possibility of exchange at any rate. This greatly expands the domain of applicability of the method. Both these chapters provide simulations of relaxation times for the U38 residue in the upper helix of TAR RNA. Chapter V carries out a simulation of the solid-state relaxation times and some line shapes of three different sites along the target cytosine of the HhaI recognition DNA, using previously published PDB structures of the molecule at three different points along a proposed flipping trajectory. In essence, this simulation aims to check whether such a flipping mechanism is a part of the natural dynamism of the molecule, or if it requires the companion protein to cause the flip. Finally, Chapter VI combines aspects of the three previous chapters by using PDB structures generated using the FARFAR program⁵⁴ as sites along the motional trajectory of TAR RNA. One of the main goals is to properly characterize the dynamic behaviour of the under-constrained, single-stranded regions of the RNA. The exchange process between

these structures is simulated using the analytical formulae derived in Chapter IV, and the data is fit using a Markov-chain Monte Carlo (MCMC) algorithm searching through rate space. The simulated results include solution relaxation times from five different sites along the molecule, aiming to provide a description of concerted motion for the entire molecule.

Chapter II

Basic NMR Theoretical Background

Given the analytical nature of the work that is to follow in later chapters, I would be remiss if I did not provide a gentle introduction to the theoretical calculations underlying the simulations of various observables. This implies covering both an elementary picture of nuclear magnetic resonance (NMR) and the basic quantum mechanical and thermodynamic background to the various spin manipulations.

1. Basics of nuclear magnetic resonance

The essence of NMR lies in the Zeeman interaction between a nuclear spin and an external magnetic field. The Hamiltonian for such an interaction is given by

$$H^{Zeeman} = -\hbar \vec{\gamma} \cdot \vec{B}_0 = -\hbar I_z \gamma B_0 = -\hbar I_z \omega \quad (1)$$

Here, the uniform (over the scale of the sample) external magnetic field $\vec{B}_0 = B_0 \hat{z}$ lies along the z-axis and exerts a torque on the spin magnetic moment, quantified by the product of the spin angular momentum $\hbar \vec{I}$ (the magnitude I is the spin quantum number of the given nucleus) and the gyromagnetic ratio γ of the spin in question. It is this interaction which causes the establishment of $2I + 1$ energy levels defined by the z-projections of the spin angular momentum. The Larmor precession of the spin around the $\hat{z}$ direction, which occurs when a magnetic moment is placed in the presence of a magnetic field, occurs at an angular frequency of $\omega = \gamma B_0$.

For reference, the gyromagnetic ratios of some common NMR-relevant nuclei are: proton (^1H): $2.675 \times 10^8 \text{ rad s}^{-1} \text{ T}^{-1}$; deuteron (^2H): $4.11 \times 10^7 \text{ rad s}^{-1} \text{ T}^{-1}$; carbon-13 (^{13}C): $6.73 \times 10^7 \text{ rad s}^{-1} \text{ T}^{-1}$; and nitrogen-15 (^{15}N): $-2.71 \times 10^7 \text{ rad s}^{-1} \text{ T}^{-1}$. NMR magnetic field strengths are usually

defined in terms of the Larmor frequency of a proton in that field. Therefore, a "500" magnet refers to a field that results in a Larmor frequency of 500 MHz for ^1H . This corresponds to a field strength of 11.74 T. Other examples of commonly used field strengths are 14.1 T (in a 600 MHz magnet), 17.6 T (in a 750MHz magnet) and 18.8 T (in an 800 MHz magnet).

It is worth noting that the term “nuclear spin” is used to refer to the total angular momentum of the nucleus, with both spin and orbital contributions, not just the quantum mechanical spin angular momentum. This dependence on the orbital angular momentum is essential to understanding the quantification of the electric quadrupolar interaction, as will be seen later.

The spin is almost always not isolated. It usually finds itself in a thermal bath or reservoir at a fixed temperature T . In this circumstance, the population of spins in each of the energy levels is proportional to the Boltzmann factor of the state:

$$\text{Probability}(I_z = m) \propto \exp(-\hbar m \omega / k_B T) \quad (2)$$

This equilibrium distribution of spin populations defines the baseline condition of the sample, with various experimental techniques perturbing the spins away from equilibrium. Before proceeding on to a description of these techniques, however, it is necessary to describe the different forms of coupling of the spin to the thermal reservoir. These couplings involve some form of electromagnetic interaction of the spin with its surroundings, with the strength of the interaction being modulated by the orientation and, in the case of a non-uniform magnetic field, the translation of the molecule or its components.

The key step to quantifying these couplings is to parse the full interaction Hamiltonian (leaving out the Zeeman term for now) into its various multipole moments:

$$H^{\text{Interaction}} = H^{\text{EM}}(0) + H^{\text{ED}}(1) + H^{\text{MD}}(2) + H^{\text{EQ}}(2) + H^{\text{MQ}}(3) + \dots \quad (3)$$

In Equation 3, EM stands for the electric monopole, ED and MD for the electric and magnetic dipoles, respectively, and EQ and MQ for the electric and magnetic quadrupole moments, respectively, and so on. Also shown in parentheses next to each term is the tensorial order of the interaction, where “0” represents a scalar interaction, “1” a vector interaction and “2” a second-rank tensor interaction. The tensorial nature of an interaction Hamiltonian relates to how the term transforms under a coordinate transformation. Thus, a tensor of rank l , with spherical tensor components $T_q^{(l)}(S)$, transforms in the following manner under a coordinate frame rotation:

$$T_q^{(l)}(S) = \sum_{q'=-l}^l D_{q'q}^{(l)}(\vec{\Omega}_{SS'}) T_{q'}^{(l)}(S') \quad (4)$$

In this equation, the terms $T_q^{(l)}(S)$ represent the $2l+1$ components ($q = -l, -l+1, \dots, +l$) of the rank l tensor in the coordinate frame S , which are given as linear combinations of the same tensor elements expressed in terms of coordinate frame S' . The coefficients of this linear transformation are the matrix elements of the Wigner matrix of order l , $D_{q'q}^{(l)}(\vec{\Omega}_{SS'})$, which are functions of the Euler transformation $\vec{\Omega}_{SS'} = \{\alpha, \beta, \gamma\}$ between the frames S and S' (using the conventions of Rose⁵⁶ and Tinkham⁵⁷):

$$D_{q'q}^{(l)}(\vec{\Omega}_{SS'}) = e^{-iq'\alpha} d_{q'q}^{(l)}(\beta) e^{-iq\gamma} \quad (5a)$$

For definitions of $d_{q'q}^{(l)}(\beta)$ elements please for a general l value, please check the above references.

For convenience, the expressions for the frequently-encountered $l = 2$ case are as follows:

$$\begin{aligned}
d_{2,\pm 2}^{(2)}(\theta) &= \left(\frac{1 \pm \cos \theta}{2} \right)^2; \\
d_{2,\pm 1}^{(2)}(\theta) &= -\frac{1 \pm \cos \theta}{2} \sin \theta; \\
d_{2,0}^{(2)}(\theta) &= \sqrt{\frac{3}{8}} \sin^2 \theta; \\
d_{1,\pm 1}^{(2)}(\theta) &= -\frac{1 \pm \cos \theta}{2} (2 \cos \theta \mp 1); \\
d_{1,0}^{(2)}(\theta) &= -\sqrt{\frac{3}{2}} \sin \theta \cos \theta; \\
d_{0,0}^{(2)}(\theta) &= \frac{3 \cos^2 \theta - 1}{2}; \\
d_{q',q}^{(l)}(\theta) &= (-1)^{q-q'} d_{q,q'}^{(l)}(\theta) = d_{-q,-q'}^{(l)}(\theta)
\end{aligned} \tag{5b}$$

Mathematically, these coefficients are the means by which the orientation dependence of an interaction is introduced into the calculation. To see how, consider the following. Firstly, the Hamiltonian of tensor rank l is expressed as a sum over products of spin-dependent parts and spatial-coordinate dependent parts:

$$H^{(l)} = \sum_{q=-l}^l (-1)^q A_q^{(l)}(\vec{I}) T_{-q}^{(l)}(\vec{r}) \tag{6}$$

Secondly, the Hamiltonian is expressed in a local frame associated with a given bond and interaction where the interaction matrix is diagonal, the so-called principal axis coordinate system (PAS) of the interaction. Thirdly, a transformation from the PAS frame to the laboratory frame defined by the uniform magnetic field is carried out, through any number of intermediate coordinate frames. The reason for this is that the spin operators are easily diagonalized in a frame with the magnetic field in a fixed, known direction. Thus, the spatial dependent parts of the Hamiltonian must be transformed into the external lab frame, resulting in sensitivity to the (potentially time-dependent) relative orientation between the associated atomic bonds and the lab frame.

The advantage of using a multipole expansion together with knowledge of the tensor nature of the various interactions can be seen by considering matrix elements of the spin operators of a given rank using a basis of the nuclear spin z-eigenstates $\langle I, M_I | \hat{O}_q^{(l)} | I', M_{I'}' \rangle$. The Wigner-Eckart theorem puts limitations on the values of l and q for which the matrix element is non-zero. The theorem states that

$$\langle \Lambda, I, M_I | \hat{O}_q^{(l)} | \Lambda', I', M_{I'}' \rangle = \langle \Lambda, I | \hat{O}_q^{(l)} | \Lambda', I' \rangle C_{lq, I' M_{I'}'}^{I M_I} \quad (7)$$

where $C_{lq, I' M_{I'}'}^{I M_I}$ is the Clebsch-Gordan coefficient for summing the angular momenta l and I' and their projections q and $M_{I'}'$, and $\langle \Lambda, I | \hat{O}_q^{(l)} | \Lambda', I' \rangle$ is known as the reduced matrix element, which quantifies the contribution to the matrix element from additional degrees of freedom labeled by Λ and Λ' . The Clebsch-Gordan coefficient is non-zero only if $q + M_{I'}' = M_I$ and $I \in \{|l - I'|, |l - I'| + 1, \dots, l + I'\}$. This restriction puts limitations on the type of interactions that can be experienced by a given spin. For example, deuterium (^2H) is a spin-1 nucleus and so for all matrix elements involving eigenstates of the deuteron alone, the only interactions that produce non-zero elements are those for which $l \leq 2I = 2$. This precludes all electric interactions beyond the quadrupolar term. Similarly, carbon-13 (^{13}C) is a spin-1/2 nucleus and so the only tensor interactions that can influence the nucleus are those with $l \leq 1$, resulting in ^{13}C being unaffected by electric terms higher than the dipolar one. The magnetic dipolar interaction has an influence in both cases, as can be seen by the fact that the Hamiltonian can be expanded in terms of spin operators of rank-1.

This naturally leads us to the question of which interactions are relevant for the spin labels of interest. In general, the electric monopole moment (the magnetic monopole still eludes

the question of its existence) does not produce an orientation dependent interaction as it is a scalar. Thus, it is relegated to the background in all following considerations. The nuclear electric dipole moment has been shown to be 0 based on all experimental evidence to date and if parity conservation holds, all odd-rank electric interaction tensors are forced to 0 as well. The relative importance of the remaining interactions depends on the specific nucleus in question. The two primary nuclei used in the work presented herein are the ones mentioned above: deuterium (^2H) labels in solid state samples and carbon-13 (^{13}C) in solution. The deuteron is influenced by both the electric quadrupolar and magnetic dipolar interactions, but the electric quadrupole moment dominates substantially and most calculations ignore the contribution of the magnetic dipole interaction (the discussion herein does as well). The carbon-13 nucleus is influenced by the magnetic dipole moment interactions, both heteronuclear and homonuclear, as well as chemical shift anisotropies (anisotropic effects of nearby electron currents induced by the external magnetic field), J-couplings with other carbons (spin-spin interactions mediated by electrons in chemical bonds), and scalar couplings with adjacent ^{13}C nuclei (isotropic J-couplings). Furthermore, analysis of relaxation times for ^{13}C is further complicated by multiple dipolar interactions with neighbouring ^{13}C and ^{15}N nuclei. However, by means of various experimental techniques and other considerations,⁵⁸ the only interactions of relevance for this work are the dipolar couplings of ^{13}C with bonded protons as well as the chemical shift anisotropy. In the following, the exact forms of these Hamiltonians are presented (for a very nice exposition see Duer⁵⁹).

a. The electric quadrupolar interaction

This energy contribution arises as a consequence of the interaction of the quadrupolar moment of the nucleus with the electric field gradients created in the vicinity by electrons and other charges.

$$\hat{H}^{EQ} = \frac{eQ}{4I(2I-1)\hbar} \left[\frac{V_0}{\sqrt{6}} (3I_z^2 - I^2) - \frac{V_{+1}}{2} (I^- I_z + I_z I^-) - \frac{V_{-1}}{2} (I^+ I_z + I_z I^+) \right. \\ \left. + \frac{V_{+2}}{2} (I^-)^2 + \frac{V_{-2}}{2} (I^+)^2 \right] \quad (8)$$

(The Hamiltonian is expressed in the computationally relevant units of rad s^{-1} , with an additional factor of $\hbar$ required to convert to SI units.)

The expression for the Hamiltonian involves components of a second-rank tensor in the spin operators. The V_q 's are spherical components of the 2nd rank electric field gradient (EFG) tensor. As mentioned above, it is the spatial dependence of these components that results in an orientation dependence of the interaction Hamiltonian. The convention used here is a combination of the ones used by Slichter,⁶⁰ and Vold and Vold.⁶¹

An interesting point associated with Expression 8 is the fact that an electric interaction term has been represented in terms of the spin operators. This brings us back to the definition of the “nuclear spin” which was described as being representative of the total nuclear angular momentum. While the electric quadrupole moment is defined as a 2nd rank distribution of the nuclear charge, the matrix elements of these spatial distributions are proportional to spin operators of the same rank and tensor component through the Wigner-Eckart theorem (Equation II.7). The constant of proportionality is defined to be $eQ/[I(2I-1)]$, where Q is the quadrupole moment of the nucleus and e is the charge of a proton. The quadrupole moment is usually found

experimentally for different nuclei. This transfer from spatial to spin operators allows the Hamiltonian to be written as in Equation 8, and is possible only because the “spin” eigenstates in the matrix elements of interest are dependent on the spatial distribution of the charge of the nucleus.

Returning to the EFG tensor, the common practice is to find the coordinate frame in which the EFG tensor, written in terms of the Cartesian coordinates, is diagonal. Thus, in this principal axis system (PAS) the only non-zero elements of the 2nd rank EFG tensor are V_{xx} , V_{yy} and V_{zz} . The spherical tensor components in general are given by

$$V_0 = \sqrt{6}V_{zz}; V_{\pm 1} = -2(V_{zx} \pm iV_{zy}); V_{\pm 2} = (V_{xx} - V_{yy}) \pm 2iV_{xy} \quad (9a)$$

and when specialized to the principal axis frame they become

$$V_0^{PAS} = \sqrt{6}V_{zz}^{PAS} = \sqrt{6}eq; V_{\pm 1}^{PAS} = 0; V_{\pm 2}^{PAS} = (V_{xx}^{PAS} - V_{yy}^{PAS}) = \eta V_{zz}^{PAS} = \eta eq \quad (9b)$$

The parameters q and $\eta = \frac{(V_{xx}^{PAS} - V_{yy}^{PAS})}{V_{zz}^{PAS}}$ are defined as the *field gradient* and *quadrupolar*

asymmetry parameter, respectively. The PAS frame is always defined such that V_{zz}^{PAS} is the largest component of the EFG tensor, and thus, η has a magnitude < 1 . In many cases, to a good approximation the EFG tensor is sufficiently close to being axially symmetric, i.e. $V_{xx}^{PAS} \approx V_{yy}^{PAS}$, that the asymmetry can be ignored.

Interaction strengths are usually quantified by the *quadrupolar coupling constant* (QCC)

$$f_Q = \frac{\omega_Q}{2\pi} = \frac{e^2 q Q}{h} \text{ (in units of Hz). The QCC value for deuterium is usually around 170 kHz, the}$$

value varying based on the electronic environment. For the case of base carbons in RNA bonded to deuterium, the asymmetry is assumed to be negligible.

These orientation dependent terms can then be rotated into a new coordinate system using the tensor transformation rule,

$$V_q^{(2)}(S) = \sum_{q'=-l}^l D_{q'q}^{(2)}(\vec{\Omega}_{SS'}) V_{q'}^{(2)}(S') \quad (10)$$

b. The magnetic dipolar interaction

This energy contribution arises due to the interaction of the nuclear spin magnetic moment with the magnetic field created by a neighbouring spin.

$$\hat{H}^{MD} = -\frac{\mu_0}{4\pi} \frac{\gamma_I \gamma_S \hbar}{r_{IS}^3} \left(3\sqrt{\frac{2\pi}{15}} \right) \left[\begin{aligned} & 2\sqrt{\frac{2}{3}} Y_{2,0} \left(I_z S_z - \frac{1}{4} (I^+ S^- + I^- S^+) \right) + \\ & \{ Y_{2,-1} (I^+ S_z + S^+ I_z) - Y_{2,1} (I^- S_z + S^- I_z) \} + \\ & Y_{2,-2} (I^+ S^+) + Y_{2,+2} (I^- S^-) \end{aligned} \right] \quad (11)$$

(The Hamiltonian is expressed in the computationally relevant units of rad s^{-1} , with an additional factor of $\hbar$ required to convert to SI units.)

Here, γ_I and γ_S are the gyromagnetic ratios of the species I and S respectively, with r_{IS} being the distance between the two spins. The magnetic permeability of free space, μ_0 , is $4\pi \times 10^{-7} \text{ N.A}^{-2}$. The $Y_{2,q}$'s are the 2nd order spherical harmonics, and the operators I_q and S_q are the spin operators for species I and S respectively.

In a similar fashion as the QCC, the *dipolar coupling constant* (DCC) is defined as

$$f_D = \frac{\omega_D}{2\pi} = \frac{\mu_0}{4\pi} \frac{\gamma_I \gamma_S \hbar}{r_{IS}^3} \text{ (in units of Hz).}$$

Considering the instance of an aromatic ^{13}C bonded to a

^1H , a case frequently encountered in the work presented herein, the DCC is 22.7 kHz when the distance r_{CH} is taken to be 1.1 Å.

The l^{th} order spherical harmonics are tensors of rank l and so transform in the now-familiar manner

$$Y_q^{(2)}(S) = \sum_{q'=-l}^l D_{q'q}^{(2)}(\tilde{\Omega}_{SS'}) Y_{q'}^{(2)}(S') \quad (12)$$

c. Chemical shift interaction

This energy contribution arises due to the interaction of the nuclear spin with the magnetic field generated by electron currents, which have been induced by the external magnetic field.

$$\hat{H}^{\text{CS}} = -\vec{\mathcal{M}}^T \cdot \vec{\delta} \cdot \vec{B}_0 \quad (13a)$$

(The Hamiltonian is expressed in the computationally relevant units of rad s^{-1} , with an additional factor of $\hbar$ required to convert to SI units.)

Here, $\vec{\delta}$ represents the *chemical shift tensor* in Cartesian coordinates. It quantifies the potentially anisotropic (and therefore, orientation dependent) impact that the induced electron currents have on the nuclear spin. The induced field is given by $\vec{\delta} \cdot \vec{B}_0$. In the most general frame, this tensor will have nine independent components. However, as in the previous two cases, the tensor will be locally expressed in the PAS frame where it is diagonal. The three independent components in this frame are usually expressed as combinations of the Cartesian components $\{\delta_{xx}^{\text{PAS}}, \delta_{yy}^{\text{PAS}}, \delta_{zz}^{\text{PAS}}\}$ as follows:

$$\begin{aligned}
\delta_{iso} &= \frac{1}{3}(\delta_{xx}^{PAS} + \delta_{yy}^{PAS} + \delta_{zz}^{PAS}) \\
\Delta_{CS} &= \delta_{zz}^{PAS} - \delta_{iso} \\
\eta_{CS} &= \frac{\delta_{xx}^{PAS} - \delta_{yy}^{PAS}}{\Delta_{CS}}
\end{aligned} \tag{14}$$

δ_{iso} is the isotropic chemical shift, Δ_{CS} is the chemical shift anisotropy (CSA) and η_{CS} is the chemical shift asymmetry.

The chemical shift interaction in Equation 13a can be recast in terms of spherical components as well. The first step is to notice that since this interaction includes an explicit dependence on the external magnetic field, we can describe the direction of the magnetic field in the PAS frame by spherical coordinates (θ, ϕ) :

$$\hat{H}^{CS} = -\gamma B_0 \left\{ \delta_{iso} \sqrt{4\pi} Y_0^{(0)} + \Delta_{CS} \left(\sqrt{\frac{4\pi}{5}} Y_0^{(2)} + \eta_{CS} \sqrt{\frac{2\pi}{15}} (Y_2^{(2)} + Y_{-2}^{(2)}) \right) \right\} \tag{13b}$$

The transformation properties of the individual terms can be inferred from Equation 13b.

For reference, the values of the elements of the chemical shift tensor in the PAS frame for some base carbons in RNA and DNA are given in Ying et al.⁶² In the present manuscript, the base carbon C6 for uridine residues is taken to have an axially symmetric tensor ($\eta_{CS} = 0$) and the anisotropy is taken to be 212 ppm.

2. Experimental observables

In this section, I will cover the theory behind the basic NMR observables relevant to the discussions in the following chapters. These include solid state and solution relaxation times and the solid state line shape.

a. Relaxation times

The general principle is to find an expression for the decay rate (the inverse of the relaxation time) of the expectation value of a particular function of the spin operators. Since we have a system that evolves under the influence of stochastic forces, the natural framework for describing the time evolution of the spin is in terms of the density matrix $\hat{\rho}$ and the Liouville operator formalism:

$$\frac{d\hat{\rho}(t)}{dt} = -i[\hat{H}(t), \hat{\rho}(t)] \quad (15)$$

(As mentioned when describing the Hamiltonians, the units of $\hat{H}$ in Equation 15 are rad s^{-1} .)

The expectation value of a function of the spin operators is given by

$$\langle \hat{A}(\vec{I}) \rangle = \text{Tr}\{\hat{\rho}\hat{A}\} \quad (16)$$

Thus, we utilize Equation 15 in solving the time evolution equation of the expectation value of the function in Equation 16.

Before proceeding to specific forms of the functions in Equation 16 and their respective relaxation rates (or times), let us first consider the solution to Equation 15. In any relaxation time analysis, the Hamiltonian consists of the Zeeman interaction and a much smaller lattice-coupling interaction, such as the electric quadrupolar coupling or the magnetic dipolar interaction. It is assumed that the relaxation of the spin system occurs after the perturbing pulse is turned off. Under these circumstances, it is easier to find a perturbative solution to the density matrix equation after the density matrix and Hamiltonian are transformed to a frame where the Zeeman interaction vanishes. This is essentially the rotating frame where the frequency of rotation is matched to the Larmor precession frequency. It can be shown that an equation analogous to

Equation 15 holds for the rotated frame (starred in the equation below) density matrix and the rotated frame Hamiltonian (which now includes only the interaction part):

$$\frac{d\hat{\rho}^*(t)}{dt} = -i[\hat{H}_{\text{int}}^*(t), \hat{\rho}^*(t)] \quad (17)$$

The solution to this equation is $\hat{\rho}^*(t) = \hat{\rho}^*(0) - i \int_0^t [\hat{H}_{\text{int}}^*(t'), \hat{\rho}^*(t')] dt'$, which can be expanded out to

2nd order in the interaction Hamiltonian as

$$\hat{\rho}^*(t) \approx \hat{\rho}^*(0) - i \int_0^t [\hat{H}_{\text{int}}^*(t'), \hat{\rho}^*(0)] dt' - \int_0^t \int_0^{t'} [\hat{H}_{\text{int}}^*(t'), [\hat{H}_{\text{int}}^*(t''), \hat{\rho}^*(0)]] dt'' dt' \quad (18)$$

The equation is turned back into a differential equation, to yield the 2nd order density matrix evolution equation

$$\frac{d\hat{\rho}^*(t)}{dt} = -i[\hat{H}_{\text{int}}^*(t), \hat{\rho}^*(0)] - \int_0^t [\hat{H}_{\text{int}}^*(t), [\hat{H}_{\text{int}}^*(t'), \hat{\rho}^*(0)]] dt' \quad (19)$$

This equation is used in conjunction with Equation 16 to yield expression for the relaxation times. When the nested commutator on the RHS of Equation 19 is expanded out, terms containing products of the rotating frame interaction Hamiltonian are obtained. Further, all Hamiltonian terms are ensemble-averaged over the stochastic degrees of freedom, with the constraints that the ensemble average of the Hamiltonian is 0. Thus, the only remaining terms of importance are the averages over the two-Hamiltonian products, which are of the form

$$\langle \langle \alpha | \hat{H}_{\text{int}}^*(t) \hat{H}_{\text{int}}^*(t') | \beta \rangle \rangle_{\text{ensemble}} \quad (20)$$

At this stage, the rotated frame Hamiltonian is converted back to the lab frame (the original transformation was done for the sake of applying perturbation theory), and the so-called *secular approximation* is applied. This approximation only accounts for transitions between

eigenstates of the Zeeman Hamiltonian that are either exactly degenerate or nearly so. This is usually enforced because the interaction Hamiltonian is almost always of a substantially smaller amplitude than the Zeeman splittings and cannot cause significant transitions between widely separated levels.

If we consider the situation where the stochastic modulation of the interaction Hamiltonian in time arises from the diffusive motion of the molecule or one of its parts, we see that the time dependence of the interaction Hamiltonian is contained in the orientation dependent spherical tensor elements discussed previously. These tensor elements, in turn, are first described in their respective PAS frames and then rotated into the lab frame using the Wigner rotation matrix elements. Therefore, in the end, the information about molecular motion is contained in two-time correlation functions over Wigner rotation matrix elements:

$$C(t) = \left\langle D_{mn}^{(2)*}(\vec{\Omega}_{PL}(0)) D_{m'n'}^{(2)}(\vec{\Omega}_{PL}(t)) \right\rangle_{motion} \quad (21)$$

The problem of describing the motion of a specific system as a function of either a continuous potential or a discrete-site jump process is formulated in terms of finding an expression for the correlation function in Equation 21. Finding an expression for this correlation function for the motion of a molecule that is both tumbling in an isotropic solution and exchanging between multiple conformations will be the focus of Chapters III and IV.

Returning to the expression for the relaxation rates, we see that during the course of re-transforming back to the lab frame, the factors of $e^{\pm i H_{Zeeman} t}$ acting on the spin eigenstates will end up producing frequency exponentials of the form $e^{\pm i \omega t}$, where the ω 's are frequencies characteristic to the particular spin functions $\hat{A}(\vec{I})$ and interaction Hamiltonians. The effect of these exponentials when integrated over time is to produce Fourier cosine transforms of the

correlation functions, the so-called *spectral density functions*, sampled at the characteristic frequencies:

$$J(\omega) = \text{Re} \left[\int_0^\infty C(t) e^{i\omega t} dt \right] = \int_0^\infty C(t) \cos(\omega t) dt \quad (22)$$

To proceed further we need to describe specific spin functions. We will consider only the four relaxation times used in this analysis.

i. Solid state deuterium T_{1Z}

This relaxation time quantifies the decay of the expectation value of the longitudinal (z) spin $\hat{I}_z$:

$$\frac{d\langle \hat{I}_z \rangle}{dt} = -\frac{\langle \hat{I}_z - \hat{I}_z^{eq} \rangle}{T_{1Z}} \quad (23)$$

Experimentally, this time is measured by inverting the Boltzmann-distributed populations and measuring the rate of re-establishment of equilibrium. The expression for the relaxation time is:

$$\frac{1}{T_{1Z}} = \frac{3}{8} \left(\frac{e^2 q Q}{\hbar} \right)^2 [J(\omega_D) + 4J(2\omega_D)] \quad (24)$$

where ω_D is the deuterium Larmor frequency.

ii. Solid-state deuterium T_{1Q}

This relaxation time quantifies the decay of the expectation value of the quadrupolar

order $3\hat{I}_z^2 - \hat{I}^2$:

$$\frac{d\langle 3\hat{I}_z^2 - \hat{I}^2 \rangle}{dt} = -\frac{\langle 3\hat{I}_z^2 - \hat{I}^2 - (3\hat{I}_z^2 - \hat{I}^2)^{eq} \rangle}{T_{1Q}} \quad (25)$$

Experimentally, this is measured by inverting one-half of the line shape and measuring the rate of decay of the intensity difference between the inverted and non-inverted horns of the line shape. The expression for the relaxation time is:

$$\frac{1}{T_{1Q}} = \frac{9}{8} \left(\frac{e^2 q Q}{\hbar} \right)^2 [J(\omega_D)] \quad (26)$$

iii. Solution carbon-13 T_{1Z}

This represents the same physical quantity as the solid-state T_{1Z} , but the expression differs on account of the interaction Hamiltonians being different:

$$\frac{1}{T_1} = R_1 = \frac{d^2}{4} [J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)] + c^2 J(\omega_C) \quad (27)$$

$$d = \left[\frac{\mu_0 \hbar \gamma_H \gamma_C}{8\pi^2 r_{CH}^3} \right], c = \left(\frac{\omega_C}{\sqrt{3}} \right) (\Delta)$$

where ω_C is the carbon-13 Larmor frequency and Δ is the chemical shift anisotropy (CSA). This form holds for the case that only one proton is attached to the carbon atom.

iv. Solution carbon-13 T_2 (equivalent to $T_{1\rho}$ under certain conditions)

This relaxation time quantifies the decay of the expectation value of the transverse

spin $\hat{I}_+ = \hat{I}_X + i\hat{I}_Y$:

$$\frac{d\langle \hat{I}_+ \rangle}{dt} = i\omega \langle \hat{I}_+ \rangle - \frac{\langle \hat{I}_+ \rangle}{T_2} \quad (28)$$

Experimentally, this is measured by measuring the decay rate of the spin echo intensity, using a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence.^{63,64} In practice for the carbons considered in this analysis, however, the rotating frame z-relaxation time $T_{1\rho}$ was measured instead.^{11,65} Under the application of a weak spin-lock field and the assumption of Lorentzian spectral densities, $T_{1\rho}$ has been shown to have the same spectral density information as T_2 .⁶⁶ The expression for T_2 is:

$$\frac{1}{T_2} = R_2 = \frac{d^2}{8} \left[\begin{array}{l} 4J(0) + J(\omega_H + \omega_C) + 3J(\omega_C) \\ + 6J(\omega_H) + 6J(\omega_H + \omega_C) \end{array} \right] + \frac{c^2}{6} [3J(\omega_C) + 4J(0)] \quad (29)$$

$$d = \left[\frac{\mu_0 h \gamma_H \gamma_C}{8\pi^2 r_{CH}^3} \right], c = \left(\frac{\omega_C}{\sqrt{3}} \right) (\Delta)$$

where ω_C is the carbon-13 Larmor frequency and Δ is the chemical shift anisotropy (CSA). This form holds for the case that only one proton is attached to the carbon atom.

v. Nuclear Overhauser Effect (NOE)

This is a form of relaxation experiment conducted under a certain constraint: the relaxation of one spin I coupled with another spin S through dipolar coupling is studied, while the S spin is continually *saturated*, i.e. while the populations of the z-projections of the S spin are equalized through the application of an rf pulse. The new equilibrium value of the I spin magnetization is different in general to the value that would be obtained if the S spin were not irradiated. The term NOE is used to refer to ratio:

$$\frac{\langle \hat{I}_Z^{SS} \rangle}{\langle \hat{I}_Z^0 \rangle} = 1 + \frac{\sigma_{IS}^{NOE}}{R_{1Z}^I} \frac{\gamma_S}{\gamma_I} \quad (30)$$

where $\langle \hat{I}_Z^{ss} \rangle$ is the new steady state magnetization value in the presence of the saturating pulse, $\langle \hat{I}_Z^0 \rangle$ is the steady state value in the absence of such a pulse, σ_{IS}^{NOE} is the rate of cross-relaxation between the two spins in the presence of the pulse, and R_{1Z}^I is the “self-relaxation” rate (i.e. the coefficient of the I magnetization at time t in the time-evolution equation for spin I itself) of spin I.

Experimentally, the NOE is measured by recording two spectra, one with saturation of the partner spin and one without, and then taking the ratio between the intensities.

The expression for this effect in terms of spectral densities is:

$$\text{NOE} = 1 + \frac{d^2}{4} \left(\frac{\gamma_H}{\gamma_C} \right) \left[6J(\omega_H + \omega_C) - J(\omega_C - \omega_H) \right] T_1 \quad (31)$$

$$d = \left[\frac{\mu_0 \hbar \gamma_H \gamma_C}{8\pi^2 r_{CH}^3} \right], c = \left(\frac{\omega_C}{\sqrt{3}} \right) (\Delta)$$

b. Solid state quadrupolar coupling line shape

The solid state line shape is the combined spectrum of Larmor frequencies across all orientations of the bond in question. In a solid state sample, the molecules may be oriented in a limited number of crystallite directions, or they may be randomly oriented. The transition frequencies in going from $M = 0 \rightarrow M = 1(+)$ and $M = -1 \rightarrow M = 0(-)$ observed for a given sample depend on the distribution of orientations of the PAS frame of the EFG tensor relative to the lab frame, and are given by 1st order perturbation theory:

$$\omega_{i,\pm} = \omega_D \pm \frac{3}{4} \frac{e^2 q Q}{\hbar} P_2(\cos(\theta_{PL}^i)) \quad (32)$$

where ω_D is the deuterium Larmor frequency, i represents the jump-site label for a particular bond orientation and θ_{PL}^i is the angle between the z-axis of the PAS frame (P) and the z-axis of the lab frame (L).

The time evolution equation under consideration for a single spin in the sample is that for the transverse magnetization (Equation 28), but for simplicity we will ignore the effects of relaxation in the analysis that follows. This is a reasonable approximation as the typical Larmor frequencies are on the time scale of nanoseconds, while the deuterium T_2 scale is on the order of milliseconds. It is possible to include the relaxation term in a more complete analysis, and in fact partially relaxed line shapes are studied, but such a discussion would distract from the presentation of the core ideas behind motional modulation of the line shape. The relevant equation of motion for the entire sample is therefore:

$$\frac{d\vec{M}_+(t)}{dt} = (i\vec{\omega} + \vec{R})\vec{M}_+(t) \quad (33)$$

In Equation 33, $\vec{M}_+$ is the vector whose components are the magnetizations at each site, $\vec{\omega}$ is the diagonal matrix with the orientation dependent frequencies as its elements and $\vec{R}$ is the jump matrix that arises due to motional exchange between various orientations of a single atomic bond. This jump matrix intermixes the magnetizations at each site, and carries information regarding the particular motional model for the system.

In some cases it is possible to ignore slower motions and apply the simplest theory of line shape modulation by a single internal motion; however, in the interest of generality, we present here a theory for two uncoupled motions. A simplification to the case of a single-axis motion would require only that the number of intermediate frames between the PAS and lab frames be reduced by one, and is not explicitly considered here.

To describe line shape modulation by two uncoupled motions, four frames are required. The ^2H solid-state line shape is dominated by the interaction of the nuclear quadrupole moment with external electric field gradients (EFGs). The EFG tensor is approximated to have a negligible asymmetry for hydrogens bonded to all carbons. The principal axis system (PAS) is defined so that the z axis is parallel to the C-D bond axis. Internal motions, which modulate the relative orientation of the PAS and the L (lab) frame, are mathematically described using an additional frame of reference for each motion. The first intermediate frame which is fixed to the molecular framework will be designated C. The orientation of the C frame depends upon the nature of the motion. Similarly, the second motion determines the orientation of second intermediate frame (M) relative to the C frame. The geometric progression is $PAS \xrightarrow{\vec{\Omega}_{PC}} C \xrightarrow{\vec{\Omega}_{CM}} M \xrightarrow{\vec{\Omega}_{ML}} L$.

Internal molecular motions modulate the Euler angles $\vec{\Omega}_{PC} = (\alpha_{PC}, \beta_{PC}, \gamma_{PC})$ and $\vec{\Omega}_{CM} = (\alpha_{CM}, \beta_{CM}, \gamma_{CM})$, which quantify the mutual orientation of the PAS and C frame and C and M frames, respectively. As a result of the internal motion, these solid angles become time dependent, e.g. $\vec{\Omega}_{PC}(t) = (\alpha_{PC}(t), \beta_{PC}(t), \gamma_{PC}(t))$. In a polycrystalline state, the line shape is averaged over all orientations of the sample; thus, the solid angle $\vec{\Omega}_{ML} = (0, \theta_{ML}, \phi_{ML})$ that relates the M and L frames is distributed randomly. (Only two angles are necessary due to the axial symmetry of the lab frame, and the definitions in terms of Euler angles and spherical coordinate angles are equivalent for the case of only two non-zero angles.) With these conventions, the static ^2H line shape intensity is a function of the frequency ω and the intervals between quadrupolar echo pulses τ_1 (period between the two 90° pulses) and τ_2 (period between the 2nd 90° pulse and the signal acquisition) according to:

$$I(\omega, \tau_1, \tau_2) = \frac{1}{8\pi^2} \int_0^{2\pi} d\varphi_{ML} \int_0^\pi d\theta_{ML} \sin\theta_{ML} I(\omega, \tau_1, \tau_2, \theta_{ML}, \varphi_{ML}) \quad (34)$$

where the orientation-dependent line shape is the Fourier transform of the projected time domain response (scalar projection of $\vec{M}_+(t)$) as seen below in Equation 36):

$$I(\omega, \tau_1, \tau_2, \vec{\Omega}_{ML}) = \text{Re} \int_{-\infty}^{+\infty} m(t) e^{-i\omega t} dt \quad (35)$$

where above right hand side summation is obtained from the expression for $m(t)$:

$$m(t) = \vec{1} \cdot e^{(t+\tau_2)\vec{A}} e^{\tau_1\vec{A}^*} \vec{M}_+(0) \quad (36)$$

For a discrete-jump motion between C-D bond orientations, the vector $\vec{M}_+(0)$ has as its components the initial magnetizations at each site. The matrix operator $\vec{A} = i\vec{\omega} + \vec{R}$ ($\vec{A}^*$ is the complex conjugate) is the time-evolution operator in Equation 33. The frequencies defined in Equation 32 are expanded as functions of the 2nd order Wigner rotation matrices for the case of the three coordinate frame transformations in going from the PAS to lab frame given earlier:

$$\omega_{i,\pm} = \omega_D \pm \frac{3}{4} \frac{e^2 q Q}{\hbar} \sum_{a,b=-2}^{+2} D_{0a}^{(2)}(\Omega_i^{PC}) D_{ab}^{(2)}(\Omega_i^{CM}) D_{b0}^{(2)}(\Omega_i^{ML}) \quad (37)$$

The quadrupolar coupling constant $QCC = e^2 q Q / \hbar$ varies only slightly among aliphatic deuterons (~170 kHz) and aromatic deuterons (~180 kHz). The time-dependent magnetization in Equation 36 can be calculated by expanding $\vec{A}$ and $\vec{A}^*$ in terms of their eigenvalues and eigenvectors (see for example, Meints et al²¹).

Having set up the theoretical background for NMR observables, we now move on to the original research that forms the crux of this manuscript.

Chapter III

A Slow Exchange Model of Non-rigid Rotational Motion in RNA for combined Solid-state and Solution NMR studies

Abstract

Functional RNA molecules are conformationally dynamic and sample a multitude of dynamic modes over a wide range of frequencies. Thus, a comprehensive description of RNA dynamics requires the inclusion of a broad range of motions across multiple dynamic rates which must be derived from multiple spectroscopies. Here we describe a slow conformational exchange theoretical approach to combining the description of local motions in RNA that occur in the ns- μ s window and are detected by solid-state NMR with non-rigid rotational motion of the HIV-1 TAR RNA in solution as observed by solution NMR. This theoretical model unifies the experimental results generated by solution and solid-state NMR and provides a comprehensive view of the dynamics of HIV-1 TAR RNA, a well-known paradigm of an RNA where function requires extensive conformational rearrangements. This methodology provides a quantitative atomic level view of the amplitudes and rates of the local and collective displacements of the TAR RNA molecule, and provides directly motional parameters for the conformational capture hypothesis of this classical RNA-ligand interaction.

1. Introduction

The transactivation response element (TAR) RNA from the human immunodeficiency virus type-1 (HIV-1) is required for HIV replication making it an important drug target, but it is also an ideal example of the way in which structural flexibility facilitates the interaction of RNA

with ligands.⁶⁷⁻⁷⁰ Binding of HIV-1 Tat protein to TAR RNA requires a large conformation change of the RNA and a significant decrease in conformational disorder in both RNA and the polypeptide.^{6,7,9,71,72} The structural properties of these conformational states are partly known,^{6,8,73,74} yet, despite some recent progress,^{14,75} the exact atomic level displacements leading to these changes and many aspects of the changes in dynamics that accompany them remain to be studied.

The local dynamics of TAR RNA have been analyzed by solution NMR relaxation methods and analyzed extensively, using for the most part the Model-Free approximation.^{14,58,72} More recently, heterogeneities in $^{13}\text{C}/^1\text{H}$ line widths and relaxation rates in and around the bulged loop of TAR and the measurement of residual dipolar couplings (RDCs) have provided strong evidence for the existence of collective motions at ns- μs rates of the distal helices about a dynamic hinge associated with the bulged loop of TAR.^{11,14} A meso-scale view of TAR functional dynamics derived from these studies suggest that TAR samples a manifold of structures that includes those observed when TAR is bound to various ligands, implying that TAR RNA is recognized through the opportunistic capture of pre-existing and possibly rare conformations (conformational capture). However, while RDCs can report on the amplitudes of motion occurring at any rate fast enough to pre-average the $^{13}\text{C}-^1\text{H}$ dipolar coupling constant ($< \text{ms}$), they do not provide any direct information on rates. Therefore, the time scale of these internal motions could only be broadly and indirectly inferred by a process of elimination when faster (sub-ns) and slower (ms) motions are not observed.^{12,14,76} Solution NMR relaxation methods can probe ms and ns-ps motions, but cannot directly detect motions within the ns- μs time frame. In contrast, ^2H solid-state NMR (ssNMR) line shape and relaxation studies provide direct information on motions occurring at this rate. By combining solution and solid-state NMR

studies, it should become possible to study the complete range of motions experienced by this paradigmatic RNA. However, this requires the development of theoretical methods to merge the results provided by the two techniques into a common interpretative model. Here we achieve this goal by first extracting detailed information on internal motions from the solid-state data, and then by incorporating these motions into a model of overall tumbling obtained from hydrodynamic considerations.

Local motions alter the global structure of the RNA, so that it is necessary to consider how the concerted motions of helical domains or single nucleotides modulate the macroscopic rotational diffusion of the RNA in solution and thus impact NMR relaxation rates.. However, the correct treatment of non-rigid tumbling of RNA poses a serious challenge because: 1) the correlation times of their motion may approach those of the overall tumbling,^{12,14} and 2) these motions may deform the tumbling molecule and affect its hydrodynamics.⁷⁷ To solve the problem of motional coupling, several creative approaches have been taken, among them mode coupling analyses and domain elongation.^{12,14,78} In this paper we consider the situation, motivated by ssNMR models,⁷⁵ where the rate of hydrodynamically-significant domain motions is substantially slower than the tumbling correlation time. This *slow-exchange* formalism considers weighted populations of conformers with arbitrary diffusion tensors, each of which contributes independently to the relaxation rates. Faster local motions that are reasonably assumed to have little effect on the tumbling rates are integrated into the separate conformer populations.

Using this slow-exchange model, we study dynamics at U-38, a helical base site in HIV-1 TAR RNA at which the assumption of motional decoupling is valid (Figure 1). Rates and amplitude of motions of the upper helix were obtained using ²H ssNMR applied to a TAR

sample where uridine-5,6- $^2\text{H}_2$ was incorporated at U38.^{75,79} While comparisons to solution data in this manuscript are limited to solution relaxation times for the U38- $^{13}\text{C6}$ site, the method is applicable to any site and residue.

We focus this manuscript on the concerted dynamics of the TAR RNA construct shown in Figure 1A. In section 2 we introduce a slow exchange theory of concerted dynamics. In section 3A, constraints on the internal motions of TAR RNA obtained from solid-state NMR are explicitly stated and subsequently incorporated into the slow exchange correlation functions. A brief summary is also provided in section 3B of the method for calculating rotational diffusion tensors as implemented by HYDRONMR. In section 4, the theory of slow conformational exchange of section 2 is used to simulate the spin lattice and transverse relaxation rates for a ^{13}C spin in the upper helix of TAR. In section 5, the validity of the assumptions made in constructing a slow conformational exchange correlation function for TAR RNA rotational diffusion is discussed.

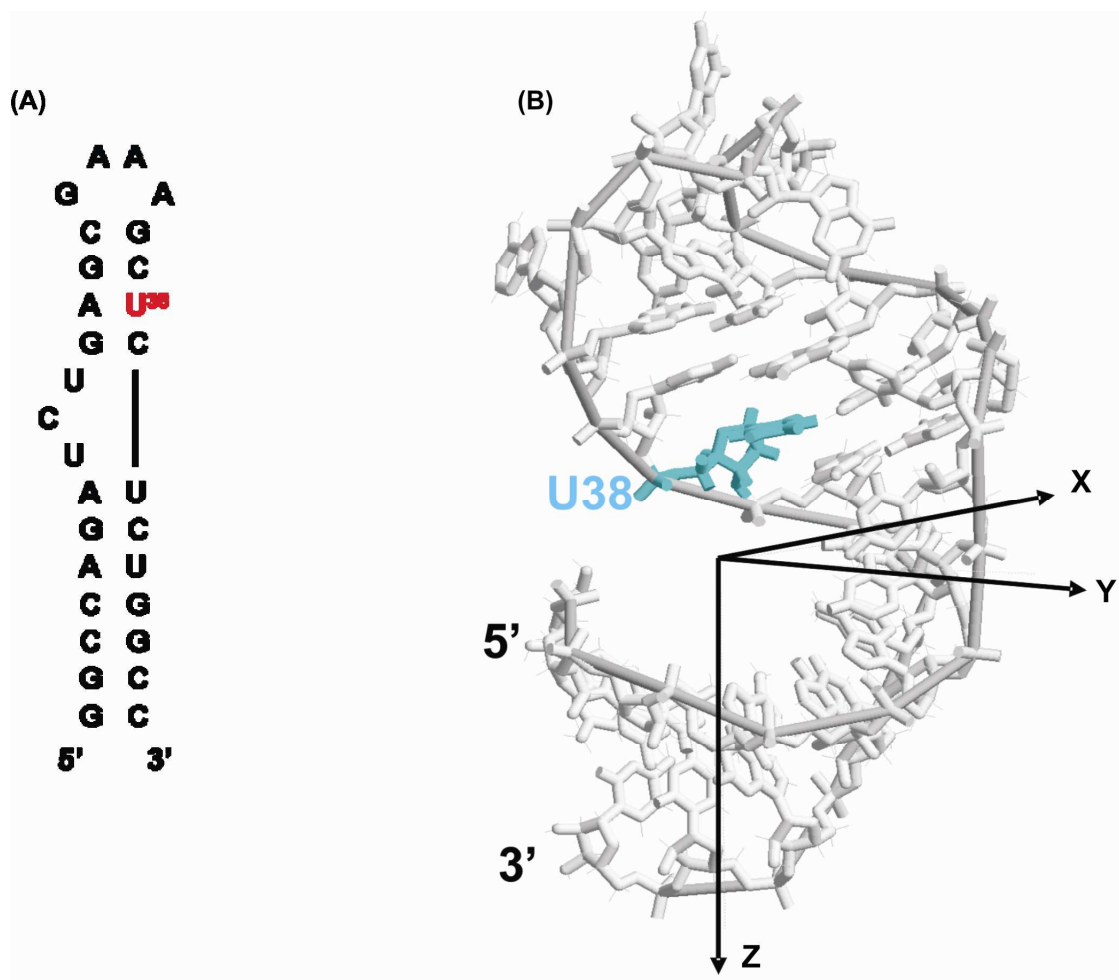


Figure III.1. A) Sequence and secondary structure of the HIV-1 TAR RNA construct used in this study, which includes a stable tetraloop for convenience instead of the native TAR loop. The U38 residue, which was deuterium labeled at the H5 and H6 sites for solid-state NMR and which is the subject of the present theoretical study, is highlighted in red. B) Lowest energy free TAR structure 1ANR. Hydrogen atoms have been removed from the figure for clarity. The Principal Axis System of the rotational diffusion tensor (PASd) is superimposed upon the molecule.

2. Theory

a. NMR Relaxation in the Presence of Slow Conformational Exchange

Structural changes that alter significantly the shape of the TAR RNA molecule can perturb the global rotational motions of TAR in solution. Here we consider the problem of how slow conformational exchange perturbs the overall molecular tumbling motion and thus affects the relaxation of nuclear spins in the TAR RNA.

In deriving the model, we assume the existence in a RNA molecule of several types of motion occurring at different time scales. These include: 1) localized motions of individual structural units (nucleotide bases, furanose rings and small segments of the phosphodiester backbone); 2) long range conformational changes involving helical domains, the bulged loop, etc.; and 3) overall rotational motions of the TAR RNA molecule in solution. Prior studies of DNA oligonucleotides indicate that localized motions of small structural units in nucleic acids occur on the ns-ps timescale and are independent of overall rotational motions,^{29,80} which for a small RNA like TAR occurs on a time scale of 5-10 ns. In the theory that follows we therefore assume the localized base librations have no effect on overall molecular rotational diffusion. However, larger conformational rearrangements such as the reorientation of the helical domains of TAR will change the molecular shape, thereby modulating the molecular rotational diffusion tensor.¹² Zhang and Al-Hashimi have similarly considered the modulation of the magnetic susceptibility tensors due to domain motion.⁸¹

The general problem of the coupling of molecular conformational changes to overall molecular tumbling has been treated recently in the context of the model-free analysis by Wong et al.⁸² for changes in the dimensions of idealized molecular shapes including the radii of spheres and the aspect ratio of cylinders. As in the theory of Wong et al.,⁸² we simplify the problem by treating the conformational motions in RNA as exchange between discrete structural forms, and we then solve the molecular diffusion problem in the presence of exchange between discrete conformers with distinct diffusion tensors.

Below we present the standard theory of correlation functions for single conformer diffusion (Equations 1 – 6), the diffusion equation for multiple conformers (Equation 7), as well as our own contribution in the form of the slow exchange formalism (Equations 8 and 9).

The free-diffusion equation can be written in terms of the diffusion tensor elements and the angular momentum operators:⁸³

$$\frac{\partial}{\partial t} P(\vec{\Omega}, t) = - \sum_{i,j=1}^3 \hat{L}_i D_{ij} \hat{L}_j P(\vec{\Omega}, t). \quad (1)$$

In equation 1, $\hat{L}_i$ is the angular momentum operator about the i^{th} axis, D_{ij} is the ij component of the rotational diffusion tensor, and $P(\vec{\Omega}, t)$ is the probability that the molecular axis system will have rotated through an Euler angle vector of $\vec{\Omega}$ at a time t relative to its initial orientation in the laboratory frame at time $t = 0$. Following Favro,⁸³ the right hand side (RHS) of equation 1 can be expanded in the principal axis frame of the diffusion tensor (PASd) as follows:

$$\sum_{i,j=1}^3 \hat{L}_i D_{ij} \hat{L}_j = D^+ \hat{L}^2 + (D_z - D^+) \hat{L}_z^2 + D^- (\hat{L}_x^2 - \hat{L}_y^2) \quad (2)$$

where $D^\pm = \frac{1}{2}(D_x \pm D_y)$, and D_x , D_y and D_z are the eigenvalues of the diffusion tensor. In this formalism, the problem is thus reduced to finding the eigenvalues and eigenfunctions of the operator shown on the RHS of Equation 2. However, $\hat{L}_z$ does not commute with the RHS operator, and a complete set of eigenfunctions cannot be found for this operator as long as $D_x \neq D_y$. It is nonetheless possible to find eigenfunctions of the RHS operator in equation 2 for the first few eigenvalues of $\hat{L}^2$ in an iterative manner. Utilizing the orthogonality of the first few resulting eigenfunctions, and the fact that the vector orientation of the C-H bond can be expressed as an expansion in these eigenfunctions, it is possible to calculate exactly two-time correlation functions of the C-H bond orientations relative to the laboratory frame, which has the form:

$$C(t) = \langle P_2(\hat{n}(t) \cdot \hat{n}(0)) \rangle = \frac{2}{5} \sum_{l=1}^5 a_l e^{-t/\tau_l}. \quad (3)$$

In equation 3, $P_2(x)$ is the second order Legendre polynomial, $\hat{n}(t)$ is the C-H bond orientation at time t , $\hat{n}(0)$ is the C-H bond orientation at $t=0$, and the amplitudes $a_l(\vec{A}^0, \vec{B}^0)$ are functions of the C-H bond orientation vector in the PASd frame, designated $\vec{A}^0 = (A_1, A_2, A_3)$ and $\vec{B}^0 = (B_1, B_2, B_3)$, where the superscript 0 indicates that the vector orientation is referenced to the PASd frame. In the absence of internal motion, $\vec{A}^0 = \vec{B}^0$. Equation 3 is equivalent to equation 6.29 in Favro⁸³ and equation 35 in Woessner.⁸⁴ The five amplitudes are defined as:

$$a_1 = \frac{3}{4}(F + G); a_2 = 3A_2A_3B_2B_3; a_3 = 3A_3A_1B_3B_1; a_4 = 3A_1A_2B_1B_2; a_5 = \frac{3}{4}(F - G) \quad (4)$$

where $F = \sum_i A_i^2 B_i^2 - \frac{1}{3}$, $G = \frac{1}{\Delta} \left[-D + \sum_{i,j \neq k} D_i \{ A_i^2 B_i^2 + A_j^2 B_k^2 + A_k^2 B_j^2 \} \right]$, and

$\Delta = (D_1^2 + D_2^2 + D_3^2 - D_1D_2 - D_2D_3 - D_3D_1)^{1/2}$. The correlation times τ_l are functions of the

diffusion tensor eigenvalues alone and are given by:

$$\tau_1^{-1} = 6D - 2\Delta, \tau_2^{-1} = 3(D_1 + D), \tau_3^{-1} = 3(D_2 + D), \tau_4^{-1} = 3(D_3 + D), \tau_5^{-1} = 6D + 2\Delta, \quad (5)$$

where $D = \frac{1}{3}(D_1 + D_2 + D_3)$, and Δ is the same as above. The diffusion tensor eigenvalues are designated such that $D_3 > D_2 > D_1$.

The occurrence of five correlation times in these expressions can be understood in terms of the number of parameters needed to describe a diffusion tensor with three different eigenvalues. Given the symmetry of the solvent, the diffusion tensor in an arbitrary frame is

symmetric; thus, five unique parameters plus the trace are required to describe the tensor completely. The trace quantifies the overall scale of the diffusion tensor eigenvalues, and occurs in the exponential term $e^{-6Dt} = e^{-2Tr\{\bar{D}\}t}$ that multiplies the entire correlation function. This leaves five unique times and a scale-setting trace term in the correlation function.

We can incorporate discrete site jump models for local motions of the C-H bond vector into the rigid tumbling correlation function (equation 3) by averaging the amplitudes (a_l 's) over the local motion. The resulting expression for the general case of N_{site} local discrete-site jumps is then:

$$\begin{aligned} \langle C(t) \rangle_{\text{internal}} &= \frac{2}{5} \sum_{l=1}^5 \left\langle a_l \left(\bar{A}^0(t), \bar{A}^0(0) \right) \right\rangle_{\text{internal}} e^{-t/\tau_l} \\ &= \frac{2}{5} \sum_{l=1}^5 e^{-t/\tau_l} \sum_{j=1}^{N_{\text{sites}}} \sum_{k=1}^{N_{\text{sites}}} a_l \left(\bar{A}^0(t, j), \bar{A}^0(0, k) \right) P \left(\bar{A}^0(t, j) \mid \bar{A}^0(0, k) \right) P_0 \left(\bar{A}^0(0, k) \right) \end{aligned} \quad (6)$$

Additional angular brackets show explicitly the averaging of the correlation function $C(t)$ over the internal degrees of freedom. In the second line of Equation 6, the sites at time t are labeled by the index j and the sites at time $t=0$ are labeled by k . As seen here more explicitly, the a_l are functions of: $\bar{A}^0(t)$, the orientation of the C-H bond in the diffusion tensor frame at time t , $\bar{A}^0(0)$, the orientation of the bond at time $t = 0$, and of the diffusion tensor eigenvalues. The probability that a C-H bond lies at the orientation site j given that it was in site k at $t=0$ is given by the transition probability $P \left(\bar{A}^0(t, j) \mid \bar{A}^0(0, k) \right)$, while the *a priori* probability is given by $P_0 \left(\bar{A}^0(0, k) \right)$.

Equation 6 implicitly assumes that small amplitude local motions of the bases in RNA do not result in significant changes in the rotational diffusion tensor, and are thus independent of the overall molecular tumbling. This assumption is supported by the experimental data for U38,

U25, and U23 in TAR RNA.⁸⁵ However, reorientational motions of the helical domains of TAR RNA observed in RDC studies and in solid-state NMR studies of U38 are expected to modify the shape of the molecule in such a way as to perturb its hydrodynamic properties, effectively making the rotational diffusion tensor time-dependent. Treating these additional motions as exchanges of the TAR RNA molecule between N discrete structures, the diffusion equation becomes⁸²

$$\frac{\partial}{\partial t} P_{\alpha}(\vec{\Omega}, t) = - \sum_{i,j=1}^3 \hat{L}_i D_{ij}^{\alpha} \hat{L}_j P_{\alpha}(\vec{\Omega}, t) + \sum_{\beta=1}^N R_{\alpha\beta} P_{\beta}(\vec{\Omega}, t) \quad (7)$$

where the second term describes conformational transitions from conformer β to conformer α with rate $R_{\alpha\beta}$. The assumption in this equation is that the individual conformers tumble rigidly, punctuated by instantaneous shifts to rigidly-tumbling conformers with different diffusion tensors, so that relaxation times are obtained as averages over these multiple conformations. Wong et al.⁸² solved equation 7 for exchange between idealized structures like spheres and cylinders, but equation 7 can also be treated in a straightforward manner for completely anisotropic diffusion tensors, provided a significant rate separation exists between the molecular tumbling and the motions responsible for conformational exchange. For example, if conformational exchange can be described as a 2-site exchange with rates that are very slow compared to the rate of overall tumbling, then two independent populations of conformers exist, each representing one of the sites in the two-site motion. The contributions to the solution relaxation parameters from each of these conformers can be calculated by weighting the correlation functions by the population fraction of the i^{th} conformational state (w_i):

$$\langle C(t) \rangle_{\text{internal}} = \frac{2}{5} \sum_{i=1}^N w_i \sum_{l=1}^5 e^{-t/\tau_l^i} \sum_{j=1}^{N_{\text{sites}}} \sum_{k=1}^{N_{\text{sites}}} a_l^i \left(\vec{A}^0(t, j), \vec{A}^0(0, k) \right) \times P(\vec{A}^0(t, j) | \vec{A}^0(0, k)) P_0(\vec{A}^0(0, k)) \quad (8)$$

where the additional indices in a_l^i and τ_l^i reflect the fact that diffusion tensor elements differ between distinct structural conformers. Fourier transformation of equation 8 results in the slow exchange expressions for solution longitudinal relaxation rate $1/T_1$, transverse relaxation rate $1/T_2$ and nuclear Overhauser relaxation, averaged over the distribution of N slow exchange conformers:

$$\begin{aligned} \left\langle \frac{1}{T_{1,2}} \right\rangle &= \sum_{i=1}^N w_i \left(\frac{1}{T_{1,2}^i} \right) \\ \langle NOE \rangle &= 1 + \langle T_1 \rangle \left\{ \sum_{i=1}^N w_i \frac{(NOE^i - 1)}{T_1^i} \right\} \end{aligned} \quad (9)$$

where the angular brackets in equation 9 indicate averaging over structural conformers. In equation 9, the following well known expressions are used in the calculation of the relaxation times and NOEs for individual rigidly-tumbling conformers:⁸⁶

$$\frac{1}{T_1} = R_1 = \frac{d^2}{4} \left[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C) \right] + c^2 J(\omega_C) \quad (10)$$

$$\frac{1}{T_2} = R_2 = \frac{d^2}{8} \left[\frac{4J(0) + J(\omega_H + \omega_C) + 3J(\omega_C)}{+6J(\omega_H) + 6J(\omega_H + \omega_C)} \right] + \frac{c^2}{6} [3J(\omega_C) + 4J(0)] \quad (11)$$

$$NOE = 1 + \frac{d^2}{4} \left(\frac{\gamma_H}{\gamma_C} \right) [6J(\omega_H + \omega_C) - J(\omega_C - \omega_H)] T_1 \quad (12)$$

$$d = \left[\frac{\mu_0 h \gamma_H \gamma_C}{8\pi^2 r_{CH}^3} \right], c = \left(\frac{\omega_C}{\sqrt{3}} \right) (\Delta) \quad (13)$$

In equations 10-13, the spectral density is the cosine Fourier transform of $C(t)$, $J(\omega) = \int_0^\infty C(t) \cos \omega t dt$,

ω_H and ω_C are the Larmor frequencies of 1H and ^{13}C respectively, μ_0 is the permeability of a vacuum, γ_H

and γ_C are the magnetogyric ratios of ^1H and ^{13}C , h is Planck's constant ($= 6.626 \times 10^{-34}$ J.sec), r_{CH} is the length of the C-H bond, and Δ is the chemical shift anisotropy (CSA). The effect of CSA tensor asymmetry has been neglected in these expressions. The errors associated with this assumption are discussed in section 4. For the $^{13}\text{C6-H6}$ bond length we used a value of 1.1 Å; the CSA of $^{13}\text{C6}$ is set at $\Delta = 212$ ppm. The choice of CSA was based on the magnitude of the CSA tensor calculated by Ying et al for a Uracil-C6 atom.⁶² While the value used here is the mean value given in that reference (CSA = 212 ± 4 ppm), Ying et al. suggest using the lower end of the error range, i.e., 208 ppm. The difference in the results caused by assuming this lower value will be discussed in section 4.

The choice of r_{CH} is of greater significance due to the r^{-6} dependence of the expressions for the relaxation times. Ying et al.⁶² suggest a value of 1.104 Å for r_{CH} based on computational and RDC analysis. A neutron diffraction study of 1-Methylthymine by Frey et al.⁸⁷ yielded a value for the C6-H6 bond length of 1.096 Å. The average of these two values gives our chosen value of 1.1 Å. However, it is important to note the different values used by other authors: Duchardt and Schwalbe⁸⁸ used a value of 1.08 Å (not averaged over zero-point motion) in their analysis, while a survey of aromatic carbon-to-hydrogen bond lengths derived from neutron diffraction experiments gave a mean value of 1.083 Å.⁸⁹ We briefly consider the effect of making a different choice of r_{CH} in section 4. The selection of the value of 1.1 Å over a lower value can be justified by the fact that the value of 1.083 Å in Reference 33 was derived as an average over several aromatic compounds and did not specifically address nucleic acid bases, while both Ying et al.⁶² and Frey et al.⁸⁷ target nucleic acids, with more specific data on thymine (similar to uracil) in the case of Frey et al. Moreover, the value used by Duchardt and Schwalbe⁸⁸ is not averaged over the vibrational motion of the bond, a choice made to ensure that their order parameters did not exceed unity.

3. Methods

a. Rates, Amplitudes and Coordinate Frames for Internal Motions in TAR RNA

To extract ^{13}C relaxation rates measured in solution from the slow exchange equation 9, the identities and weightings of the structural conformers of TAR involved in the exchange process(es) are required. In accordance with the conformational capture hypothesis, exchange occurs between the unbound TAR and a state that is structurally similar to the form of TAR in complexes with Tat-like peptides. We assume that one of the conformers involved in the slow exchange can be modeled as the lowest energy unbound TAR RNA structure reported by Aboul-ela et al.,⁸ hereafter referred to as 1ANR-1 (Figure 1B). On the basis of model free solution relaxation and RDC studies,^{12,14} the other exchange conformers are hypothesized to differ from 1ANR-1 in the mutual orientation of the upper and lower helices. Therefore, starting with the 1ANR-1 structure, other conformational forms of TAR RNA were generated by twisting and bending the distal helices about the bulged loop.

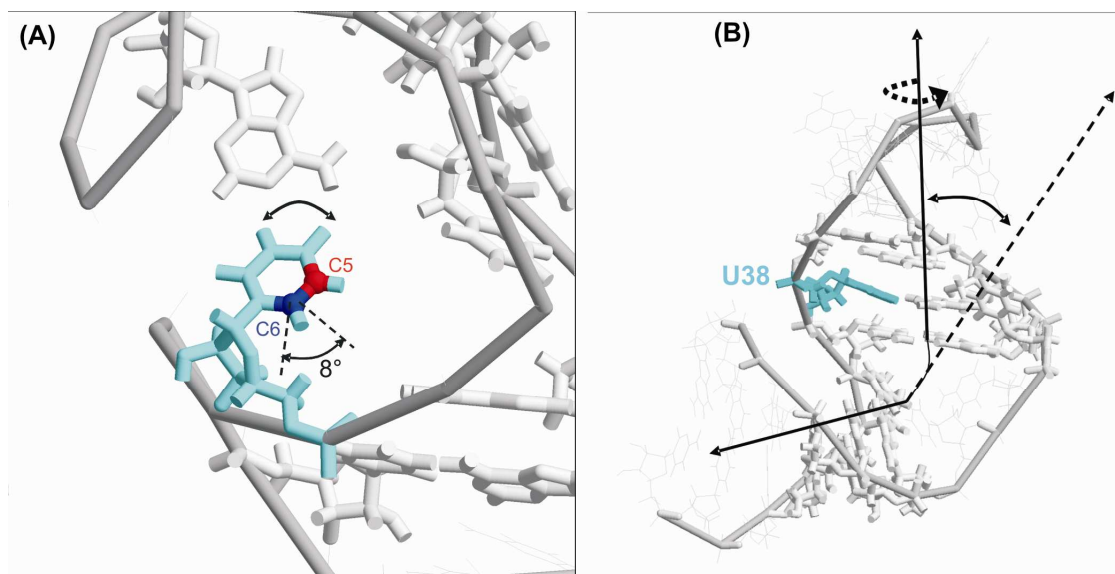


Figure III.2. A) Base-libration of U38 (light blue) relative to its pair-based nucleotide A27; the C5 (red) and C6 (dark blue) atoms are shown. The total amplitude of the best-fit base-libration model was found to be $\pm 4^\circ$, corresponding to an 8° total angular excursion. B) Conformational exchange carries U38 (blue) within the entire upper helix between two distinct conformational states. Shown here is the transformation from the C frame to the M frame. The z axis of the C frame is indicated by the dotted line, and it is approximately perpendicular to the plane of the U38 base in 1ANR. The localized libration of the U38 base shown in Figure 2A is generated by a $\pm 4^\circ$ rotation about the z_C axis. A bend of the entire upper helix is rendered by a rotation about y_M and a twist and is affected by a rotation about z_M . Both rotations are indicated in this figure.

Further information on the rate of conformational exchange is provided by a recent solid-state deuterium NMR (ssNMR) study of selectively deuterated TAR, that monitored the solid-state exchange dynamics of uridine-5,6- $^2\text{H}_2$ incorporated at position U38 in the upper helix (Figure 1A) and at positions U25 and U23 in bulged loop, respectively.⁷⁵ Of the three sites for which data were collected, only the dynamics of U38 is amenable to a slow exchange theoretical analysis. The repositioning of U23 and U25 observed by solid-state NMR are almost coincident with the time scale of overall molecular rotations, violating the separation of motional time scales used to derive equations 8 and 9.

The localized dynamics of U38 in the upper helix at a hydration of 16 waters per nucleotide include small-amplitude local motions ($\pm 4^\circ$) of the base at a rate of $2.2 \times 10^8 \text{ s}^{-1}$, superimposed upon a combination of 15° twisting and 9° bending of the entire nucleotide

occurring at a much lower rate of $1.4 \times 10^6 \text{ s}^{-1}$ (Figure 2). Because of its slow rate and the motional restrictions imposed upon U38 by its position within the upper helix, the combined twisting and bending motion is interpreted as a collective movement of the entire upper helix and likely corresponds to the domain motions observed in RDC studies.^{14,81} Since solid-state relaxation rates and the ^2H NMR line shape for U38 do not change markedly at hydration levels of at least 30 waters per nucleotide,⁸⁵ we assume that the rate of upper helix reorientation remains longer than the time scale of overall rotational motion in solution as well. Thus, the rates of collective dynamics of U38 and of the upper helix of TAR are separated in time scale from the rate of overall molecular rotation and are therefore amenable to analysis using equations 8 and 9. An order of magnitude estimate substantiating the claim of rate separation is provided in the following section (section 3B). Previously, Zhang, Al-Hashimi and co-workers have artificially produced such a separation in conformational and diffusion time-scales by elongating alternately the two TAR-RNA helices.^{12-14,77}

To implement equation 8, relative orientations of the principal axis frame of the $\text{C6-}^1\text{H6}$ dipolar tensor, the dynamic axis system C for local libration of the U38 base, and the axis frame M for the collective motion of the upper helix, were all inferred from the ^2H ssNMR line shape and relaxation studies.⁷⁵ The principal axis systems for the $\text{C6-}^1\text{H}$ dipolar and the $^2\text{H6}$ EFG tensor are axially symmetric about the $\text{C6-}^1\text{H6}$ bond axis. The dynamic axis system C used to produce the small amplitude fast libration shown in Figure 2A has its z_C axis normal to the plane of the base, while the x_C axis bisects the two $\text{C6-}^1\text{H6}$ bond orientations and the y_C axis completes a right-handed coordinate system. The local motion of the U38 base is then treated as a two site exchange about the z_C axis, making the solid angle $\Omega_{PC}(t) = (0, \beta_{PC}, \gamma_{PC}(t))$ time dependent. Based on the analysis of the solid-state data, the difference between γ_{PC} in each of the two

exchanging sites is 8 degrees and the rate of exchange is $2.2 \times 10^8 \text{ s}^{-1}$ and β_{PC} is set to 90° in the above coordinate frame definition.⁷⁵

For the collective motions of the upper helix, the M frame coincides with the crystal frame of one of the sites. To bring the crystal frame of the second site into coincidence with the M frame, two transformations are required. A bend of the entire upper helix is rendered by a rotation about y_{M} and a twist is effected by a rotation about the z' axis that results from the bend transformation, which together carry the upper helix into its orientation within the second conformer. The direction of y_{M} determines the sense of the upper helix rotation relative to the bulged loop. Both rotations are depicted in Figure 2B.

b. Rotational Diffusion Tensors

To complete the definition of dynamic axes relevant to the TAR RNA relaxation problem, we need to establish the orientation of the principal axis frame of the rotational diffusion tensor (PASd) relative to the M, C, and C6-H6 bond axis frames. This in turn requires a determination of the eigenvalues and eigenvectors for the diffusion tensors of each of the two slowly exchanging conformers, which are characterized by different orientations of the upper helix relative to the lower helix. Equations 14 – 19 provide an overview of the standard method of calculating the diffusion tensors using a bead analysis. The well-known methodology is included in this paper to provide the reader with a clear picture of where the atomic element radius (AER) occurs in the calculation. The AER is defined below and is considered at some length in the interpretation of subsequent results.

Eigenvalues and eigenvectors for the rotational diffusion tensors of TAR are derived formally from a generalized Einstein relation⁹⁰

$$D = \begin{pmatrix} D_{tt} & D_{tr}^\dagger \\ D_{tr} & D_{rr} \end{pmatrix} = k_B T \begin{pmatrix} \Xi_{tt} & \Xi_{tr}^\dagger \\ \Xi_{tr} & \Xi_{rr} \end{pmatrix}^{-1} \quad (14)$$

where D_{tr} and D_{tt} are 3×3 rotational and translational diffusion tensors, respectively. The off diagonal terms D_{tr} express the influence on diffusion of translational-rotational coupling.

Similarly, Ξ_{tt} and Ξ_{rr} are the 3×3 translational and rotational friction tensors, respectively, and the matrix Ξ_{tr} represents the friction due to translational-rotational coupling. The rotational friction tensor is derived from the relation:

$$\Xi_{rr} = \sum_{i,j} U_i \cdot C_{ij} \cdot U_j \quad (15)$$

where U is a matrix with elements composed of non-hydrogenic atomic coordinates:

$$U_i = \begin{pmatrix} 0 & -z_i & y_i \\ z_i & 0 & -x_i \\ -y_i & x_i & 0 \end{pmatrix} \quad (16)$$

and $C=B^{-1}$ is a $3N \times 3N$ supermatrix (for a molecule composed of N discrete elements) composed of 3×3 blocks derived as follows

$$B_{ij} = \begin{cases} T_{ij} & \text{for } i \neq j \\ (1/\zeta_i)I & \text{for } i = j \end{cases} \quad (17)$$

where $\zeta_i = 6\pi\eta_0\sigma_i$ is the Stokes law friction, η_0 is the solvent viscosity and T_{ij} is the ij element of the hydrodynamic interaction tensor between two beads:^{90,91}

$$T_{ij} = (8\pi\eta_0 R_{ij})^{-1} \left(I + \frac{\vec{R}_{ij} \vec{R}_{ij}}{R_{ij}^2} + \frac{\sigma_i^2 + \sigma_j^2}{R_{ij}^2} \left(\frac{I}{3} - \frac{\vec{R}_{ij} \vec{R}_{ij}}{R_{ij}^2} \right) \right) \quad (18)$$

Equation 18 expresses the fact that elements ij of the Oseen tensor are obtained by assuming non-hydrogenic atoms i and j can be treated as spherical beads with radii σ_i and σ_j , the centers of which are separated by a displacement vector $\vec{R}_{ij}$ and a distance R_{ij} . I represents the 3×3

identity matrix. The atomic radii must however account for covalently attached protons and hydration. For $i = j$, $B_{ij} = (1/\zeta_i)I$. For the case in which the spheres are overlapping ($R_{ij} < \sigma_i + \sigma_j$), equation 18 becomes

$$T_{ij} = (6\pi\eta_0 R_{ij})^{-1} \left(\left(1 - \frac{9}{32} \frac{R_{ij}}{\sigma} \right) I + \frac{3}{32} \frac{\vec{R}_{ij} \vec{R}_{ij}}{\sigma R_{ij}} \right) \quad (19)$$

where the simplification $\sigma_i = \sigma_j = \sigma$ has been made.

Several numerical algorithms based on modified Oseen tensors and bead models have been developed to obtain eigenvalues and eigenvectors for diffusion tensors.^{90,91} Here we use the public domain version of HYDRONMR.⁹² An important simplification in calculating the diffusion tensors with HYDRONMR is the assumption that a single value for σ , also called the Atomic Element Radius (AER), can be applied to all non-hydrogenic atoms in the molecule, as in equation 19. Although this is a significant simplification, it is questionable to use multiple independent atomic radii given a lack of detailed information on hydration in RNA on an atom-by-atom basis. The choices of AER used in this study are evaluated and described in Section 4 below.

As will be shown below in Figures 4A and 4B, the rotational diffusion eigenvalues for TAR RNA are on the order of $3 \times 10^8 \text{ rad}^2 \text{ s}^{-1}$. An order of magnitude estimate for the tumbling rate can be obtained by assuming spherical symmetry: rate $\sim 6D_0 \sim 10^9 \text{ s}^{-1}$, where D_0 is the diffusion coefficient for a sphere. This is almost three orders of magnitude higher than the exchange rate, justifying the use of a slow exchange methodology for motion of the U38 residue. Moreover, the effects of time-scale overlap start to become significant for a CE rate greater than about $1 \times 10^7 \text{ sec}^{-1}$, as established using a more general theory that does not require separation of time scales.

4. Results

Calculation of Relaxation Times of TAR RNA from the Slow Conformational Exchange Model

To apply the slow conformational exchange relaxation model to the calculation of relaxation times we assume that $^{13}\text{C6}$ relaxation is dominated by the fluctuations of its dipolar coupling to H6 and of its own CSA. For the $^{13}\text{C6}$ -H6 bond length we used a value of 1.1 Å; the CSA of $^{13}\text{C6}$ is set at $\Delta = 212$ ppm. The asymmetry of the CSA as well as the non-collinearity of the principal CSA axis and the dipole interaction axis were neglected in the relaxation equations. An estimate for the percentage error introduced by these approximations for pyrimidine C6 atoms is about 6-7%, for a CSA of 212 ppm.⁶⁵ The error bars shown in all figures do not include this systematic error.

The values of the CSA and r_{CH} chosen for this study were chosen from references discussed in section 2. To consider the effect of changing these parameters, we lowered the CSA down to 208 ppm, as recommended by Ying et al.⁶² The resulting T_1 times increased by ~5 ms and the T_2 times increased by ~0.4 ms. The magnitudes of these changes were considered to be sufficiently close to the error bars to be negligible.

An independent variation of r_{CH} down to 1.096 Å⁸⁷ or up to 1.104 Å⁶² resulted in a decrease or increase, respectively for the two changes, of T_1 by ~5 ms and T_2 by ~0.4 ms, again within experimental error. Lowering r_{CH} to 1.08 Å,⁸⁹ on the other hand, caused T_1 to decrease by ~25 ms, and T_2 by ~1.7 ms. These variations are significant enough to alter the best-fit parameters considered below. However, for reasons mentioned in section 2 above, we chose a higher value of 1.1 Å for all of the following analysis.

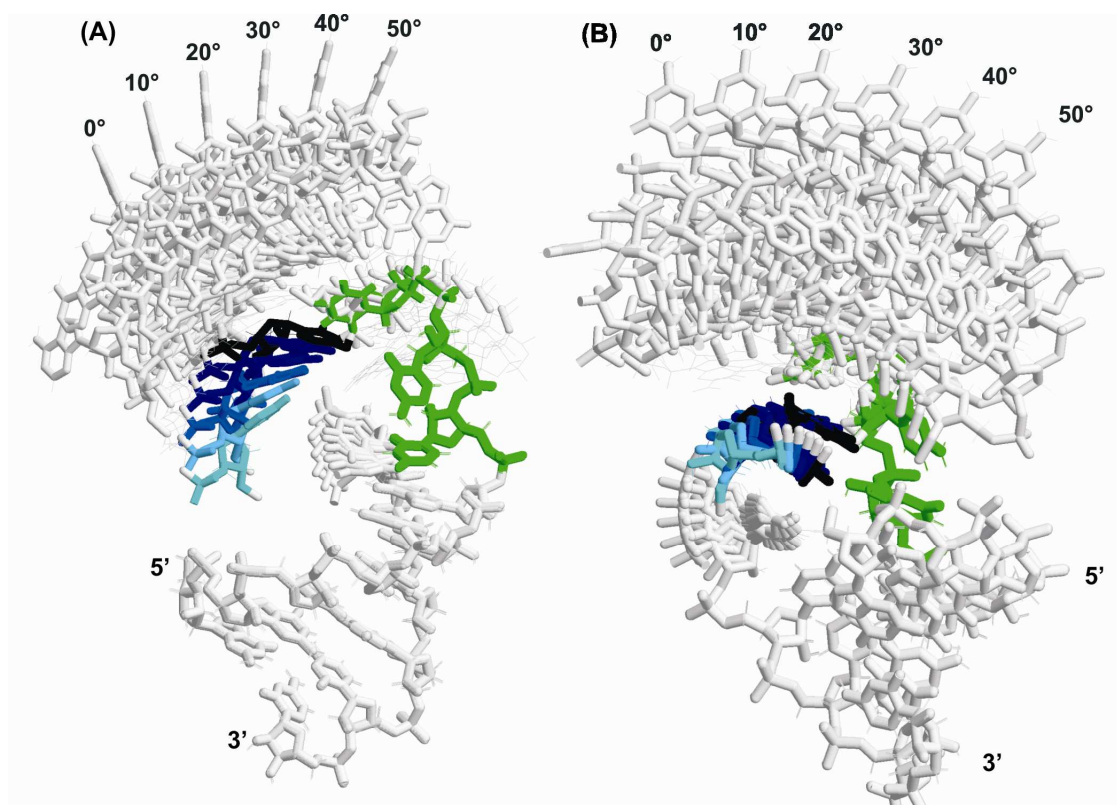


Figure III.3. Representation of the family of TAR structures differing in the rotation of the upper helix of TAR RNA relative to the lower helix starting from 1ANR-1 and systematically increasing the bend angle. The conformational exchange (CE) angle was varied in 10° increments from 0° to 50°. Both sets of structures are oriented in the figures such that the axis of helix reorientation is approximately out of the plane of the page. A) The “90 degree series”; and B) the “30 degree series”; conformational exchange in the two sets of structures is represented as a rotation about axes that are 60° apart in a plane perpendicular to the lower helix (The coloring of the U38 residue is varied from light blue to black for clarity in going from a CE angle of 0° to 50°. The bulge residues are highlighted in green in both panels.)

As discussed in section 3, one of the slowly exchanging conformers is very reasonably assumed to be represented by 1ANR-1, the lowest energy NMR structure reported for unbound TAR. To obtain the structure of the second TAR conformer, we systematically probed the conformational space sampled by the upper helix to test the possibility that, in solution, the angular displacement of the upper helix may differ from the bend angle of $\sim 10^\circ$ observed by ssNMR (the twist angle was assumed to have a negligible effect on the hydrodynamic properties, when considering an idealized upper helix with near-cylindrical symmetry). Specifically, the upper helix was reoriented relative to a fixed lower helix, starting with 1ANR-1 and

systematically increasing the inter-helical angle in 10° increments up to 50°, which is the difference between the interhelical angles of free TAR and TAR bound to Tat (Figure 3). The change in the inter-helical bend angle relative to 1ANR-1 is referred to here as the conformation exchange (CE) rotation angle. For each structural model associated with a specific CE rotation angle, we evaluated the diffusion tensor using HYDRONMR and incorporated the localized base-libration motion (amplitude of $\pm 4^\circ$, rate of $2.2 \times 10^8 \text{ s}^{-1}$) into the relaxation time calculations according to equation 8. The relaxation times for each structure associated with a CE rotation angle were subsequently averaged with those of 1ANR-1 using the slow exchange formalism of equation 9, assuming an equal population of conformers, as observed in solid-state NMR studies.⁷⁵

Several sets of rotated structures were generated with the y_M reorientation axis chosen perpendicular to the lower helix. To study how changing the direction of helical bending would change hydrodynamics and relaxation, the orientation of the y_M axis (within the plane perpendicular to the lower helix) was varied through 30° increments. The arbitrarily chosen starting orientation was designated as the “0 degree series,” with subsequent increments labeled similarly as the “30 degree series,” “60 degree series,” etc. A superposition of atomic-detail TAR RNA structures of 1ANR-1 with CE rotation angles up to 50°, are shown for the “90 degree series” and “30 degree series” in Figures 3A and 3B, respectively. The following discussion will focus mainly on these two sets of structures for the sake of brevity (except for an inclusion of the “150 degree series” in Figures 7A and 7B). In practice, these structures were generated by shearing one of the bonds, and rotating the upper helix about the designated y_M axis, keeping the lower helix and trinucleotide bulge positions fixed. It is important to note that the procedure of artificially modifying the 1ANR-1 TAR structure by rotating the upper helix was done only to

approximate the hydrodynamic effect of rotating one helix relative to the other. We justify this approximation by observing that the hydrodynamic beads used in the diffusion tensor calculations overlapped sufficiently to wash out details of atomic-level bonding. Moreover, the U38 residue is rotated along with the upper helix and is thus not affected by the shearing and reattachment of the bonds. We have attempted to avoid steric clashes as much as possible, but for higher CE angles of 50° and above there are minimal instances of overlapping bonds and atoms (results for CE angles > 50°, while available were not included in the following analysis for this reason). These overlaps were not deemed to be significant enough to affect the sets of results shown in this manuscript.

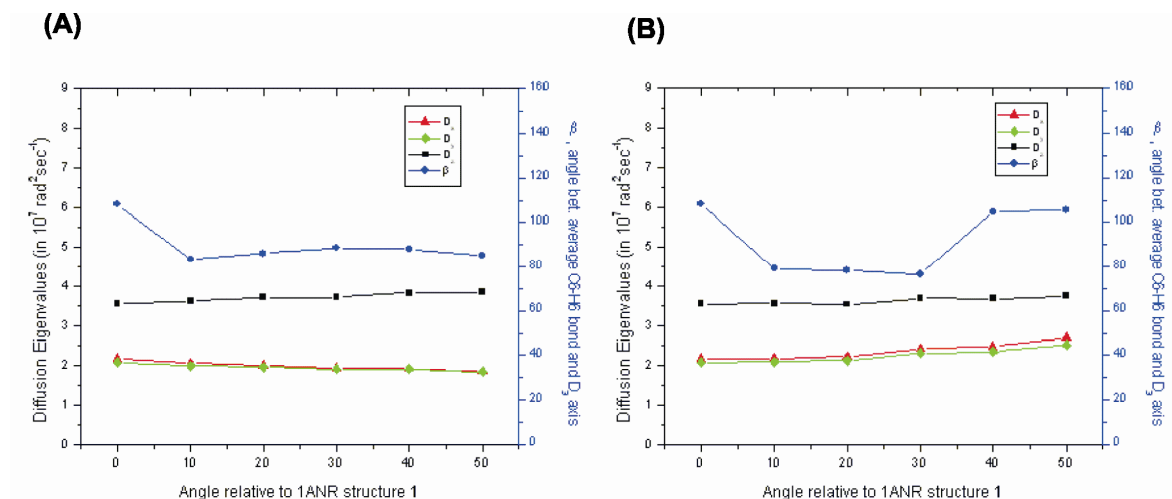


Figure III.4. Eigenvalues of the rotational diffusion tensor and the angle β between the $^{13}\text{C6-}^2\text{H6}$ bond axis of U38 and the largest eigenvalue axis of the principal axis frame of the diffusion tensor (PASd) plotted as a function of rotation angle for the two sets of structures generated by bending the TAR structures through two independent directions. The calculations use an AER of 2.3 Å. A) corresponds to the structures of Figure 3A, while B) corresponds to the structures of Figure 3B.

Figures 4A and 4B display the eigenvalues of the diffusion tensor as functions of the CE rotation angle for the two choices of helical bending axes of Figures 3A and 3B, respectively. Also shown is the orientation of the $^{13}\text{C6-}^1\text{H6}$ bond vector relative to the principal axis frame of

the diffusion tensor (PASd), expressed as the angle β between the C6-H6 bond and the axis of PASd associated with the largest eigenvalue D_3 . The magnitudes of the diffusion eigenvalues are high enough to amply support the case for using a slow exchange theory (see estimate of tumbling rate in section 3B).

The model parameters that were adjusted independently of the experimental results include primarily the bead radius or atomic element radius (AER) and the sense of the direction of the upper helix rotation relative to the lower helix of TAR. Based on the solution conditions,¹¹ the solvent viscosity in HYDRONMR was chosen to be the viscosity of 99.9% D₂O at 25° C, $\eta_0 = 0.01096$ Poises.⁹³ Simulated T_1 and T_2 relaxation times were compared to solution NMR experimental relaxation times reported for the ¹³C6 spin of U38 as: $T_1 = 354 \pm 3.2$ ms, $T_2 = 24.6 \pm 0.5$ ms, and a heteronuclear NOE of 1.13 ± 0.01 .¹¹

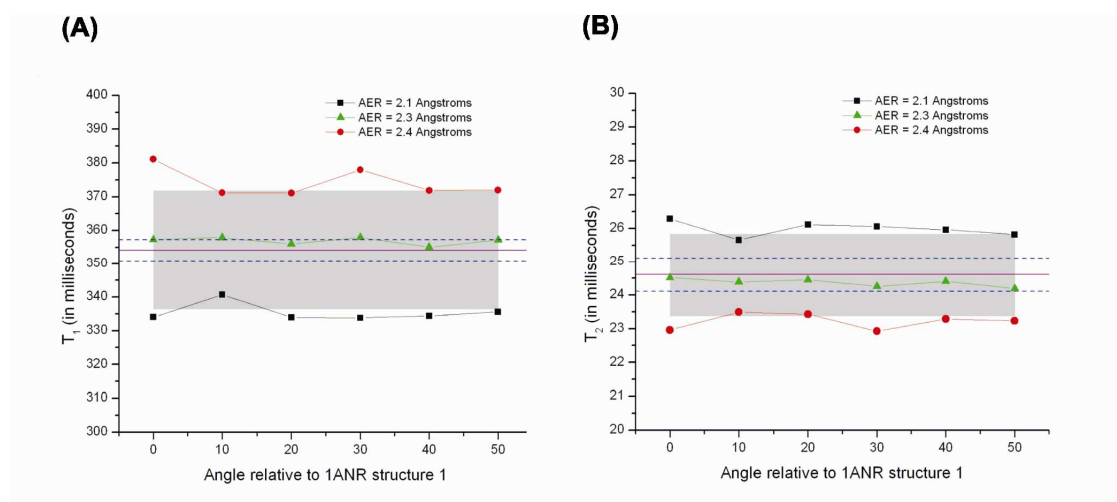


Figure III.5. A) T_1 and B) T_2 values derived using equation 25 for the “90 degree series” (Figure 3A) as a function of AER and CE rotation angle. The solid purple line and dashed blue lines represent the experimental values and error bars respectively of A) $T_1 = 354 \pm 3.2$ ms, and B) $T_2 = 24.6 \pm 0.5$ ms. The shaded regions in both panels represent error regions of $\pm 5\%$.

a. AER dependence

Previous studies into the hydration radii of atoms in nucleic acids have yielded a range of results. In a study of several short oligonucleotides using HYDROPRO, a bead-modeling hydrodynamics program, Fernandes et al.⁹⁴ showed that fits to hydrodynamic properties required different AER values for different molecules ranging from 1.6 to 3.8 Å. The authors assumed an average van der Waals radius of 1.7 Å for non-hydrogenic atoms. Aragon et al.⁹⁵ found good agreement with optical Kerr effect rotational correlation times of tRNA^{phe} using solvent-accessible surface-based hydrodynamic simulations with the BEST⁹⁶ program, that included uniform hydration 1.1 Å in thickness. When added to the van der Waals radius of about 1.7 Å above, this yields an atomic radius of 2.8 Å.

In view of this diversity of values for nucleic acids, we chose to vary the AER in order to get the best-fit to the solution data and the resulting range of values is from 2.1 to 2.4 Å (which lies within the range quoted by Fernandes et al.⁹⁴). In Figures 5A and 5B, values of T_1 and T_2 are shown for three AER values of 2.1, 2.3 and 2.4 Å for the “90 degree series” (Figure 3A; AER = 2.2 Å was left out for clarity). In Figures 5A and 5B the experimental T_1 and T_2 values for ¹³C6 of 354 ms and 24.6 ms, respectively, are indicated by the solid, purple line, with dotted lines indicating the experimentally-determined uncertainties. Based on these calculations we see T_1 values increase and T_2 values decrease with increasing AER due to a corresponding increase in the tumbling times. However, for a given AER there is no discernible trend for this series as a function of CE angle, and relaxation times for all CE angles $\leq 50^\circ$ fit the experimental data. A visual inspection of the structures as well as the trend in diffusion tensor eigenvalues in Figure 4A show that the rotational diffusion tensors remain fairly cylindrical throughout the reorientation process, and the β angle does not change significantly. Thus, no drastic changes in

the rotational diffusion of the C-H bond are to be expected and the relaxation times do not vary.

In the physical sample, however, as the upper helix begins to line up with the lower helix, the bulge would have to reconfigure itself and would possibly lead to a discernible asymmetry in the diffusion tensor.

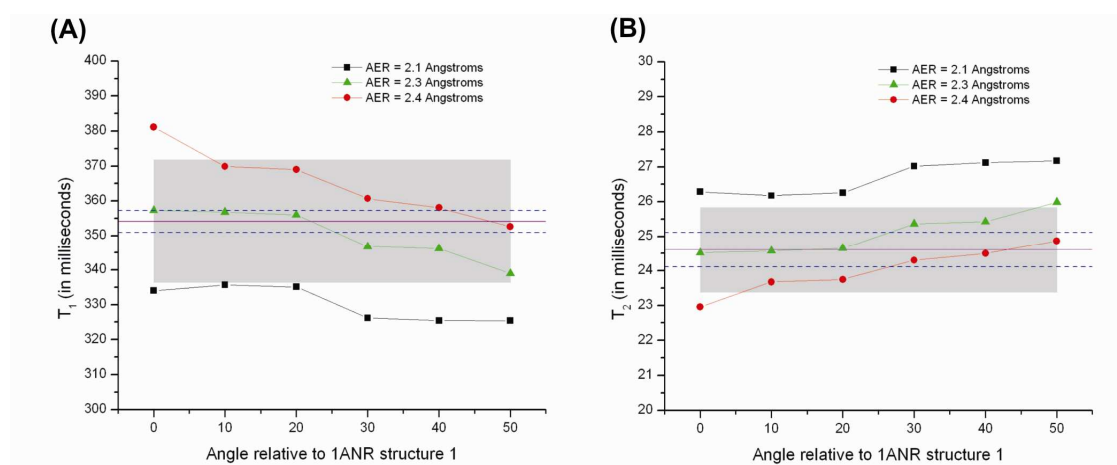


Figure III.6. A) T_1 and B) T_2 values derived using equation 25 for the “30 degree series” (Figure 3B) as a function of AER and CE rotation angle. The solid purple line and dashed blue lines represent the experimental values and error bars respectively of A) $T_1 = 354 \pm 3.2$ ms, and B) $T_2 = 24.6 \pm 0.5$ ms. The shaded regions in both panels represent error regions of $\pm 5\%$.

We repeated the same calculations for the “30 degree series” (Figure 3B). The results for T_1 and T_2 are shown in Figures 6A and 6B, respectively. The relaxation times for this series show more variation as a function of CE angle. This increased variation is due not only to changes in the principal elements of the rotational diffusion tensor, but also because of changes in the angle β which orients the C6-H6 bond relative to the PASd frame. Best fits to experimental T_1 and T_2 values occur for AER=2.3 Å and for smaller bend angles of $< 20^\circ - 30^\circ$.

There is a correlation between the AER and the CE angle where an increase in the AER allows higher bend angles to be fit. For example, in addition to the fit parameters above, an AER of 2.4 Å together with CE angles between 30° and 40° fit the experimental data fairly well.

Continuously increasing the AER from 2.3 Å to 2.4 Å and more will result in increasingly larger CE angles fitting the experimental data. It is possible to explain the correlation between the AER and CE angles in terms of the effect of increasing the AER upon the hydrodynamic models: if a model is composed of larger spheres, the anisotropies are partially “washed out.” Therefore, by scaling up the AER, increasingly bent molecules with large CE angles can start to approximate the experimental data better (while at the same time lower CE angle structures will hydrodynamically approach the ideal spherical case). In general, however, we believe that the results for smaller CE angles are more reliable due to the fact that the bulge configuration, frozen in the 1ANR-1 state, would be less representative of the physical structure of the bulge as the upper helix deviated further and further from the 1ANR-1 orientation.

Due to the correlation between the AER and the CE angle, to isolate best-fit CE angles an independent assessment of the AER is needed. This can be achieved by considering more explicit hydration models for nucleotides, attempts at which have been described at the beginning of this sub-section. An establishment of the AER would then allow the selection of best-fit CE angles. Conversely, if we consider the CE angles to be constrained by ssNMR models, it is possible to establish the AER associated with a given molecule: the ssNMR best-fit value for the upper helix bend of 9° supports an AER of 2.3 Å. Again, if we allow the angle to vary under solution conditions, it is possible to extract slightly different AERs.

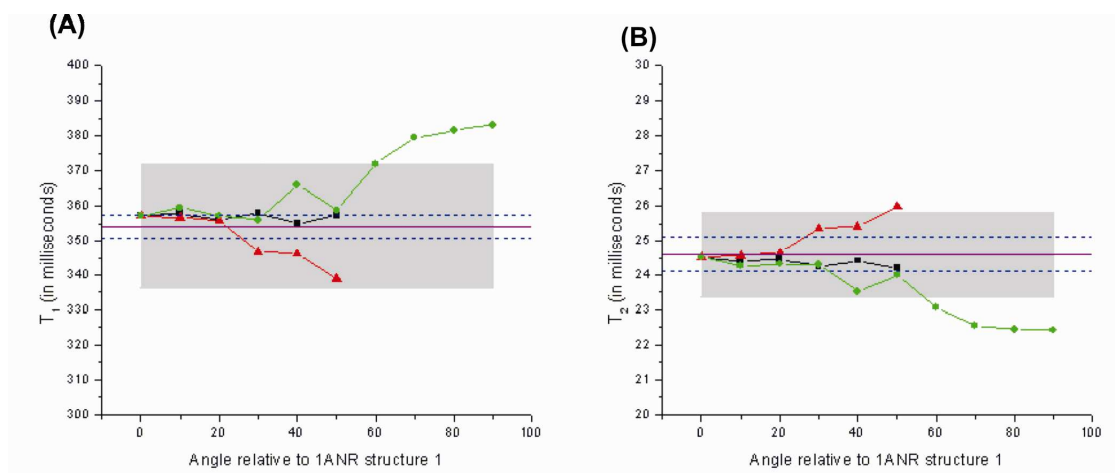


Figure III.7. A) T_1 and B) T_2 relaxation times calculated for the $^{13}\text{C6}$ spin of U38 with an AER of 2.3 Å for the TAR RNA structures designated as the “90 degree series” (Figure 3A; black), the “30 degree series” (Figure 3B; red), and for the “150 degree series” (green; the “150 degree series” has been extended to a CE angle of 90 degrees as there were no steric clashes even for high CE angles). The solid purple line and dashed blue lines represent the experimental values and error bars respectively of A) $T_1 = 354 \pm 3.2$ ms, and B) $T_2 = 24.6 \pm 0.5$ ms. The shaded regions in both panels represent error regions of $\pm 5\%$.

b. Dependence on direction of helical bending

Figure 7A shows calculated T_1 values as a function of CE rotation angle for the set of structures of Figures 3A and 3B, as well as the “150 degree series” with an AER of 2.3 Å. The results for the “150 degree series” were extended to a CE angle of 90 degrees as there were no steric clashes in this series, even at high CE angles. For the structures of Figure 3A, the calculated T_1 relaxation time is almost invariant as a function of CE angle and lies partly within experimental uncertainty. For the set of structures of Figure 3B, only CE angles $< 30^\circ$ produce simulated T_1 values that are within experimental error of the measured value. The data for the “150 degree series” shows some variation in the range of CE angles between 30° and 50° degrees, and then shows a clear deviation away from the experimental data for angles $> 50^\circ$. This series is similar to the structures in Figure 3A (the “90 degree series”) in that the upper helix “straightens out” relative to the lower helix (not shown) and is thus a good representation of the solid-state models which predict such a helical realignment in the transition from free to bound

TAR. The “150 degree series” does however differ from the series in Figure 3A in the position of the bulge relative to the upper helix during the reorientation. From Figure 7A one can therefore conclude that the CE angles that match the data are sensitive to the direction of reorientation of the helix, as evidenced by the graphs for both the structures in Figure 3B and the “150 degree series.” Similar considerations apply to the results for the T_2 calculation shown in Figure 7B.

If we were to select a range of CE angles based on ssNMR simulations, it would be possible to select out a preferred helical reorientation direction. Conversely, an independent determination of the direction of reorientation, perhaps based on solid-state models and/or steric considerations might allow the CE angles to be determined from the above data.

In the above analyses the NOE values were not considered as the values varied negligibly over the entire range of parameter values.

5. Discussion

We have developed a slow exchange model to describe the solution behavior of TAR RNA in the presence of internal motions that alter the shape of the RNA and therefore its rotational diffusion. Our theory of solution NMR relaxation accounts for the non-rigid rotational motion of a molecule of arbitrary structure possessing a rotational diffusion tensor with arbitrary eigenvalues. The application of the model requires a significant separation in time scales between internal and global motion, a condition that is satisfied in the present system. The slow exchange correlation function given in equation 8, when parameterized exactly according to models of the internal dynamics of TAR RNA obtained by prior solid-state NMR studies, simulates to within experimental error the T_1 and T_2 data for a ^{13}C spin in U38 within the upper helix of TAR RNA, the same nucleotide for which the solid-state models were developed. The results of the slow exchange model applied to TAR in section 4, together with prior ^2H ssNMR

and solution relaxation/RDC studies, indicate that conformational capture is a feasible paradigm for describing recognition between TAR RNA and Tat-like peptides.

By its very nature, a theory of relaxation that involves non-rigid rotation will have more parameters than a theory of rigid rotation, even when the diffusion equation is simplified by treating motion as exchange between discrete conformers with distinct rotation tensors, as was done in equation 7 and in the theory of Wong et al.⁸² Thus, a serious concern is that relaxation and line shape data are better simulated by assuming extra motions and additional fitting parameters. However, the motions that were introduced into the present model of TAR dynamics are supported by other data including solid-state relaxation times and line shapes, as well as RDC data, all of which indicate the presence of collective motions of the helices of TAR RNA about a putative flexible bulged loop. It can be also seen that even the single conformer 1ANR-1 (the CE angle = 0° data point in all graphs) fits the solution experimental data well, without recourse to the conformational exchange formalism. This is to be expected if the change in the hydrodynamic behavior of the molecule is not large in the transition to the proposed exchange partner. Also, we do not believe that a single conformer can accurately represent the physical condition of the molecule in solution conditions given that the ssNMR results support the presence of such a motion, and it is unlikely that there would be a reduction in the types of motion available to the molecule in going from solid-state to solution sample conditions.

It is certainly possible that the values of the slow exchange model parameters derived from solid-state NMR might not accurately portray the amplitudes and rates of these motions in solution. However, solid-state relaxation rates for the relaxation of the base deuterons of U38, U25, and U23 in TAR RNA change very slowly at high levels of hydration,⁸⁵ indicating that internal motions converge to stable amplitudes and rates long before the bulk solution state is

reached. In order to further address this issue, we probed a wide range of models of conformational exchange generated by systematically increasing the conformational exchange angle from 0° (starting structure) to a distorted structure with a rotation angle of 50° (the maximum CE angle can be raised to any arbitrary value as long as physically reasonable conformers are available). The results of section 5 show that solution experiments are in agreement with different CE angles depending on the AER and the direction of helical reorientation.

With regard to the rates of conformational exchange, it was assumed the time scale separation between conformational exchange and overall rotational motion exists in solution, but the solid-state value of $1.4 \times 10^6 \text{ s}^{-1}$ was not explicitly used. To test the assumption of slow conformational exchange, a more general solution would require solving equation 7 for arbitrarily structured molecules without a time scale separation. Such a theory would allow the analysis of the relaxation of spins of the U23 and U25 nucleotides as well. We have recently obtained such a solution and will report the ensuing results in the near future.

In order to compare our results to previously published results by Zhang, Al-Hashimi and co-workers^{12,14} we examined the order tensor-derived inter-helical bending angles, as well as the ensemble-derived motional parameters in reference 13. The order tensor analysis yields average inter-helical bending angles of 25° (short upper helix relative to elongated lower helix) and 54° (short lower helix relative to elongated upper helix), with the difference being interpreted as the effect of twisting motions affecting the two helices. While solid-state results reported in Olsen et al.¹⁰ disagree with the magnitude of the order tensor-derived bending motions, they do indicate the presence of simultaneous twisting and bending motions, and the results reported in this manuscript do not exclude inter-helical bend angles of 25°, or even 54°, for appropriate choices

of the AER and the direction of helical bending. Furthermore, the three-conformer ensemble study in reference 13 yielded an overall helical bend of 94° along with an upper helical twist of 110° and lower helical twist of 53° . While these angles are much larger than the equivalent solid-state parameters¹⁰, they can potentially be reconciled with the results of this paper, albeit with an AER outside of the range of values considered here ($> 2.4 \text{ \AA}$). Finally, we considered the timescale of domain motions reported in reference 17 to be on the order of 1.5 ns to 1.9 ns. These are several orders of magnitude faster than the slow exchange suggested by the solid-state models of reference 14. Thus, if it is assumed that the solid-state models are a fair representation of the solution motions, then there is a sharp discord between the results presented herein and those of references 13 and 17. It is possible that the solid state matrix imposes some constraints on the motions compared to solution, although a thorough study of dynamics as a function of hydration in the solid state indicates that this is unlikely. Additional experiments and theoretical studies will be required to address the origin of this difference.

Finally, the structures generated above assume a frozen configuration for the bulge. In the physical sample, however, as the upper helix begins to line up with the lower helix, the bulge would have to reconfigure itself possibly leading to a discernible asymmetry in the diffusion tensor. This is a significantly more difficult problem given the number of under-constrained bulge degrees of freedom, and will have to be addressed in the future.

In this manuscript we have modeled only the U38 residue. Our justification in doing so is the relative degree of certainty in the available degrees of motion of this residue which allowed us to demonstrate the slow exchange methodology. Nevertheless, solution NMR relaxation studies show that for all pyrimidine $^{13}\text{C6}$ spins in the upper helix of TAR, T_1 values lie within the narrow range of $354 \pm 14 \text{ ms}$ and T_2 values are similarly distributed narrowly within $25.2 \pm 0.2 \text{ ms}$.

The slow exchange model can therefore account for the relaxation of all the $^{13}\text{C6}$ pyrimidine spins by adjustment of the orientation of the base relative to the frame of the diffusion tensor.

6. Conclusions

We have simulated solution NMR experimental relaxation times with models derived from solid-state NMR studies for the U38 residue in HIV-1 TAR RNA. In addition to incorporating base-libration into single conformers, this also involved using a formalism of conformational exchange between conformers at a rate slower by over two orders of magnitude than tumbling rates. The slow conformational exchange simulation procedure is completely general, and for sufficient rate separation, can easily be applied to a discrete set of conformers. Starting with solid-state models, and incorporating them into molecular tumbling equations, it is possible to address the question of whether the same motions are common to both sample conditions: if the solution results can be simulated through this method, then we have extended the domain of support for such motions to a larger data set; however, even if ssNMR models are at odds with solution NMR experiments, we can iterate the procedure for modified internal motional rates and/or amplitudes, or change the motional models altogether. We have endeavored to provide atomic detail in the combined modeling of solid-state and solution NMR experiments without recourse to hydrodynamic symmetry. Future research will attempt to characterize situations of conformational exchange with rates overlapping tumbling rates, as well as accounting for accompanying modifications of bulge configurations.

Appendix III.A: Sample Mathematica code for the slow exchange method

//Function and parameter definitions common to all structures

//Function definitions based on Equations III.4

```
Fprime[{μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = -(1/3) + (f1^2)*(μ1^2) + (f2^2)*(μ2^2) +
(f3^2)*(μ3^2);
Δ[{D1_, D2_, D3_}] = Sqrt[D1^2 + D2^2 + D3^2 - D1*D2 - D2*D3 - D3*D1];
Dcoeff[{D1_, D2_, D3_}] = (1/3)*(D1 + D2 + D3);
Gprime[{D1_, D2_, D3_}, {μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = (1/Δ[{D1, D2, D3}])*(-
Dcoeff[{D1, D2, D3}] + D1*((f1*μ1)^2 + (μ2*f3)^2 + (μ3*f2)^2) + D2*((f2*μ2)^2 +
(μ3*f1)^2 + (μ1*f3)^2) + D3*((f3*μ3)^2 + (μ1*f2)^2 + (μ2*f1)^2));
```

//Coefficients of the time-dependent exponentials defined in Equation III.4

```
a1[{D1_, D2_, D3_}, {μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = 0.75*(Fprime[{μ1, μ2, μ3}, {f1, f2, f3}]
+ Gprime[{D1, D2, D3}, {μ1, μ2, μ3}, {f1, f2, f3}]);
a2[{μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = 3*(μ2*μ3)*(f2*f3);
a3[{μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = 3*(μ1*μ3)*(f1*f3);
a4[{μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = 3*(μ1*μ2)*(f1*f2);
a5[{D1_, D2_, D3_}, {μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = 0.75*(Fprime[{μ1, μ2, μ3}, {f1, f2, f3}] -
Gprime[{D1, D2, D3}, {μ1, μ2, μ3}, {f1, f2, f3}]);
avec[{D1_, D2_, D3_}, {μ1_, μ2_, μ3_}, {f1_, f2_, f3_}] = {a1[{D1, D2, D3}, {μ1, μ2, μ3}, {f1,
f2, f3}], a2[{μ1, μ2, μ3}, {f1, f2, f3}], a3[{μ1, μ2, μ3}, {f1, f2, f3}], a4[{μ1, μ2, μ3}, {f1, f2,
f3}], a5[{D1, D2, D3}, {μ1, μ2, μ3}, {f1, f2, f3}]};
```

//Characteristic times defined in terms of the diffusion tensor eigenvalues (Equation III.4)

```
tauvec[{D1_, D2_, D3_}] = {(6*Dcoeff[{D1, D2, D3}] - 2*Δ[{D1, D2, D3}])^-1, (
3*(Dcoeff[{D1, D2, D3}] + D1))^-1, ( 3*(Dcoeff[{D1, D2, D3}] + D2))^-1, (
3*(Dcoeff[{D1, D2, D3}] + D3))^-1, (6*Dcoeff[{D1, D2, D3}] + 2*Δ[{D1, D2, D3}])^-1};
```

```
specd = {};
```

//NMR parameters

```
γC = 6.728*10^7; //Gyromagnetic ratio of Carbon-13
```

```
γH = 2.675*10^8; //Gyromagnetic ratio of Hydrogen
```

```
rCH = 1.1*10^-10; //Carbon-hydrogen bond length for an aromatic carbon atom
```

```
δσ = 212*10^-6; //Chemical shift anisotropy (CSA) for a uridine base carbon atom
```

```
rate = 2.15*10^8; //Local base libration rate for U38 in TAR RNA
```

```
d2 = (0.1*10^-14)*((1.05451*10^-34)*γH*γC/rCH^3)^2 //Squared dipolar coupling constant
```

```
c2[ω_] = (2/15)*(δσ^2)*ω^2; //Coefficient of CSA contribution to relaxation
```

```
FT[ω_, time_] = time/(1 + (ω*time)^2); //Fourier transform function
```


//Spectral density function for the case when the site vectors $\{\mu_1, \mu_2, \mu_3\}$ and $\{f_1, f_2, f_3\}$ are the same

$$\text{Jeq}[\omega, \text{rate}, \{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{f_1, f_2, f_3\}] = 0.25 * \text{Sum}[\text{avec}[\{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{f_1, f_2, f_3\}][[k]] * (\text{FT}[\omega, \text{tauvec}[\{D1, D2, D3\}][[k]]] + \text{FT}[\omega, \text{tauvec}[\{D1, D2, D3\}][[k]]) / (1 + 2 * \text{rate} * \text{tauvec}[\{D1, D2, D3\}][[k]])], \{k, 5\}];$$

//Spectral density function for the case when the site vectors $\{\mu_1, \mu_2, \mu_3\}$ and $\{f_1, f_2, f_3\}$ are different

$$\text{Juneq}[\omega, \text{rate}, \{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{f_1, f_2, f_3\}] = 0.25 * \text{Sum}[\text{avec}[\{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{f_1, f_2, f_3\}][[k]] * (\text{FT}[\omega, \text{tauvec}[\{D1, D2, D3\}][[k]]] - \text{FT}[\omega, \text{tauvec}[\{D1, D2, D3\}][[k]]) / (1 + 2 * \text{rate} * \text{tauvec}[\{D1, D2, D3\}][[k]])], \{k, 5\}];$$

//Total spectral density function

$$\text{J0}[\omega, \text{rate}, \{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{f_1, f_2, f_3\}] = \text{Jeq}[\omega, \text{rate}, \{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{\mu_1, \mu_2, \mu_3\}] + \text{Juneq}[\omega, \text{rate}, \{D1, D2, D3\}, \{\mu_1, \mu_2, \mu_3\}, \{f_1, f_2, f_3\}] + \text{Juneq}[\omega, \text{rate}, \{D1, D2, D3\}, \{f_1, f_2, f_3\}, \{\mu_1, \mu_2, \mu_3\}] + \text{Jeq}[\omega, \text{rate}, \{D1, D2, D3\}, \{f_1, f_2, f_3\}, \{f_1, f_2, f_3\}];$$

$w0 = \{\};$

$wCmH = \{\};$

$wCpH = \{\};$

$wCar = \{\};$

$wHyd = \{\};$

//Magnetic field strength

$B_{\text{field}} = 11.74;$

//Larmor frequencies of Carbon-13 and Hydrogen

$\omega_C = \gamma_C * B_{\text{field}}$

$\omega_H = \gamma_H * B_{\text{field}}$

//Example of calculation of spectral densities for a single structure

$\text{mod} = \{\};$

//Diffusion tensor parameters

$\text{Diff} = \{2.065 * 10^7, 2.167 * 10^7, 3.547 * 10^7\};$ *//Diffusion eigenvalues*

//Diffusion eigenvectors

$D2_{\text{vec}} = \{0.9621, 0.0194, -0.2720\};$

$D1_{\text{vec}} = \{0.2716, 0.0216, 0.9622\};$

$D3_{\text{vec}} = \{0.0245, -0.9996, 0.0155\};$

$D1_{\text{vec}} /= \text{Norm}[D1_{\text{vec}}];$

$D2_{\text{vec}} /= \text{Norm}[D2_{\text{vec}}];$

$D3_{\text{vec}} /= \text{Norm}[D3_{\text{vec}}];$

```

//Assignment of diffusion tensor eigenvectors to Cartesian basis vectors
Dxvec = D2vec;
Dyvec = D1vec;
Dzvec = D3vec;

//Check to see if the diffusion tensor frame is a right-handed coordinate system (HYDRONMR
does NOT always produce a right-handed coordinate system)
Cross[Dxvec,Dyvec].Dzvec

//Information about the atomic bond in a given PDB file (either original or modified)
(* U38 *)
//Bond orientations for the two sites of the base-libration in the PDB coordinate system
site1p= {-0.7391,-0.1744,0.6506};
site2p={-0.6399,-0.1736,0.7486};

//Recalculating the sites in terms of the diffusion tensor principal axis system
site1 = {site1p.Dxvec, site1p.Dyvec, site1p.Dzvec};
site2 = {site2p.Dxvec, site2p.Dyvec, site2p.Dzvec};
site1 /= Norm[site1];
site2 /= Norm[site2];

//Calculation of the spectral densities relevant to the solution relaxation times and NOEs as
defined by Equations III.10-12
Jw0 = J0[0, rate, Diff, site1, site2];
JCmH = J0[ $\omega_C - \omega_H$ , rate, Diff, site1, site2];
JCpH = J0[ $\omega_C + \omega_H$ , rate, Diff, site1, site2];
JH = J0[ $\omega_H$ , rate, Diff, site1, site2];
JC = J0[ $\omega_C$ , rate, Diff, site1, site2];

AppendTo[mod, {"Model 1", Jw0, JCmH, JCpH, JH, JC}];

AppendTo[specd, mod];
Print[TableForm[SetPrecision[mod, 6]]];
AppendTo[w0, Jw0];
AppendTo[wCmH, JCmH];
AppendTo[wCpH, JCpH];
AppendTo[wHyd, JH];
AppendTo[wCar, JC];
1/tauvec[Diff]
*****
//Pairwise combination of the spectral densities from different structures

pop1 = 0.5;
pop2 = 1-pop1;
T1mat = {};

```

```

T2mat = {};

NOEmat = {};

For[i = 0, i < 11, i++;
  T1int = {};
  T2int = {};
  NOEint = {};
  For[j = 0, j < 11, j++;

    //Linear combinations of spectral density constributions from each member in the pair
    Jw0avg = pop1*w0[[i]] + pop2*w0[[j]];
    JcmHavg = pop1*wcmH[[i]] + pop2*wcmH[[j]];
    JCpHavg = pop1*wCpH[[i]] + pop2*wCpH[[j]];
    JHavg = pop1*wHyd[[i]] + pop2*wHyd[[j]];
    JCavg = pop1*wCar[[i]] + pop2*wCar[[j]];

    //T1 calculation
    T1avg = (d2*(JcmHavg + 3*JCavg + 6*JCpHavg) + c2[ωC]*JCavg)^-1;
    AppendTo[T1int, 1000*T1avg];
    //T2 calculation
    T2avg = (0.5*d2*(4*Jw0avg + JcmHavg + 3*JCavg + 6*JHavg + 6*JCpHavg) +
    (1/6)*c2[ωC]*(4*Jw0avg + 3*JCavg))^ -1;
    AppendTo[T2int, 1000*T2avg];

    //NOE calculation
    NOEavg = 1 + (γH/γC)*d2*(6*JCpHavg - JcmHavg)*T1avg;
    AppendTo[NOEint, NOEavg];

  ];
  AppendTo[T1mat, T1int];
  AppendTo[T2mat, T2int];
  AppendTo[NOEmat, NOEint];
]
//Output of the relaxation times and NOEs only for exchange of all other structures with the
lowest energy structure of TAR RNA, 1ANR-1
T1mat[[1]]
T2mat[[1]]
NOEmat[[1]]

*****

```

Chapter IV

Theory of Non-rigid Rotational Motion Applied to NMR Relaxation in RNA

Abstract

Solution NMR spectroscopy can elucidate many features of the structure and dynamics of macromolecules, yet relaxation measurements, the most common source of experimental information on dynamics, can only sample certain ranges of dynamic rates. A complete characterization of motion of a macromolecule therefore requires the introduction of complementary experimental approaches. Solid-state NMR successfully probes the nanoseconds to microseconds (ns to μ s) time scale, a dynamic window where solution NMR results have been deficient, and probes conditions where the averaging effects of rotational diffusion of the molecule are absent. Combining the results of the two distinct techniques in a single framework provides greater insight into dynamics, but this task requires the common interpretation of results recorded under very different experimental conditions. Here we provide a unified description of dynamics that is robust to the presence of large-scale conformational exchange, where the diffusion tensor of the molecule varies on a time scale comparable to rotational diffusion in solution. We apply this methodology to the HIV-I TAR RNA molecule, where conformational rearrangements are both substantial and functionally important. The formalism described herein is of greater generality than earlier combined solid state/solution NMR interpretations, if detailed molecular structures are available, and can offer a more complete description of RNA dynamics than either solution or solid state NMR alone.

1. Introduction

The last few years have seen a growing interest in the roles played by motion and conformational adaptation in mediating protein-nucleic acid interactions. It is now widely recognized that many RNAs^{2,5-7,9,67,69-72,97-99} and some DNAs^{18,22,100,101} function by undergoing large conformational changes when binding to proteins or small molecule ligands. In some cases, fluctuations of the native structure may pre-dispose an RNA or DNA sequence to form interactions with other molecules, suggesting that a protein may in effect “capture” a structurally labile nucleic acid as it fluctuates through a conformation that is optimized for binding. Therefore a thorough description of the physical basis for specific recognition requires that static structural data obtained by crystallography and/or NMR are augmented by information on the conformational changes that each component undergoes to form a complex. Such information includes the dynamic amplitudes associated with functionally relevant structural changes and the rates at which such structural changes occur.

A classical paradigm for the importance of dynamics in RNA function is provided by the transactivation response element (TAR) RNA (Figure 1) from the human immunodeficiency virus type-1 (HIV-1). This RNA plays a role in transcription elongation and is thus required for HIV replication, making it a potential drug target. Relevant to the present investigation, binding of TAR RNA to the HIV-1 Tat protein requires extensive structural rearrangement of the RNA.^{6,7,9,67,69-72,97} The structural properties of the bound and unbound conformational states of TAR RNA are known,^{6,8,73,74} yet, despite recent progress,^{14,75} there is still considerable uncertainty regarding how these structural changes occur. For example, it is not yet certain whether the TAR RNA is driven into its bound structural state only upon actual contact with Tat,

or if TAR fluctuates through a structural state optimized for binding that is subsequently “captured” by the Tat protein, or some combination of the two mechanisms.

The local dynamics of TAR RNA have been analyzed using a variety of NMR approaches, including solution NMR relaxation methods.^{14,58,72} Detailed studies of $^{13}\text{C}/^1\text{H}$ line widths and relaxation rates in and around the bulged loop of native TAR, and the measurement of residual dipolar couplings (RDCs), have provided strong evidence for the existence of ns- μs time scale collective motions of the distal helices about a dynamic hinge associated with the bulged loop of TAR.^{11,14} TAR mutants consisting of elongated distal helices were used to demonstrate that these concerted helical motions occur at rates nearly coincident with the time scale of the overall tumbling in native TAR.^{12,14} The proposed helical motions are significant enough in amplitude to deform the tumbling TAR molecule and affect its hydrodynamics.⁷⁷

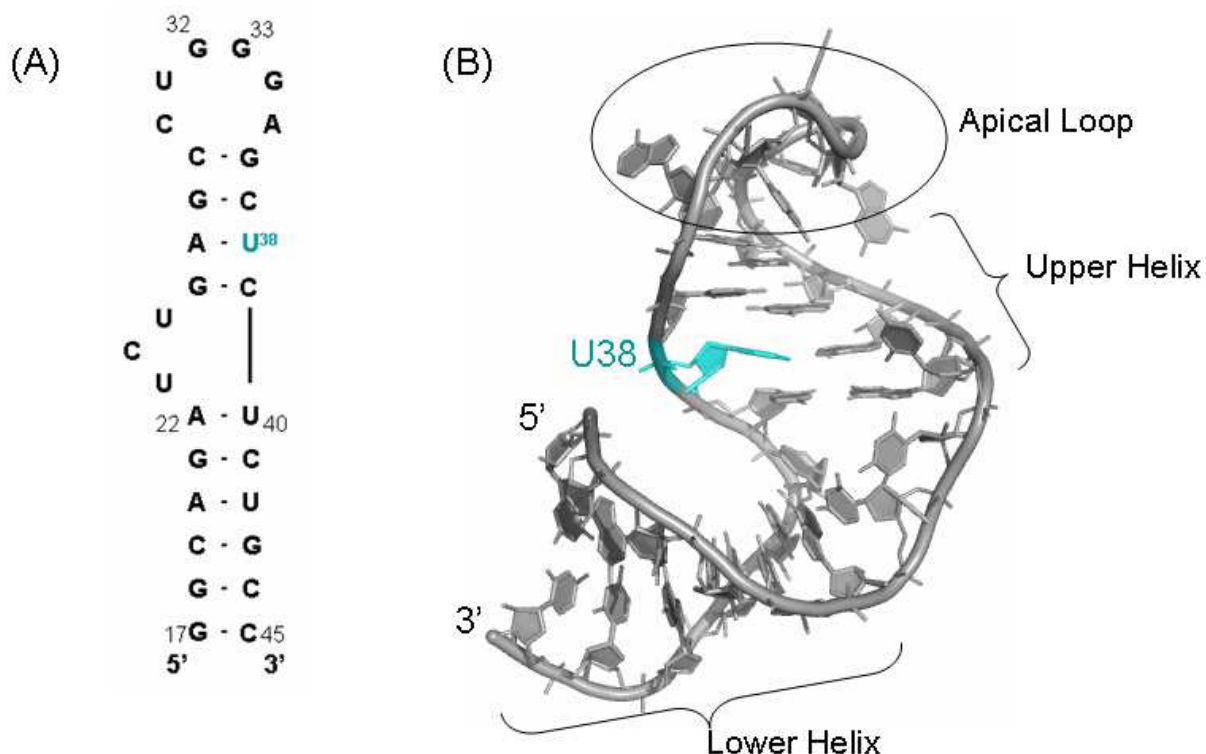


Figure IV.1. A) Sequence and secondary structure of the HIV-1 TAR RNA construct used in this study. The U38 residue, which was deuterium labeled at the H5 and H6 sites for solid-state NMR studies and is the subject of the present theoretical study, is highlighted in cyan. Watson-Crick base pairings are shown with short dashes. B) Lowest energy free TAR structure 1ANR. The U38 residue is shown in cyan. The lower helix, upper helix and apical hairpin loop regions are also indicated. Hydrogen atoms have been removed from the figure for clarity.

These solution NMR relaxation and RDC data have been interpreted as supporting a conformational capture mechanism as the basis for the TAR-Tat interaction, but quantitative interpretation of solution NMR data in terms of a specific sequence of structural changes of TAR has been elusive due in part to the complexity of the dynamic mode-coupling problem.^{12,14,78} In a recent paper, we considered the problem of the non-rigid tumbling of TAR, motivated by solid-state NMR (ssNMR) results⁷⁵ suggesting that the rate of hydrodynamically-significant helical domain motions occur on a timescale significantly longer than the characteristic molecular tumbling times. By extrapolating to the case of infinitely slow exchange between discrete TAR conformers in solution, we simplified the problem to two independently tumbling molecular

conformers, each of which contributed to a fraction of the solution T_1 and T_2 relaxation rates determined by its probability of occurrence. This *slow-exchange* (SE) formalism was applied to the dynamics observed for residue U38, a base paired residue in TAR RNA, for which the rate and amplitude of conformational exchange were obtained using ^2H ssNMR by incorporating uridine-5,6- $^2\text{H}_2$ at the U38 residue.^{75,79} Not only did the SE formalism account for the relaxation properties of the ^{13}C spins of U38, but with small adjustments to account for minor structural variations, the solution relaxation properties of all pyrimidine ^{13}C spins in the upper helix of TAR could be accounted for as well.

In the current manuscript, we develop a theory of non-rigid biomolecular rotation that allows for an arbitrary rate of exchange between conformers, in effect relaxing the SE assumption of our previous work. We apply the expressions for the solution T_1 and T_2 relaxation rates derived herein to the motion of C6-H6 bond of the U38 residue. While comparisons to solution data in this manuscript are limited to relaxation times for the U38- $^{13}\text{C}6$ site, the method is applicable to any site and residue for which motional models (ssNMR-derived or otherwise) are available. Expressions are obtained for the case of instantaneous jumps between discrete conformers with arbitrary rotational diffusion tensors, and solution relaxation rates are calculated for various motional parameters.

In section 2, we present the derivation of the expressions for the solution T_1 and T_2 relaxation times for a general rate of exchange between conformers. We also reiterate the solid-state models used in the simulations. In section 3, we present the results obtained when applying our theory to sets of structures generated by the artificial modification of upper helical angles relative to the lower helix, after showing the correspondence of results of the general theory with those of the slow exchange formalism from the previous approach. In section 4, we discuss these

results and their implications for understanding the full range of dynamics of TAR-RNA under two different sets of sample conditions.

2. Theory

In this section, we derive an expression for the correlation function for a reorienting macromolecule, modeled as a Brownian rotator, and characterized by a fully anisotropic diffusion tensor. As a result of conformational changes occurring within the RNA, the molecular shape changes in such a way as to make the diffusion tensor time-dependent. This time dependence of the diffusion tensor is modeled as an exchange between discrete structural conformers, as described in our prior work.⁷⁵ However, in the present work, the assumption of time scale separation between the rate of conformational exchange and molecular reorientation is not made. The two types of motion may therefore occur at arbitrary time scales.

Rotational diffusion for a fully-anisotropic diffusion tensor has been previously considered in a number of cases. Perrin derived the time correlation function for rotational diffusion about three perpendicular axes of an ellipsoid and used the expression to calculate the fluorescence depolarization that results from the rotation of ellipsoidal molecules in solution.¹⁰² Woessner⁸⁴ extended Perrin's formalism to calculate spectral densities and then the rate of Zeeman relaxation of dipole-coupled spins attached to a reorienting ellipsoid. Favro⁸³ exploited the similarity of the rotational diffusion equation to Schroedinger's equation for a rigid rotor. For a rigid body with an axis of symmetry, Favro showed that the Green's function (i.e. conditional probability) is a linear combination of rigid rotor eigenfunctions. For a fully asymmetric diffusion tensor, this expansion is only approximate, but the first few eigenfunctions and their coefficients were calculated. Huntress extended the isotropic diffusion equation to account for

fully anisotropic rotation and solved for the relaxation time of a nuclear spin undergoing fully anisotropic rotational diffusion assuming a number of spin interaction mechanisms.¹⁰³ Huntress also showed how the measurement of the relaxation times of appropriate nuclei in the molecule can give the full rotational diffusion tensor. A similar approach was used by Freed to calculate the rate of electron spin relaxation resulting from anisotropic rotational diffusion.¹⁰⁴

These approaches all aimed to solve the diffusion equation for fully anisotropic rotational diffusion tensors, but differed in the specific mathematical techniques used to reach this end. For example, Woessner used the direction cosine formalism to construct a correlation function for rotation of ellipsoidal molecules, whereas Favro formulated basically the same problem using Cayley-Klein coordinates. We have repeated the analysis as follows, and explicitly present the derivation in a slightly different manner for clarity.

The evaluation of relaxation times in solid-state and solution NMR requires the evaluation of ensemble-averaged two-time correlations functions of the orientations of the labeled site. The orientation function often takes the form of a Legendre polynomial whose argument is the dot product between the unit vectors $\hat{n}(\tau)$ at time $\tau = 0$ and time $\tau = t$. This is the case for a molecule that is freely diffusing in a solvent. The expression for this polynomial can be expanded using the well-known addition theorem for spherical harmonics (see, for example, Jackson¹⁰⁵):

$$\langle P_l(\hat{n}(0) \cdot \hat{n}(t)) \rangle = \frac{4\pi}{2l+1} \sum_{m=-l}^l \langle Y_{lm}(\hat{n}(0)) Y_{lm}^*(\hat{n}(t)) \rangle \quad (1)$$

The unit vector orientations of the labeled site in the above equation are given in the laboratory frame of the experiment. In order to simplify the evaluation of the given expression we can define a molecule-fixed frame that rotates relative to the lab frame and express the

orientation functions in terms of the vector position of the labeled site in this molecule-fixed frame. This is achieved by expressing the spherical harmonics of the laboratory frame orientations as rotated functions of spherical harmonics in the molecule-fixed frame. Moreover, since the motion considered here is the free-diffusion of the tumbling molecule in the solvent, it is reasonable to use the diffusion tensor principal axis frame (PASd) as the molecule-fixed frame.

$$\begin{aligned} \langle P_l(\hat{n}(0) \cdot \hat{n}(t)) \rangle &= \frac{4\pi}{2l+1} \sum_{m=-l}^l \langle Y_{lm}(\hat{n}(0)) Y_{lm}^*(\hat{n}(t)) \rangle \\ &= \frac{4\pi}{2l+1} \sum_{m=-l}^l \left(\sum_{a,a'=-l}^l \langle D_{ma'}^{(l)*}(\vec{\Omega}_0) Y_{la'}(\theta_0, \phi_0) D_{ma}^{(l)}(\vec{\Omega}_t) Y_{la}^*(\theta_t, \phi_t) \rangle \right) \end{aligned} \quad (2)$$

Here, $\vec{\Omega}_0$ and $\vec{\Omega}_t$ are the Euler angle rotations required to rotate the PASd frame into the laboratory frame, the (θ, ϕ) angles are the orientations of the labeled site vectors in the PASd frame, the $D_{mn}^{(l)}$'s are the Wigner rotation matrices of angular momentum l following the convention of Rose⁵⁶ and Tinkham⁵⁷:

$$D_{mn}^{(l)}(\alpha, \beta, \gamma) = e^{-im\alpha} d_{mn}^{(l)}(\beta) e^{-in\gamma} \quad (3)$$

In this work, we want to consider the effect of including exchanges between conformers with different diffusion tensors. The effect of such exchange on the correlation function is calculated below by generalizing the method of Wong, Case and Szabo⁸² to include diffusion tensors with three distinct eigenvalues (i.e. a fully-anisotropic tensor). As a first step, we can explicitly separate out the dependence of the ensemble average on the relative populations of the various conformers, with the conformer at time $\tau = 0$ being labeled as α , and the conformer at time $\tau = t$ being labeled as β :

$$\begin{aligned} & \langle P_l(\hat{n}(0) \cdot \hat{n}(t)) \rangle \\ &= \frac{4\pi}{2l+1} \sum_{m, a, a'=-l}^l \sum_{\alpha, \beta=1}^{N_{\text{Conformers}}} \langle D_{ma'}^{(l)*}(\vec{\Omega}_0) D_{ma}^{(l)}(\vec{\Omega}_t) \rangle_{\alpha\beta} Y_{la'}(\theta_\alpha, \phi_\alpha) Y_{la}^*(\theta_\beta, \phi_\beta) P_{eq}(\alpha) \end{aligned} \quad (4)$$

Here $P_{eq}(\alpha)$ is the *a priori* the probability of finding the molecule in conformation α .

In this formalism, the problem is reduced to an evaluation of $\langle D_{ma'}^{(l)*}(\vec{\Omega}_0) D_{ma}^{(l)}(\vec{\Omega}_t) \rangle_{\alpha\beta}$ for tumbling motion that occurs while the molecule also exchanges between various conformers with different diffusion tensors. First, we will find the probability for a *single conformer* to transition from an orientation of $\vec{\Omega}_0$ at time $\tau = 0$ to an orientation of $\vec{\Omega}_t$ at time $\tau = t$; the results of this calculation will then be generalized to incorporate exchange between different conformers. Considering that problems of interest in solid-state and solution NMR mostly, if not always, involve tensors of order $l = 2$, we will confine attention in the following to the evaluation of correlation functions of $l = 2$ rotation matrices.

a. General solution to single conformer diffusion equation

The free-diffusion equation for a rigid-body rotor is:

$$\frac{\partial}{\partial t} P(\vec{\Omega}, t) = - \sum_{i,j=1}^3 \hat{L}_i D_{ij} \hat{L}_j P(\vec{\Omega}, t) \quad (5)$$

Here the operators $\hat{L}_i$ represent the angular momentum operator about the i^{th} axis and $P(\vec{\Omega}, t)$ is the probability that the molecular axis system will have rotated through an Euler angle vector $\vec{\Omega}$ at a time t relative to its initial orientation in the laboratory frame at time $t = 0$. The coefficients D_{ij} are the elements of the 2^{nd} order rotational diffusion tensor for the molecule.

The right hand side (RHS) of the equation can be expanded in the principal axis frame of the diffusion tensor (PASd), where the diffusion tensor is diagonal and has eigenvalues $\{D_x, D_y, D_z\}$:

$$\sum_{i,j=1}^3 \hat{L}_i D_{ij} \hat{L}_j = D_x \hat{L}_x^2 + D_y \hat{L}_y^2 + D_z \hat{L}_z^2 \quad (6)$$

This expression can be further rewritten in terms of the operators $\hat{L}^2, \hat{L}_z$ and $\hat{L}_{\pm} = \hat{L}_x \pm i\hat{L}_y$, the lowering/raising (+/-) operators for the z-angular momentum (the $\hat{L}_+$ operator is the lowering operator when the operators are expressed in the PASd frame as opposed to the laboratory frame; this occurs because the commutation relations between operators change sign in transforming from the laboratory to the PASd frame):

$$\begin{aligned} \sum_{i,j=1}^3 \hat{L}_i D_{ij} \hat{L}_j &= D_x \hat{L}_x^2 + D_y \hat{L}_y^2 + D_z \hat{L}_z^2 \\ &= D_z \hat{L}_z^2 + D_{\perp} \left(\hat{L}^2 - \hat{L}_z^2 \right) + \frac{\Delta}{4} (\hat{L}_+^2 + \hat{L}_-^2) \end{aligned} \quad (7)$$

In Eq 7, $D_{\perp} = \frac{D_x + D_y}{2}$ and $\Delta = D_x - D_y$. The assumption in this formalism is that the component D_z is the largest eigenvalue of the diffusion tensor, followed by D_x and finally D_y . It is easy to redefine the axes to suit any of the situations ranging from a prolate ellipsoid to an oblate ellipsoid and for all the fully-anisotropic cases in between these two extremes.

The problem simplifies considerably in the cylindrically-symmetric case of $D_x = D_y$. The eigenfunctions of the diffusion operator in this instance are eigenfunctions of the operators

$\hat{L}^2$ and $\hat{L}_z$, and can be expressed in terms of the Wigner rotation matrices:

$$|l, m, k\rangle = \left(\frac{2l+1}{8\pi^2} \right)^{1/2} D_{mk}^{(l)*}(\alpha, \beta, \gamma) \quad (8)$$

where $\{\alpha, \beta, \gamma\}$ are the Euler angles that transform from the PASd frame to the laboratory frame.

In order to proceed to the fully-anisotropic case, it is useful to understand the Wigner rotation matrices from a geometric perspective: as $D_{mk}^{(l)}(\vec{\Omega} = \{\alpha, \beta, \gamma\}) = \langle l, m | e^{-i\hat{L} \cdot \vec{\Omega}} | l, k \rangle$, the Wigner rotation matrices may be considered as the overlap between the eigenkets of the $\hat{L}^2$ operator expressed in the laboratory ($\{\hat{L}_x, \hat{L}_y, \hat{L}_z\}$) and diffusion frames ($\{\hat{L}_x, \hat{L}_y, \hat{L}_z\}$), where the two frames are connected by an Euler transformation denoted as $\vec{\Omega} = \{\alpha, \beta, \gamma\}$. In the expression $e^{-i\hat{L} \cdot \vec{\Omega}}$, it is understood that the angular momentum operator is expanded in an Euler angle basis, to match the symbolic representation of $\vec{\Omega} = \{\alpha, \beta, \gamma\}$. The angular momentum operators in one of the frames commute with all the operators in the other frame. Thus, given that the rotational diffusion operator, as expressed in Eq 6, commutes with both $\hat{L}^2$ and with $\{\hat{L}_x, \hat{L}_y, \hat{L}_z\}$, the eigenvalues l and m from Eq 8, where m is associated with the laboratory frame operator $\hat{L}_z$, will be good eigenvalues even for the fully-anisotropic diffusion equation.

Therefore, we can conclude that the general solution to the fully-anisotropic problem can be expressed as a linear combination of the Wigner rotation matrices:

$$P(\vec{\Omega}, t | \vec{\Omega}_0) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \left(\sum_{p_l=1}^{2l+1} \left(\sum_{k=-l}^l F_{lk}^{(p_l)}(\vec{\Omega}_0, t) D_{mk}^{(l)*}(\vec{\Omega}) \right) \right) \quad (9)$$

where the index p_l represents the label of the eigenvalue of the diffusion operator for a fixed value of the angular momentum l , and runs from 1 to 5 in the present work. In order to show how this expansion may be achieved, we next consider the operator expressed for a value of $l = 2$, since this is the case relevant to solid state NMR applications.

Using the notation $\hat{H} = D_z \hat{L}_z^2 + D_\perp \left(\hat{L}^2 - \hat{L}_z^2 \right) + \frac{\Delta}{4} (\hat{L}_+^2 + \hat{L}_-^2)$, we can define the matrix

$H_{k'k}^{(l)} = \langle l, m, k' | \hat{H} | l, m, k \rangle$ and express it for the case of $l = 2$ as:

$$\tilde{H}^{(2)} = \begin{pmatrix} 4D_z + 2D_\perp & 0 & \Delta\sqrt{6}/2 & 0 & 0 \\ 0 & D_z + 5D_\perp & 0 & 3\Delta/2 & 0 \\ \Delta\sqrt{6}/2 & 0 & 6D_\perp & 0 & \Delta\sqrt{6}/2 \\ 0 & 3\Delta/2 & 0 & D_z + 5D_\perp & 0 \\ 0 & 0 & \Delta\sqrt{6}/2 & 0 & 4D_z + 2D_\perp \end{pmatrix} \quad (10)$$

The rows and columns are labeled from right to left and top to bottom by $k = -2, -1, 0, 1, 2$.

The off-diagonal elements have been computed using the properties of the lowering and

raising operators $\hat{L}_\pm = \hat{L}_x \pm i\hat{L}_y$:

$$\langle l, m, k | \hat{L}_+^2 + \hat{L}_-^2 | l, m, k \pm 2 \rangle = \sqrt{l(l+1) - (k \pm 2)(k \pm 1)} \cdot \sqrt{l(l+1) - (k \pm 1)k} \quad (11)$$

The eigenvalues and eigenvectors of the above operator matrix will be linear combinations of the Wigner rotation matrices of order $l = 2$. Moreover, as the matrix is real and symmetric, and given that the Wigner rotation matrices are orthogonal with respect to the indices l, m and k , the matrix $\tilde{H}^{(2)}$ will yield eigenvectors that are orthogonal for non-degenerate eigenvalues.

Table IV.1: Eigenvalues and eigenvectors of the matrix $\vec{H}^{(2)}$ defined in Eq 10. The quantity Σ is defined as $\Sigma = (D_x^2 + D_y^2 + D_z^2) - (D_x D_y + D_y D_z + D_z D_x)$.

Eigenvalue	Eigenvector elements $f_{l,k}^{(Eigenvalue)}$ (only non-zero elements are listed)
$\lambda_1 = 4D_z + 2D_\perp = 4D_z + D_x + D_y$	$f_{2,k=\pm 2}^{(1)} = \pm \frac{1}{\sqrt{2}}$
$\lambda_2 = 2D_z + 4D_\perp + \sqrt{4(D_z - D_\perp)^2 + 3\Delta^2}$ $= 2(D_x + D_y + D_z) + 2\sqrt{\Sigma}$	$f_{2,k=\pm 2}^{(2)} = \frac{\sqrt{6}(D_x - D_y)}{4\sqrt{2\Sigma + (D_x + D_y - 2D_z)\sqrt{\Sigma}}},$ $f_{2,k=0}^{(2)} = \frac{(D_x + D_y - 2D_z) + 2\sqrt{\Sigma}}{2\sqrt{2\Sigma + (D_x + D_y - 2D_z)\sqrt{\Sigma}}}$
$\lambda_3 = 2D_z + 4D_\perp - \sqrt{4(D_z - D_\perp)^2 + 3\Delta^2}$ $= 2(D_x + D_y + D_z) - 2\sqrt{\Sigma}$	$f_{2,k=\pm 2}^{(3)} = \frac{\sqrt{6}(D_x - D_y)}{4\sqrt{2\Sigma - (D_x + D_y - 2D_z)\sqrt{\Sigma}}},$ $f_{2,k=0}^{(3)} = \frac{(D_x + D_y - 2D_z) - 2\sqrt{\Sigma}}{2\sqrt{2\Sigma - (D_x + D_y - 2D_z)\sqrt{\Sigma}}}$
$\lambda_4 = D_z + 5D_\perp + \frac{3}{2}\Delta$ $= D_z + 4D_x + D_y$	$f_{2,k=\pm 1}^{(4)} = \frac{1}{\sqrt{2}}$
$\lambda_5 = D_z + 5D_\perp - \frac{3}{2}\Delta$ $= D_z + D_x + 4D_y$	$f_{2,k=\pm 1}^{(5)} = \pm \frac{1}{\sqrt{2}}$

The eigenvalues and eigenvectors of the matrix $\vec{H}^{(2)}$ are listed in Table 1, where the eigenvector elements are listed as the coefficients of the 2nd order Wigner rotation matrix

$D_{mk}^{(l)}(\vec{\Omega})$ for $k = -2, -1, 0, 1, 2$. The quantity Σ is defined as:

$$\Sigma = (D_x^2 + D_y^2 + D_z^2) - (D_x D_y + D_y D_z + D_z D_x).$$

The same procedure may be in principle repeated for all values of $l = 1, \dots, \infty$. However, due to the orthogonality of the Wigner matrices, the correlation functions of Wigner matrices of order l may be also calculated exactly with an evaluation of the eigenvalues and eigenvectors of the matrix $\tilde{H}^{(l)}$ alone. We define the functions

$$A_{mp_l}^{(l)*}(\vec{\Omega}) = \left(\frac{2l+1}{8\pi^2} \right)^{1/2} \sum_{k=-l}^l f_{l,k}^{(p_l)} D_{mk}^{(l)*}(\vec{\Omega}) \quad (12)$$

which satisfy the orthogonality relation

$$\begin{aligned} \int d\vec{\Omega} A_{mp_l}^{(l)*}(\vec{\Omega}) A_{m'p_l'}^{(l')}(\vec{\Omega}) &= \left(\frac{2l+1}{8\pi^2} \right)^{1/2} \left(\frac{2l'+1}{8\pi^2} \right)^{1/2} \sum_{k=-l}^l \sum_{k'=-l'}^{l'} f_{l,k}^{(p_l)} f_{l',k'}^{(p_l')} \int d\vec{\Omega} D_{mk}^{(l)*}(\vec{\Omega}) D_{m'k'}^{(l')}(\vec{\Omega}) \\ &= \sum_{k=-l}^l \sum_{k'=-l'}^{l'} f_{l,k}^{(p_l)} f_{l',k'}^{(p_l')} \delta_{ll'} \delta_{mm'} \delta_{kk'} = \delta_{ll'} \delta_{mm'} \sum_{k=-l}^l f_{l,k}^{(p_l)} f_{l,k}^{(p_l')} \end{aligned} \quad (13a)$$

where the orthogonality of the Wigner matrices was used

$$\int d\vec{\Omega} D_{mk}^{(l)*}(\vec{\Omega}) D_{m'k'}^{(l')}(\vec{\Omega}) = \frac{8\pi^2}{2l+1} \delta_{ll'} \delta_{mm'} \delta_{kk'} \quad (14)$$

Given that the eigenvectors whose elements are represented by the $f_{l,k}^{(p_l)}$'s are orthonormal, the summation in the second line of Equation 13a contracts to a Kronecker delta in the indices p_l and p_l' , yielding:

$$\int d\vec{\Omega} A_{mp_l}^{(l)*}(\vec{\Omega}) A_{m'p_l'}^{(l')}(\vec{\Omega}) = \delta_{ll'} \delta_{mm'} \delta_{p_l p_l'} \quad (13b)$$

Using this relation, and the initial condition $P(\vec{\Omega}, 0 | \vec{\Omega}_0) = \delta(\vec{\Omega} - \vec{\Omega}_0)$, it can be shown that the transition probability takes the form

$$P(\vec{\Omega}, t | \vec{\Omega}_0) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{p_l=1}^{2l+1} e^{-\lambda_{p_l}^{(l)} t} A_{mp_l}^{(l)}(\vec{\Omega}_0) A_{mp_l}^{(l)*}(\vec{\Omega}) \quad (15)$$

b. Solution of the conformational exchange problem

Having solved the problem of single diffusion for a molecule with a single conformer, we follow the approach of Wong, Case and Szabo⁸² in formulating the problem of a molecule exchanging between different conformers with different diffusion tensors while also undergoing diffusive motion. The equation of rotational motion is a coupling of the free-diffusion operator for a single conformer with discrete, instantaneous jumps between different conformations:

$$\frac{\partial}{\partial t} P(\bar{\Omega}, \beta, t | \bar{\Omega}_0, \alpha) = - \sum_{i,j=1}^3 \hat{L}_i D_{ij}^{\beta} \hat{L}_j P(\bar{\Omega}, \beta, t | \bar{\Omega}_0, \alpha) + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} P(\bar{\Omega}, \gamma, t | \bar{\Omega}_0, \alpha) \quad (16)$$

Here $P(\bar{\Omega}, \beta, t | \bar{\Omega}_0, \alpha)$ is the transition probability from the previous section with the additional condition that the molecule starts in conformation α at time $\tau = 0$ and is in conformation β at time $\tau = t$.

The $R_{\beta\gamma}$'s are the weighted rates of transition from conformations γ to β , = Jump

rate $\times \sqrt{\frac{P_{eq}(\beta)}{P_{eq}(\gamma)}}$, where the weighted rate with $\beta = \gamma$ is the negative of the sum of the weighted

rates of transition away from β into other conformations. The weighted rates are used instead of the bare jump rates (which are the numbers regularly quoted in this manuscript) in order to satisfy the condition of detailed balance. It is assumed that the transitions between conformations are instantaneous, and that the diffusion tensor principal axes of each conformer coincide with those of the previous conformer at the instant of transition. These assumptions can be mitigated to a certain extent by increasing the number of structural conformations in the trajectory of the molecule, thereby approximating the situation of continuous underlying atomic motion more closely.

In order to generalize to the case of exchanging and tumbling conformers, we need to (a) make the dependence of the eigenfunctions on the choice of conformer explicit; (b) choose the eigenfunctions of a single conformer as the primary basis (labeled by V); and (c) add an additional index to allow for mixing between eigenfunctions for different eigenvalues (the eigenfunctions of one conformer are not orthogonal to those of another, and so the system of equations represented by Eq. 16 mixes the eigenvalue indices):

$$P(\vec{\Omega}, \beta, t | \vec{\Omega}_0, \alpha) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \left(\sum_{p, p'=1}^{2l+1} c_{lm}^{pp', \beta\alpha}(t) A_{mp}^{(l)*}(\vec{\Omega}, V) A_{mp'}^{(l)}(\vec{\Omega}_0, V) \right) \quad (17)$$

where the initial condition is given by

$$P(\vec{\Omega}, \beta, 0 | \vec{\Omega}_0, \alpha) = \delta^{\beta\alpha} \delta(\vec{\Omega} - \vec{\Omega}_0) \quad (18)$$

Note that in formulating expression 17 we used the fact that the operators in the time evolution equation 16 commute with the operators $\hat{L}^2$ and the lab frame $\hat{L}_z$, and so the most general expansion in terms of $\vec{\Omega}$ and $\vec{\Omega}_0$ is given by expression 17 (i.e. the expansions over an additional set of l and m indices are contracted out by delta functions).

Applying the initial condition (Eq. 18) to expression 17 leads to the following constraint on the coefficients $c_{lm}^{pp', \beta\alpha}(t)$:

$$c_{lm}^{pp', \beta\alpha}(0) = \delta^{pp'} \delta^{\beta\alpha} \quad (19)$$

In order to utilize this expansion in Eq. 16, we need to find the impact of the rotational diffusion operator for an arbitrary conformer β on the common eigenfunction basis. For the sake of clarity we shall use the representation,

$$A_{mp}^{(l)*}(\vec{\Omega}, V) \equiv |A_p(V)\rangle$$

The l and m indices are left implicit due to the fact that an inner product with the eigenstate of any other conformer will force the l and m indices to be the same, i.e.

$$\langle A_p(\beta) | A_p(V) \rangle = \int A_{MP}^{(L)}(\vec{\Omega}, \beta) A_{mp}^{(l)*}(\vec{\Omega}, V) d\vec{\Omega} = \delta_{Ll} \delta_{Mm} \sum_{k=-L}^L f_{L,k}^{(P)}(\beta) f_{L,k}^{(P)}(V), \quad (20a)$$

having utilized the expansions of the respective $A_{mp}^{l*}(\vec{\Omega}, V)$'s in terms of Wigner matrices (Eqs. 12 and 13).

If $\beta = V$, i.e. for the same conformer, expression 20a reduces to a product of delta functions over all indices (Equation 13b),

$$\langle A_p(\beta) | A_p(\beta) \rangle = \delta_{Ll} \delta_{Mm} \delta_{P,p} \quad (20b)$$

Plugging expression 17 into Eq. 16,

$$\begin{aligned} \sum_{l,m,p,p'} A_{mp'}^l(\vec{\Omega}_0, V) \frac{dc_{lm}^{pp',\beta\alpha}(t)}{dt} |A_p(V)\rangle = \\ - \sum_{l,m,p,p'} A_{mp'}^l(\vec{\Omega}_0, V) c_{lm}^{pp',\beta\alpha}(t) \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) |A_p(V)\rangle + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} \sum_{l,m,p,p'} A_{mp'}^l(\vec{\Omega}_0, V) |A_p(V)\rangle c_{lm}^{pp',\gamma\alpha}(t) \end{aligned} \quad (21)$$

Next, we can multiply both sides by $A_{MP'}^{L*}(\vec{\Omega}_0, V)$ and integrate over all values of $\vec{\Omega}_0$, leading to

$$\sum_p \frac{dc_{LM}^{pP',\beta\alpha}(t)}{dt} |A_p(V)\rangle = - \sum_p c_{LM}^{pP',\beta\alpha}(t) \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) |A_p(V)\rangle + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} \sum_p |A_p(V)\rangle c_{LM}^{pP',\gamma\alpha}(t) \quad (22)$$

In the LHS term and the 2nd term on the RHS, expression 20b was used.

We now left-multiply both sides of Eq. 22 with $\langle A_p(V) |$ to yield

$$\begin{aligned} \sum_p \frac{dc_{LM}^{pP',\beta\alpha}(t)}{dt} \langle A_p(V) | A_p(V) \rangle \\ = - \sum_p c_{LM}^{pP',\beta\alpha}(t) \langle A_p(V) | \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) | A_p(V) \rangle + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} \sum_p \langle A_p(V) | A_p(V) \rangle c_{LM}^{pP',\gamma\alpha}(t) \end{aligned} \quad (23)$$

The LHS and the 2nd term on the RHS yield delta functions over the subscripts of the eigenfunctions,

$$\frac{dc_{LM}^{PP',\beta\alpha}(t)}{dt} = -\sum_p c_{LM}^{PP',\beta\alpha}(t) \langle A_p(V) \left| \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) \right| A_p(V) \rangle + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} c_{LM}^{PP',\gamma\alpha}(t) \quad (24)$$

The final procedure is to find an expression for the 1st term on the RHS of Eq. 24. The eigenfunctions of any single conformer form a complete basis w.r.t Euler angle space, and so we can insert the identity operator

$$\hat{I} = \sum_{L',M',Q} |A_Q(\beta)\rangle \langle A_Q(\beta)| \quad (25)$$

on either side of the rotational diffusion operator in Eq. 24:

$$\begin{aligned} & \langle A_p(V) \left| \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) \right| A_p(V) \rangle \\ &= \sum_Q \sum_R \langle A_p(V) | A_Q(\beta) \rangle \langle A_Q(\beta) \left| \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) \right| A_R(\beta) \rangle \langle A_R(\beta) | A_p(V) \rangle \end{aligned} \quad (26)$$

Once again, the L' and M' indices are forced to be equal to L and M, respectively, by equation 20a.

Given that the operator $\left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right)$ is diagonal in the eigenfunction basis for conformer β , we obtain

$$\langle A_p(V) \left| \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) \right| A_p(V) \rangle = \sum_Q \langle A_p(V) | A_Q(\beta) \rangle \lambda_Q(\beta) \langle A_Q(\beta) | A_p(V) \rangle \quad (27)$$

We can use result 20a to simplify this expression in terms of known coefficients:

$$\langle A_p(V) \left| \left(\sum_{i,j=1}^3 \hat{L}_i D_{ij}^\beta \hat{L}_j \right) \right| A_p(V) \rangle = \sum_Q \left(\sum_{k_1=-L}^L f_{L,k_1}^{(P)}(V) f_{L,k_1}^{(Q)}(\beta) \right) \lambda_Q(\beta) \left(\sum_{k_2=-L}^L f_{L,k_2}^{(Q)}(\beta) f_{L,k_2}^{(P)}(V) \right) \quad (28)$$

Thus, the equations for the time dependent coefficients are

$$\frac{dc_{LM}^{PP',\beta\alpha}(t)}{dt} = -\sum_p \sum_Q \left(\sum_{k_1=-L}^L f_{L,k_1}^{(P)}(V) f_{L,k_1}^{(Q)}(\beta) \right) \lambda_Q(\beta) \left(\sum_{k_2=-L}^L f_{L,k_2}^{(Q)}(\beta) f_{L,k_2}^{(P)}(V) \right) c_{LM}^{PP',\beta\alpha}(t) + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} c_{LM}^{PP',\gamma\alpha}(t) \quad (29)$$

In general, this results in a $(2L+1)^2 N_{\text{Conformers}} \times (2L+1)^2 N_{\text{Conformers}}$ matrix for the time-dependent coefficients. However, if we confine our attention to the $L = 2$ case, we notice from Table 1 that the $f_{2,k}^{(P)}$'s are diffusion tensor-dependent only for two of the diffusion eigenvalues, $P = 2$ and $P = 3$. Moreover, the subspace of coefficients for $P = 1, 4$ and 5 is orthogonal to that for $P = 2$ and 3 . This implies that

$$\sum_{k=-2}^2 f_{2,k}^{(P)}(\beta) f_{2,k}^{(Q)}(V) = \delta_{PQ}; \text{ for } P, Q = 1, 4, 5 \quad (30)$$

and so for $P = 1, 4, 5$

$$\frac{dc_{2M}^{PP',\beta\alpha}(t)}{dt} = -\lambda_P(\beta) c_{2M}^{PP',\beta\alpha}(t) + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} c_{2M}^{PP',\gamma\alpha}(t) \quad (31a)$$

In addition, as the initial condition 19 forces the time-dependent coefficients to be diagonal in the indices P and P' at time $t = 0$, the form of Eq. 31a implies that these coefficients will remain diagonal for all time:

$$\frac{dc_{2M}^{PP',\beta\alpha}(t)}{dt} = -\lambda_P(\beta) c_{2M}^{PP',\beta\alpha}(t) + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} c_{2M}^{PP',\gamma\alpha}(t) \text{ for } P = 1, 4, 5 \quad (31b)$$

The equations block-diagonalize into $N_{\text{Conformers}} \times N_{\text{Conformers}}$ blocks for each of these three values of P .

The remaining $4N_{\text{Conformers}} \times 4N_{\text{Conformers}}$ subspace can be diagonalized separately (using Eq. 29):

$$\frac{dc_{LM}^{PP',\beta\alpha}(t)}{dt} = -\sum_p \sum_Q \left(\sum_{k_1=-L}^L f_{L,k_1}^{(P)}(V) f_{L,k_1}^{(Q)}(\beta) \right) \lambda_Q(\beta) \left(\sum_{k_2=-L}^L f_{L,k_2}^{(Q)}(\beta) f_{L,k_2}^{(P)}(V) \right) c_{LM}^{PP',\beta\alpha}(t) + \sum_{\gamma=1}^{N_{\text{Conformers}}} R_{\beta\gamma} c_{LM}^{PP',\gamma\alpha}(t)$$

for $P, P' = 2, 3$ (32)

These represent $4N_{\text{Conformers}}$ equations for the coefficients $c_{2m}^{22,\beta\alpha}(t)$, $c_{2m}^{23,\beta\alpha}(t)$, $c_{2m}^{32,\beta\alpha}(t)$ and $c_{2m}^{33,\beta\alpha}(t)$.

A further simplification is possible considering that in Eq. 32 only those coefficients are coupled that share the same P' index. So it is possible to separately diagonalize the equations with $P' = 2$ and those with $P' = 3$ (i.e. the equations for $\{c_{2m}^{22,\beta\alpha}(t), c_{2m}^{32,\beta\alpha}(t)\}$ are separate from those for $\{c_{2m}^{23,\beta\alpha}(t), c_{2m}^{33,\beta\alpha}(t)\}$). This yields two separate $2N_{\text{Conformers}} \times 2N_{\text{Conformers}}$ matrix equations.

In conclusion, the $(2L+1)^2 N_{\text{Conformers}} \times (2L+1)^2 N_{\text{Conformers}}$ -dimensional rate equations represented by Eq. 29 simplify into three $N_{\text{Conformers}} \times N_{\text{Conformers}}$ -dimensional matrix equations (Eqs. 31b) and two $2N_{\text{Conformers}} \times 2N_{\text{Conformers}}$ -dimensional matrix equations (Eqs. 32).

It is worth briefly mentioning the numerical procedures we used for solving the equations for the time-dependent coefficients. For Eqs. 31b, we set up the $N_{\text{Conformers}} \times N_{\text{Conformers}}$ evolution matrices on the RHS for the coefficients (for each value of P), and then solved for the eigenvalues and eigenvectors. For the initial condition 19, it can easily be shown that

$$c_{2M}^{PP,\beta\alpha}(t) = (\vec{T} e^{-\vec{\Lambda}t} \vec{T}^{-1})_{\beta\alpha} \quad (33)$$

where $\vec{T}$ is the matrix with the eigenvectors of the evolution matrix as its columns, and $\vec{\Lambda}$ is the diagonal matrix of eigenvalues. A similar procedure holds for the cases where $P, P' = 2, 3$. Here, we set up vectors of the form

$$\vec{V}_{2,\alpha}(t) = \{c_{2m}^{22,1\alpha}(t), c_{2m}^{32,1\alpha}(t), c_{2m}^{22,2\alpha}(t), c_{2m}^{32,2\alpha}(t), \dots, c_{2m}^{22,N_{\text{Conformers}}\alpha}(t), c_{2m}^{32,N_{\text{Conformers}}\alpha}(t)\}^T, \text{ for each value}$$

of α and solved the coupled equations (just as for the $P, P' = 1, 4, 5$ cases) by setting up the evolution matrix on the RHS of Eq. 32 and finding its eigenvalues and eigenvectors. However, applying the initial condition to obtain an expression similar to Eq. 33 requires an additional step, as the initial conditions force all the $P \neq P'$, i.e. all the even-numbered coefficients of the

vector above to be 0 at time $t = 0$. If you consider the case with $P' = 3$, i.e. the solution for the vector $\vec{V}_{3,\alpha}(t) = \{c_{2m}^{23,1\alpha}(t), c_{2m}^{33,1\alpha}(t), c_{2m}^{23,2\alpha}(t), c_{2m}^{33,2\alpha}(t), \dots, c_{2m}^{23,N_{\text{Conformers}}\alpha}(t), c_{2m}^{33,N_{\text{Conformers}}\alpha}(t)\}^T$, the initial condition requires all the odd-numbered coefficients to be 0. Next, we noticed that the evolution matrices for both $P' = 2$ and $P' = 3$ are the same, due to the lack of any explicit P' dependence in the matrix elements. In light of these facts, it is possible to find the combined solutions to both sets of coefficients as

$$\begin{aligned} c_{2M}^{22,\beta\alpha}(t) &= \left(\tilde{T} e^{-\tilde{\Lambda}t} \tilde{T}^{-1} \right)_{2\beta-1,2\alpha-1}; c_{2M}^{32,\beta\alpha}(t) = \left(\tilde{T} e^{-\tilde{\Lambda}t} \tilde{T}^{-1} \right)_{2\beta,2\alpha-1}; \\ c_{2M}^{23,\beta\alpha}(t) &= \left(\tilde{T} e^{-\tilde{\Lambda}t} \tilde{T}^{-1} \right)_{2\beta-1,2\alpha}; c_{2M}^{33,\beta\alpha}(t) = \left(\tilde{T} e^{-\tilde{\Lambda}t} \tilde{T}^{-1} \right)_{2\beta,2\alpha} \end{aligned} \quad (34)$$

$\alpha, \beta = 1, \dots, N_{\text{Conformers}}$

$\tilde{T}$ is a $2N_{\text{Conformers}} \times 2N_{\text{Conformers}}$ -dimensional matrix with the eigenvectors of the evolution matrix as its columns.

c. Evaluation of the correlation functions between Wigner rotation matrices

Returning to the central problem of evaluating $\left\langle D_{ma'}^{(l)*}(\vec{\Omega}_0) D_{ma}^{(l)}(\vec{\Omega}_t) \right\rangle_{\alpha\beta}$, we can use Eq 17 along with orthogonality relations between Wigner Matrices and a uniform *a priori* probability

of $P(\vec{\Omega}_0) = \frac{1}{8\pi^2}$ to obtain the final result:

$$\begin{aligned} \left\langle D_{ma'}^{(l)*}(\vec{\Omega}_0) D_{ma}^{(l)}(\vec{\Omega}_t) \right\rangle_{\alpha\beta} &= \int d\vec{\Omega}_0 \int d\vec{\Omega} P(\vec{\Omega}_0) P(\vec{\Omega}_t, \beta, t | \vec{\Omega}_0, \alpha) D_{ma'}^{(l)*}(\vec{\Omega}_0) D_{ma}^{(l)}(\vec{\Omega}_t) \\ &= \int d\vec{\Omega}_0 \int d\vec{\Omega}_t P(\vec{\Omega}_0) D_{ma'}^{(l)*}(\vec{\Omega}_0) D_{ma}^{(l)}(\vec{\Omega}_t) \sum_{L=0}^{\infty} \sum_{M=-L}^L \sum_{P,P'=1}^{2L+1} c_{LM}^{PP',\beta\alpha}(t) A_{MP'}^{(L)}(\vec{\Omega}_0, V) A_{MP}^{(L)*}(\vec{\Omega}_t, V) \\ &= \sum_{L=0}^{\infty} \sum_{M=-L}^L \sum_{P,P'=1}^{2L+1} c_{LM}^{PP',\beta\alpha}(t) \left(\frac{1}{2L+1} \right) \sum_{K,K'=-L}^L f_{L,K}^{(P)}(V) f_{L,K'}^{(P')}(V) \delta_{Ll} \delta_{K'a'} \delta_{Ka} \delta_{Mm} \\ &= \left(\frac{1}{2L+1} \right) \sum_{P,P'=1}^{2L+1} c_{lm}^{PP',\beta\alpha}(t) f_{l,a}^{(P)}(V) f_{l,a'}^{(P')}(V) \end{aligned} \quad (35)$$

Using the final expression of Eq 35 in Eq 4, and noting that the summation over m gives a factor of $(2l + 1)$:

$$\begin{aligned} & \langle P_l(\hat{n}(0).\hat{n}(t)) \rangle \\ &= \frac{4\pi}{2l+1} \sum_{a,a'=-l}^l \sum_{\alpha,\beta=1}^{N_{\text{Conformers}}} \sum_{P,P'=1}^{2l+1} c_{lm}^{PP',\beta\alpha}(t) f_{l,a}^{(P)}(V) f_{l,a'}^{(P')}(V) Y_{la}^*(\theta_\beta, \phi_\beta) Y_{la'}(\theta_\alpha, \phi_\alpha) P_{eq}(\alpha) \end{aligned} \quad (36)$$

The right hand side of the above equation can be simplified further by noting that the products $f_{l,a}^{(P)} f_{l,a'}^{(P')}$ are non-zero only for $a' = a \pm 2i, i \in \text{Integers}$, where $(a \pm 2i) \in \{-l, -l+1, \dots, +l\}$. In the following, the expression is explicitly evaluated for the case $l = 2$. The spherical harmonics and those $f_{2,a}^{(P)} f_{2,a'}^{(P')}$'s that are diffusion tensor-independent are expressed explicitly (the $f_{2,a}^{(P)}$'s are diffusion tensor-dependent for $P = 2, 3$), yielding the final result for the correlation function:

$$\begin{aligned} C(t) = \langle P_2(\hat{n}(0).\hat{n}(t)) \rangle = & \sum_{\alpha,\beta=1}^{N_{\text{Conformers}}} \left[\begin{aligned} & c_{2M}^{11,\beta\alpha}(t) \left\{ \frac{3}{4} \sin^2 \theta_\beta \sin^2 \theta_\alpha \sin(2\phi_\beta) \sin(2\phi_\alpha) \right\} \\ & + \frac{3}{2} \left\{ c_{2M}^{22,\beta\alpha}(t) f_{2,2}^{(2)}(V) f_{2,2}^{(2)}(V) + c_{2M}^{32,\beta\alpha}(t) f_{2,2}^{(3)}(V) f_{2,2}^{(2)}(V) \right\} \sin^2 \theta_\beta \sin^2 \theta_\alpha \cos(2\phi_\beta) \cos(2\phi_\alpha) \\ & + c_{2M}^{23,\beta\alpha}(t) f_{2,2}^{(2)}(V) f_{2,2}^{(3)}(V) + c_{2M}^{33,\beta\alpha}(t) f_{2,2}^{(3)}(V) f_{2,2}^{(3)}(V) \left\{ \sin^2 \theta_\beta \cos(2\phi_\beta) (3 \cos^2 \theta_\alpha - 1) \right. \\ & + \frac{\sqrt{3}}{2\sqrt{2}} \left[\left\{ c_{2M}^{22,\beta\alpha}(t) f_{2,2}^{(2)}(V) f_{2,0}^{(2)}(V) + c_{2M}^{32,\beta\alpha}(t) f_{2,2}^{(3)}(V) f_{2,0}^{(2)}(V) \right\} \sin^2 \theta_\beta \cos(2\phi_\beta) (3 \cos^2 \theta_\alpha - 1) \right. \\ & + \left. \left\{ c_{2M}^{23,\beta\alpha}(t) f_{2,0}^{(2)}(V) f_{2,2}^{(3)}(V) + c_{2M}^{33,\beta\alpha}(t) f_{2,0}^{(3)}(V) f_{2,2}^{(3)}(V) \right\} \sin^2 \theta_\alpha \cos(2\phi_\alpha) (3 \cos^2 \theta_\beta - 1) \right] \\ & + \frac{1}{4} \left\{ c_{2M}^{22,\beta\alpha}(t) f_{2,0}^{(2)}(V) f_{2,0}^{(2)}(V) + c_{2M}^{32,\beta\alpha}(t) f_{2,0}^{(3)}(V) f_{2,0}^{(2)}(V) \right\} (3 \cos^2 \theta_\beta - 1) (3 \cos^2 \theta_\alpha - 1) \\ & + c_{2M}^{23,\beta\alpha}(t) f_{2,0}^{(2)}(V) f_{2,0}^{(3)}(V) + c_{2M}^{33,\beta\alpha}(t) f_{2,0}^{(3)}(V) f_{2,0}^{(3)}(V) \left\{ 3 \cos^2 \theta_\beta - 1 \right\} (3 \cos^2 \theta_\alpha - 1) \\ & + c_{2M}^{44,\beta\alpha}(t) \left\{ \frac{3}{4} \sin(2\theta_\beta) \sin(2\theta_\alpha) \sin(\phi_\beta) \sin(\phi_\alpha) \right\} + c_{2M}^{55,\beta\alpha}(t) \left\{ \frac{3}{4} \sin(2\theta_\beta) \sin(2\theta_\alpha) \cos(\phi_\beta) \cos(\phi_\alpha) \right\} \end{aligned} \right] P_{eq}(\alpha) \end{aligned} \quad (37)$$

The result of Eq 37 for the correlation function can then be inserted into the expression for the

spectral density, which is the cosine Fourier transform of $C(t)$, $J(\omega) = \int_0^\infty C(t) \cos \omega t dt$. Finally, the

spectral density occurs in the familiar equations for the relaxation times and the NOE for the case of relaxation mediated by carbon-proton dipole interaction and the carbon chemical shift anisotropy:

$$\frac{1}{T_1} = R_1 = \frac{d^2}{4} [J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)] + c^2 J(\omega_C) \quad (38)$$

$$\frac{1}{T_2} = R_2 = \frac{d^2}{8} \left[4J(0) + J(\omega_H + \omega_C) + 3J(\omega_C) \right] + \frac{c^2}{6} [3J(\omega_C) + 4J(0)] \quad (39)$$

$$\text{NOE} = 1 + \frac{d^2}{4} \left(\frac{\gamma_H}{\gamma_C} \right) [6J(\omega_H + \omega_C) - J(\omega_C - \omega_H)] T_1 \quad (40)$$

$$d = \left[\frac{\mu_0 h \gamma_H \gamma_C}{8 \pi^2 r_{CH}^3} \right], c = \left(\frac{\omega_C}{\sqrt{3}} \right) (\Delta) \quad (41)$$

In Eqns 38-41, ω_H and ω_C are the Larmor frequencies of ^1H and ^{13}C respectively, μ_0 is the permeability of a vacuum, γ_H and γ_C are the magnetogyric ratios of ^1H and ^{13}C , h is Planck's constant ($= 6.626 \times 10^{-34}$ J.sec), r_{CH} is the length of the C-H bond, and Δ is the chemical shift anisotropy (CSA). The effect of CSA tensor asymmetry is generally neglected in these expressions because it is small. The value chosen for r_{CH} in the following analysis was 1.1 Å and Δ was chosen to be 212 ppm. The choices for these parameters are the same as those in previous work⁷⁵ and will not be further justified.

d. Solid-state models

In the following section, we will utilize motional models that have been derived from solid-state NMR experiments¹⁰ to describe T_{1Z} , T_{1Q} relaxation measurements and the solid state lineshape. The observation of these motions was the motivation to develop the theory describe in this manuscript, so we provide a brief summary of the models first (a more detailed summary is provided in our previous work⁷⁵).

Experimental results for the U38 base could be modeled through two motions around two independent axes: the first axis is perpendicular to the base plane and is termed a “base-libration” motion, and the best-fit to the data was obtained with a two-site jump with an amplitude of 8° ($\pm 4^\circ$) and a rate of $2.15 \times 10^8 \text{ s}^{-1}$. The second, termed a “conformational exchange” motion, was interpreted as resulting from a bend in the upper helix of 9° and a twist of 15° about the upper helical axis at a rate of $1.38 \times 10^6 \text{ s}^{-1}$. We note here that the original reference¹⁰ quotes these values as a bend and a twist of 13° each, but one of the authors indicated that better fits were subsequently obtained with these new parameters. Because the solid-state experiments are insensitive to whether the twist is clockwise or counterclockwise, twist angles of either $\pm 15^\circ$ fit the solid-state NMR data equally well.

The models consist of equal populations of the two base-libration sites, as well as equal populations of the two conformers. For the analysis presented here, this means that $P_{eq}(\alpha)$ is 0.5 in all simulations.

The solid-state results were obtained for a hydration of 16 waters per nucleotide. The choice of this value was based on results which showed that the spectral densities of labeled sites did not change substantially for hydrations of 16 waters per nucleotide and higher.⁸⁵

Note: The results that follow in Figures 3 through 11 were based on equations some of which were later found to be erroneous. Corrections were obtained for the equations and some of the simulations were redone to find out the discrepancies. The incorrect equations have been replaced with the correct ones in the “Theory” section above. The new, corrected simulations, however, did not show significant deviations from the ones presented in the original manuscript, and so the “Results” section from the original published manuscript are given here without modification, with the qualitative arguments still being valid. Subsequently, we provide the redone simulations in the section “Results based on corrected equations” and attempt to quantify the discrepancies between the two sets of results.

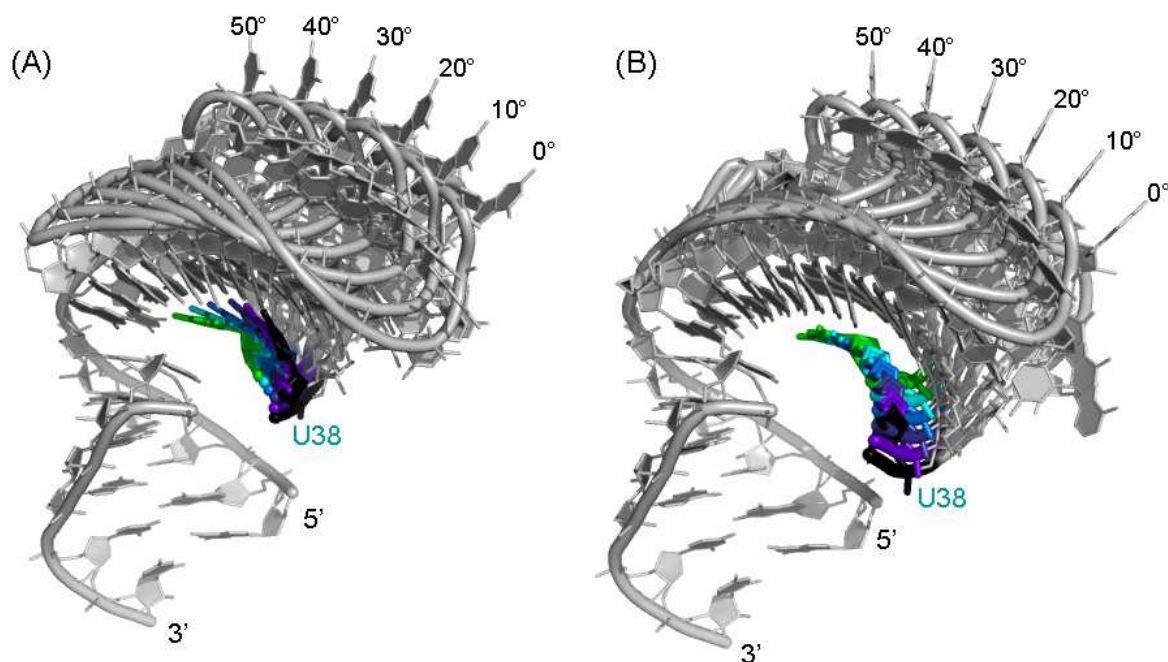


Figure IV.2. Models of TAR generated from the lowest energy TAR RNA structure (1ANR-1) by bending the upper helix relative to its orientation in 1ANR-1. The bend angles are indicated next to the corresponding structures. (A) These models have a twist of 15° applied to the upper helix in addition to the reported bend angles; these structures are denoted as the “Tw15 series.” (B) These models have a twist of -15° applied to the upper helix in addition to the bend angles; they are denoted as the “Tw-15 series.” The labeled U38 base is shown in colors ranging from black (0°) to green (50°). The hydrogen atoms and the C39 residues have been removed from both figures for greater clarity.

3. Results

We present the results of applying Eqs 37 – 40 to calculate solution ^{13}C relaxation times for the HIV-1 Trans-Activation Response (TAR) RNA element.⁷⁵ In order to apply the theory of this manuscript, molecular structures are required to define the exchanging conformers.

According to the results of the solid state NMR data, we employed two sets of PDB files: (1) those available from our previous work,⁷⁵ which reflect a modification of the bend angle of the upper helix relative to the lower helix about various axes perpendicular to the lower helix; and (2) new structures that were generated from the lowest energy TAR structure to include both a bend and a twist of the upper helix relative to lower helix (Figure 2). The structures shown in

Figure 2 have been generated with twists of $\pm 15^\circ$ (Figure 2A corresponds to $+15^\circ$, and Figure 2B to -15°) applied to the upper helix about its own axis, in addition to the various bend angles. All structures considered here were generated from lowest energy structure of the unbound HIV-1 TAR RNA (PDB code 1ANR),⁸ hereafter referred to as 1ANR-1, by rotating the upper helix relative to the lower helix about axes perpendicular to the lower helix, and, in the case of the second set, also twisting the upper helix about its axis of symmetry. In both cases, the bulge orientation was also kept fixed in its 1ANR-1 conformation. The second set also differs from the first set of structures (that was previously described) in the characterization of lower and upper helical axes, which was accomplished using the software package 3DNA.¹⁰⁶ The 1ANR-1 PDB file was rotated such that the z-axis corresponded with the lower helical axis. Any subsequent Euler rotation about the line of nodes would thus be perpendicular to the lower helix. The choice of the line of nodes within the plane perpendicular to the lower helix was made visually by selecting an orientation that resulted in the straightening of the upper helix relative to the lower helix. All Euler transformations were carried out manually by means of the UCSF Chimera program.¹⁰⁷

To be consistent with the previous simulations, we limit our attention to bend angles $\leq 50^\circ$. This restriction is necessary because, as a result of keeping the bulge fixed, structures with bend angles greater than 50° introduce steric clashes between the upper helix and the bulge. Even the 40° and 50° structures in Figure 2A show close residue-residue encounters and many of the structures in Figure 2B contain steric clashes. More accurate models allowing for the rearrangement of the bulge may alleviate these problems, but these local details are very unlikely to affect the global properties of RNA that are studied in the present work.

In order to incorporate the base-libration motion, without any assumptions about the separation of the magnitudes of rates, each structure containing a particular site of the base-libration for each of the conformational exchange partners was treated as a separate conformer. Thus, a two-site base-libration coupled to a two-site conformational exchange process yielded four conformers to be used in Equations 31b and 32. The weighted rates $R_{\beta\gamma}$ were associated with either base-libration jump rates or conformational exchange jump rates, depending on which two conformers were being inter-related.

All simulations in this manuscript were conducted with essentially the same parameters used in the previous work. This included the input parameters for HYDRONMR,⁹² the public-domain program used in the calculation of the diffusion tensor eigenvalues and eigenvectors: an atomic element radius (AER) of 2.3 Å, a temperature of 298 K and solvent viscosity of $\eta_0 = 0.01096$ Poises (the viscosity of 99.9% D₂O at 25° C),⁹³ with the last two parameters determined based on the solution NMR experimental conditions.¹¹ It is worth noting that no variation of the AER was considered in the current work (a discussion of the AER dependence of relaxation times can be found in the previous work⁷⁵), for the sake of brevity. The carbon-hydrogen bond length was set at 1.1 Å and the CSA parameter was taken to be 212 ppm. As in the previous work, we did not examine the heteronuclear NOEs in this work due to the minimal variation of values.

a. Comparison with Slow Exchange formalism

The solution T₁ and T₂ relaxation times were generated for the case of a two-site conformational exchange process between each of the rotated constructs and the original lowest energy unbound structure (1ANR-1), as motivated by solid-state NMR results.¹⁰ The *a priori* probability of each conformer, $P_{eq}(\alpha)$, was taken to be 0.5, as mentioned in the previous section.

Of course, this value of $P_{eq}(\alpha)$ was chosen purely based on the solid-state models and can be varied if so desired (this is not done in the current manuscript). The solid-state NMR results for U38 were well fit using a conformational exchange rate of $1.38 \times 10^6 \text{ s}^{-1}$ between the two conformers, and the best-fit angular excursions consisted of a 9° bend and a twist 15° in amplitude in the upper helix. The solution times were simulated in our previous work under the assumption that this exchange rate between the two conformers, with different diffusion tensors, was substantially slower than the mean tumbling rate of the individual conformers and any other internal motion occurring in each of the two states. A slow exchange formalism was therefore proposed to this effect and implemented for multiple series of structures generated through rotations. In the current manuscript, we repeat the T_1 and T_2 simulations, but removing the assumption of slow exchange.

In the work published previously,⁷⁵ the series of constructs were given the labels “90 degree series” and “30 degree series”, referring to the choice of the upper helix-rotation axis within the plane perpendicular to the lower helix. Therefore, the labels indicate that, in the second series, the upper helix was rotated through an axis that is 60 degrees away from that of the first series within the plane perpendicular to the lower helix. The rotation axes pass through the backbone at the hinge between U40 in the lower helix and C39 in the upper helix. The lower helix and bulge retain their conformations as observed experimentally in 1ANR-1.

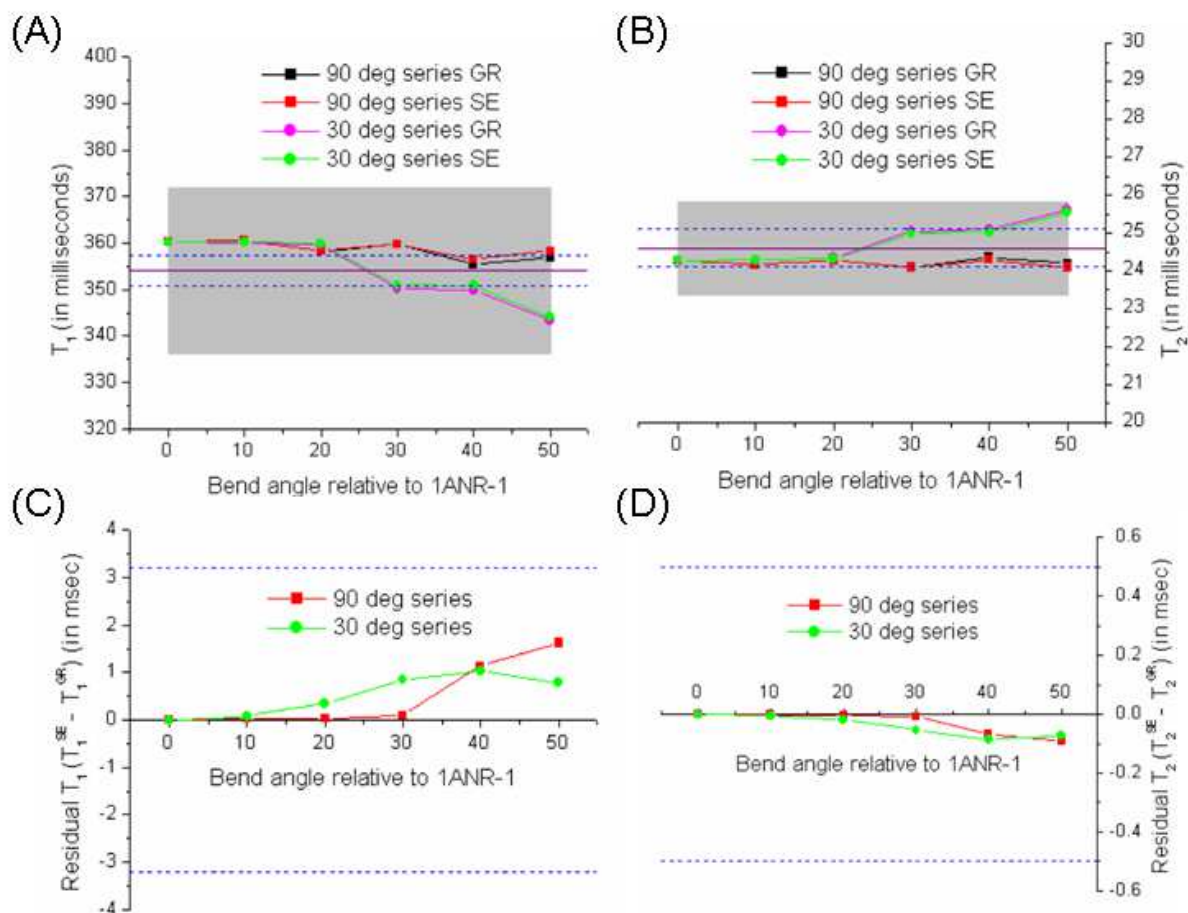


Figure IV.3. Comparison of relaxation times calculated using slow exchange (SE) and general rate (GR) formalisms. (A) T_1 times for the “90 degree series” and “30 degree series.” (B) T_2 times for the “90 degree series” and “30 degree series.” Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines; also shown in grey is the $\pm 5\%$ error region. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3$ ms, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5$ ms. The residuals ($T_{1,2}^{\text{SE}} - T_{1,2}^{\text{GR}}$) are shown in (C) (for T_1) and (D) (for T_2). The experimental error bars are shown again as dashed blue lines.

We first applied the solid-state NMR-derived rate of $1.38 \times 10^6 \text{ sec}^{-1}$ to Eqs 31b and 32. The T_1 and T_2 values are shown in Figures 3A and 3B respectively, as a function of bend angle of the upper helix relative to its orientation in 1ANR-1 (the lower helix and bulge remain fixed), along with the corresponding relaxation times from the slow exchange formalism. The results for both the “90 degree series” (square symbols) and the “30 degree series” (circular symbols) are included in the same graphs. It must be noted that an important correction has been made to the results published previously in producing Figure 3. Namely, the relaxation times from that

reference were mistakenly calculated with the x and y coordinates of the base-libration sites interchanged, an error that was discovered only after the publication of the manuscript. Given the small amplitude of the base-libration, the discrepancy in T_1 values was at most 4.3 ms and that in T_2 was less than 0.4 ms; these differences are small, so the previous conclusions would not change.

The residuals between the relaxation times obtained using the SE formalism and the general rate (GR) theory presented in Eqs. 38 and 39 ($T_{1,2}^{\text{SE}} - T_{1,2}^{\text{GR}}$) are shown in Figures 3C and 3D. For comparison, the experimental relaxation times are $T_1 = 354 \pm 3$ ms and $T_2 = 24.6 \pm 0.5$ ms.¹¹ The experimental times are indicated by the solid purple line, while the error bars are indicated by dashed blue lines. Also shown in grey in Figures 3A and 3B are the $\pm 5\%$ error regions for each of the relaxation times. These regions are included to provide room for additional systematic sources of uncertainty in the simulations arising from choices of the carbon-proton bond length and CSA, and are only visual guidelines

As indicated by the residuals, the two calculations agree within the experimental error at the solid-state derived rate, supporting the SE formalism as the limiting case of the more general rate exchange theory. Conversely, this may be seen as confirming the behavior of the general rate theory in the infinitely slow limit that is familiar from other weighted population studies, such as those used for RDCs (see, for example, Blackledge¹⁰⁸ and Stelzer et al¹⁰⁹).

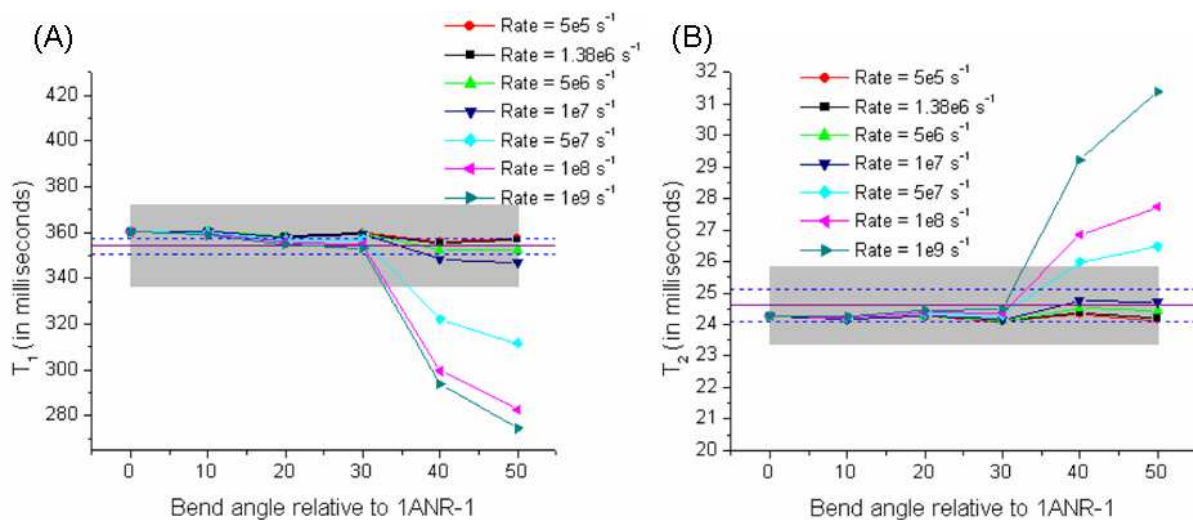


Figure IV.4. Relaxation times vs. bend angle for structures within the “90 degree series,” considering various exchange rates from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$: (A) T_1 times; and (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3 \text{ ms}$, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5 \text{ ms}$. Also shown in grey is the $\pm 5\%$ error region.

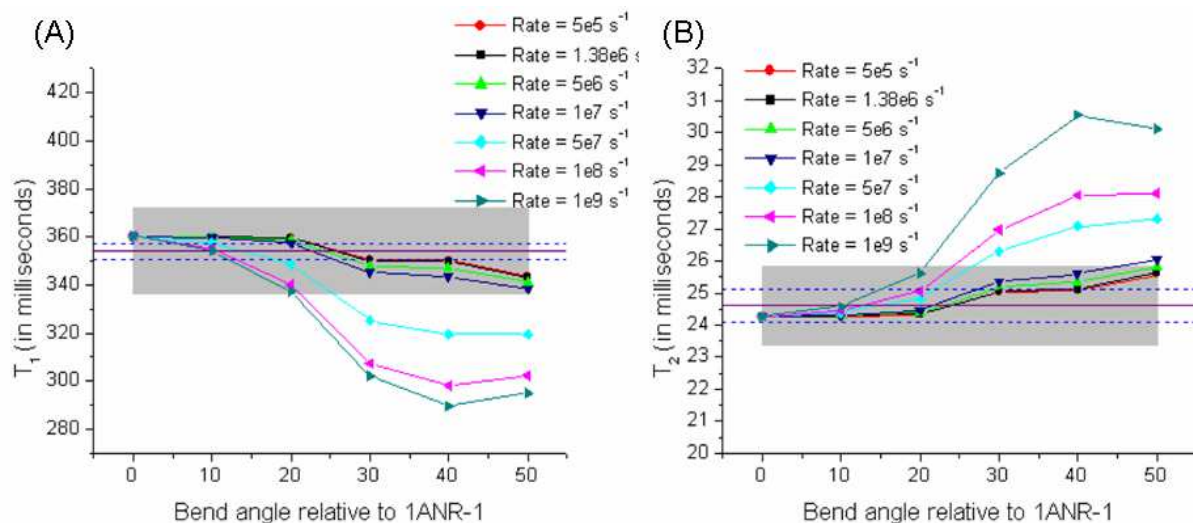


Figure IV.5. Relaxation times vs. bend angle for structures within the “30 degree series,” considering various exchange rates from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$: (A) T_1 times; and (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3 \text{ ms}$, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5 \text{ ms}$. Also shown in grey is the $\pm 5\%$ error region.

b. Rate dependence

We investigated the effect of changing the conformational exchange rate from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$ by applying the general rate theory to the structures from the previous work as well as to the new structures generated for this manuscript. Figure 4 shows the T_1 (Fig. 4A) and T_2 (Fig. 4B) dependence on the conformation rate for the “90 degree series,” while Figure 5 shows the T_1 (Fig. 5A) and T_2 (Fig. 5B) dependence on the conformation rate for the “30 degree series.” It can be seen that deviations from the slow limit values become apparent for rates of $1 \times 10^7 \text{ s}^{-1}$ and higher, the regime where the rates of rotational diffusion, base libration and conformational exchange become sufficiently close to have a cumulative impact upon the relaxation times. Due to the small amplitude of the base libration, the overlap of rates between tumbling and conformational exchange is likely to be the more significant determinant of the rate of relaxation.

Another important feature of these results is that the changes in relaxation times with increases in conformational exchange rates up to $1 \times 10^7 \text{ s}^{-1}$ occur only for conformational bend angles $\geq 30^\circ$. In other words, the amplitude of the change in the structure of the molecule needs to be substantial in order for the change to be reflected in the relaxation times. Exchange between structures that have similar diffusion tensors will understandably be difficult to discern. It is clear, however, that upon increasing the rate even further, even structures at lower bend angles begin to deviate from the slow limit behavior, implying that the effect of the conformational rate increase can amplify even small differences in diffusion tensors (note, for example, the 20° bend angle structure for rates $\geq 5 \times 10^7 \text{ s}^{-1}$ in Figures 5A and 5B).

In addition, the relaxation times for an exchange rate of $1 \times 10^9 \text{ s}^{-1}$, a rate higher than the base libration and tumbling observed experimentally, are also worth noting. While the trend of decreasing T_1 values and increasing T_2 values continues even up to this point, it will be seen in

the following subsection that this trend is not monotonic for the T_1 times and eventually reverses itself at higher rates.

Concerning the experimental solution relaxation data, there is an interdependence of the values of the two parameters of exchange rate and bend angle that provide the closest fit to the data. Thus, for rates less than or equal to $1 \times 10^7 \text{ s}^{-1}$, the “90 degree series” graphs show reasonable agreement for all bend angles $\leq 50^\circ$. For rates $> 1 \times 10^7 \text{ s}^{-1}$, structures with bend angles $> 30^\circ$ begin to deviate from the experimental values. The “30 degree series” graphs are more restrictive in the fit parameters, where an increase in rate results in a decrease in the upper limit of the bend angles that fit the data. Higher bend angles of between 40° and 50° are inconsistent with the experimental data for almost all rates considered.

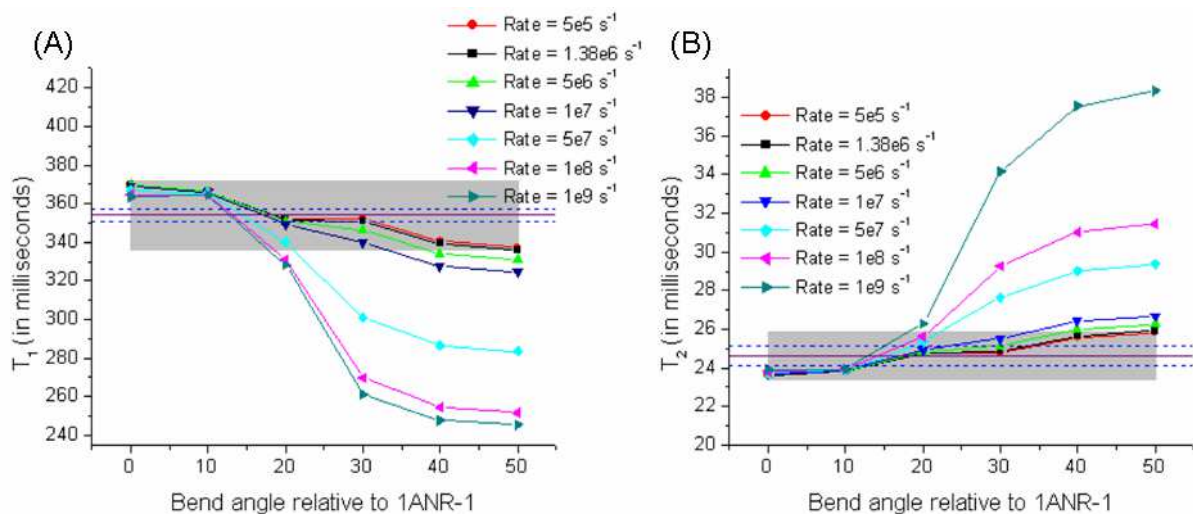


Figure IV.6. Relaxation times vs. bend angle for the “Tw15 series,” considering various exchange rates from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$: (A) T_1 times; and (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3 \text{ ms}$, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5 \text{ ms}$. Also shown in grey is the $\pm 5\%$ error region.

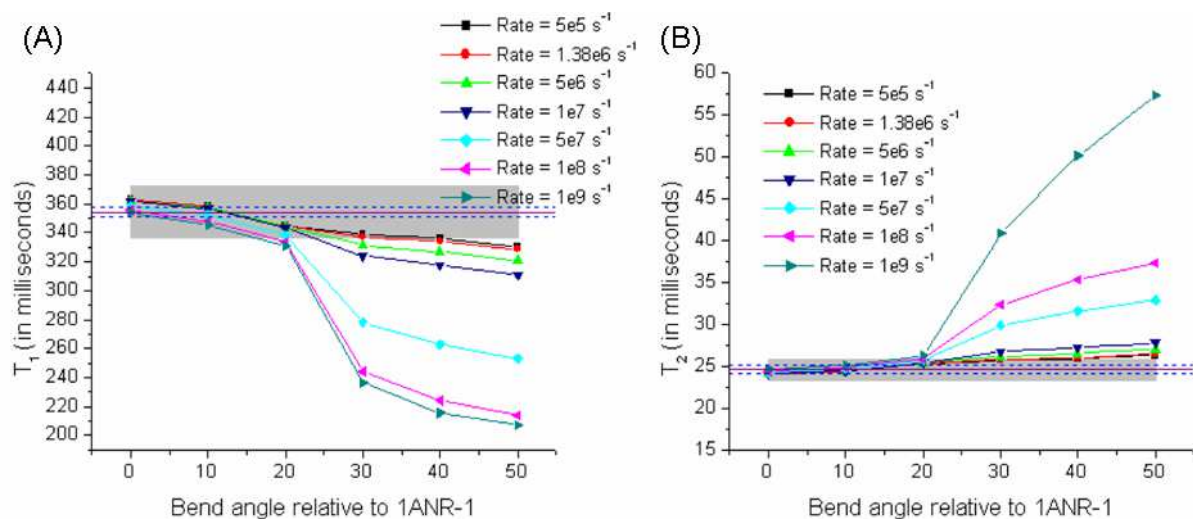


Figure IV.7. Relaxation times vs. bend angle for the “Tw-15 series,” considering various exchange rates from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$: (A) T_1 times; and (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3 \text{ ms}$, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5 \text{ ms}$. Also shown in grey is the $\pm 5\%$ error region.

We repeated the same calculations using the models in Figure 2, whose upper helices are rotated by $\pm 15^\circ$ about their own axes in addition to the bend angles $\leq 50^\circ$. The series of models in Figure 2A will hereafter be referred to as the “Tw15 series,” and those in Figure 2B as the

“Tw-15 series.” The results for both series are shown in Figures 6 (“Tw15 series”) and 7 (“Tw-15 series”).

Before describing the results, it is essential to note that the structure with a 0° bend angle is not the 1ANR-1 structure. In order to consider the effect of applying a twist alone to the upper helix, this initial structure was generated with a +15° or -15° twist, and subsequently averaged with 1ANR-1. Moreover, the 1ANR-1 model used in this manuscript was rotated relative to the 1ANR-1 used in the previous work, as a result of the different characterization of the helical axes with 3DNA in our current work, and with having to align the lower helix with the z-axis. Ideally, this overall rotation should not affect the diffusion eigenvalues. However, HYDRONMR, the program used to calculate the eigenvalues, outputs two slightly different sets of diffusion eigenvalues, with the result that the relaxation times changed slightly (on the order of 5 ms for T_1 and 0.3 ms for T_2). We surmise that this difference is due to discrepancies in the extrapolation of the diffusion eigenvalues towards zero bead size and to rounding-off errors. This discrepancy was lessened by using many more points in the extrapolation process, but, in order to match the results of the previous work, the same lower number of points (6 extrapolation points) was used here as well.

Figure 6A reports the simulated T_1 values for the “Tw15 series,” that fit the experimental data best for bend angles between 10° and 30°. Bend angles between 10° and 30° fit the data for exchange rates $< 1 \times 10^7 \text{ s}^{-1}$, while for higher rates the best-fit range is confined to angles between 10° and 20°. The simulated T_2 values in Figure 6B accord with the data for bend angles $< 30^\circ$ for rates $\leq 1 \times 10^7 \text{ s}^{-1}$. Higher rates fit the experimental data only for bend angles $\leq 20^\circ$. As mentioned before, due to uncertainties in the calculation of the diffusion eigenvalues, these ranges may have to be altered slightly, but there does seem to be a trend of lower rates increasing the range of

bend angles that match the data. Moreover, the simulated relaxation times seem to better fit the experimental values for lower bend angles in general. Considering the results for the “Tw-15 series” in Figures 7A and 7B, for rates $\leq 1 \times 10^7 \text{ s}^{-1}$, models with bend angles between 10° and 20° fit the data better, while for higher rates the best-fit bend angles are limited to $\leq 10^\circ$.

In an attempt to separate the various combinations of parameters that fit the experimental data, we must utilize the findings of other experiments. Since the models considered here are based on solid-state NMR data, we can check against the best-fit parameters mentioned in Section 2D. Ignoring for a moment the twist of the helix, Figures 4 and 5 show that the T_1 values for the solid-state bend and rate parameters are slightly outside of the experimental error bars, while they are within the error bars for the T_2 times. Given that there may be round-off errors in the evaluation of diffusion tensors, the discrepancy in the T_1 values may not be significant.

Considering the twist angles as well, for the “Tw15 series” Figure 6 shows that the simulated T_1 is outside the error bars once again, and so are the T_2 ’s, although the discrepancy for the latter is fairly small. However, a much better fit is obtained for the “Tw-15 series” (Figure 7) using the solid-state parameters, with the T_1 value being just outside the experimental errors bars and the T_2 value well within them. While this may be construed as a means of breaking the degeneracy of the solid-state results with respect to the direction of twist of the upper helix, we reiterate that there are errors due to steric clashes in the models used and uncertainties in the other free parameters in the problem. The tentative conclusion is that it is plausible that the solid-state parameters continue to be valid under solution sample conditions: previous studies of the RNA as a function of hydration have shown that relaxation times remain fairly constant above a hydration of $W = 16$.⁸⁵ Again, we have not repeated our previous studies of the impact of the

choice of the hydrodynamic bead radius (or, atomic element radius AER) on the simulations, and a consideration of the appropriate value of this parameter also seems essential.

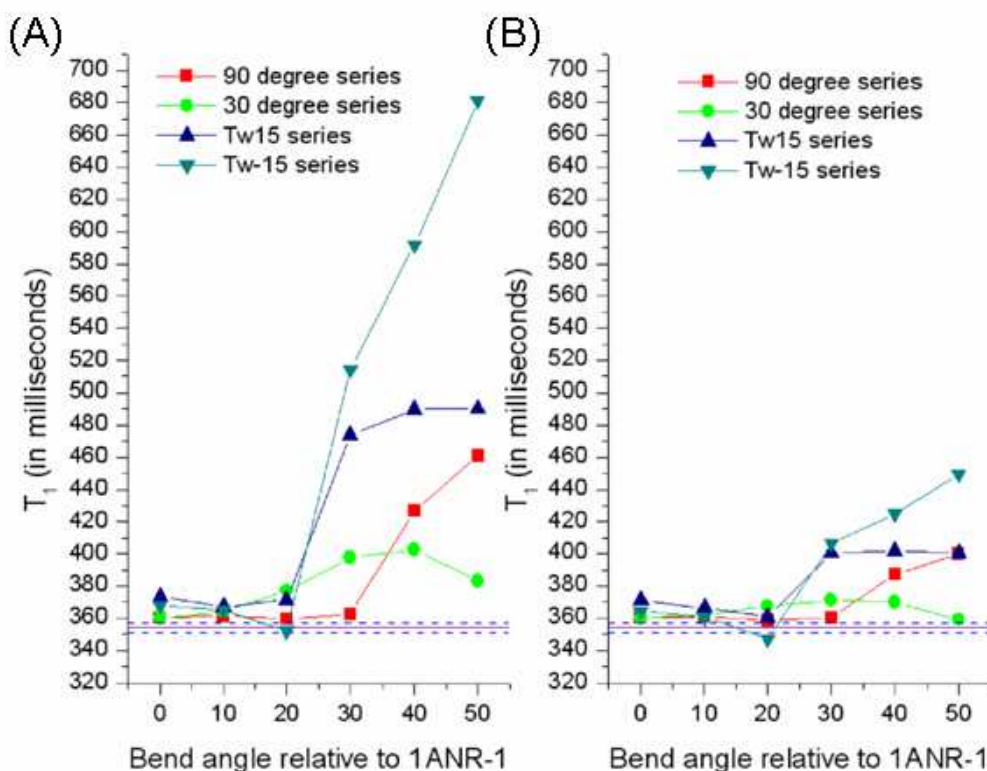


Figure IV.8. T_1 vs. bend angle for the four sets of structures for an exchange rate of: (A) $1 \times 10^{10} \text{ s}^{-1}$ and (B) $5 \times 10^9 \text{ s}^{-1}$. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation time is $T_1^{\text{Expt}} = 354.1 \pm 3.2 \text{ ms}$. The $\pm 5\%$ error region has been left out for clarity.

c. T_1 turnover

Having observed the changes in relaxation times as a function of the conformational exchange rate up to a value $1 \times 10^9 \text{ s}^{-1}$, we examined if the trend continued indefinitely to even higher rates of conformational exchange. To that end, we simulated relaxation times for all the four sets of structures used in Figures 4 – 7, with an exchange rate of $1 \times 10^{10} \text{ s}^{-1}$. The T_1 graph is shown in Figure 8A. Here the dependence on the bend angle has been completely inverted, with

an increase in T_1 for increasing bend angles. The T_1 values have swung over to much higher values, signaling that there is a turnover point in the range between $1 \times 10^9 \text{ s}^{-1}$ and $1 \times 10^{10} \text{ s}^{-1}$, where the T_1 times go through a minimum and reverse their dependence on exchange rate. To localize the minimum further, we also calculated the T_1 for a rate of $5 \times 10^9 \text{ s}^{-1}$. The results are shown in Figure 8B. As the graphs seem to have already crossed the turnover point from a downward trend to an upward one, the T_1 minimum is seen to be at slightly less than $5 \times 10^9 \text{ s}^{-1}$.

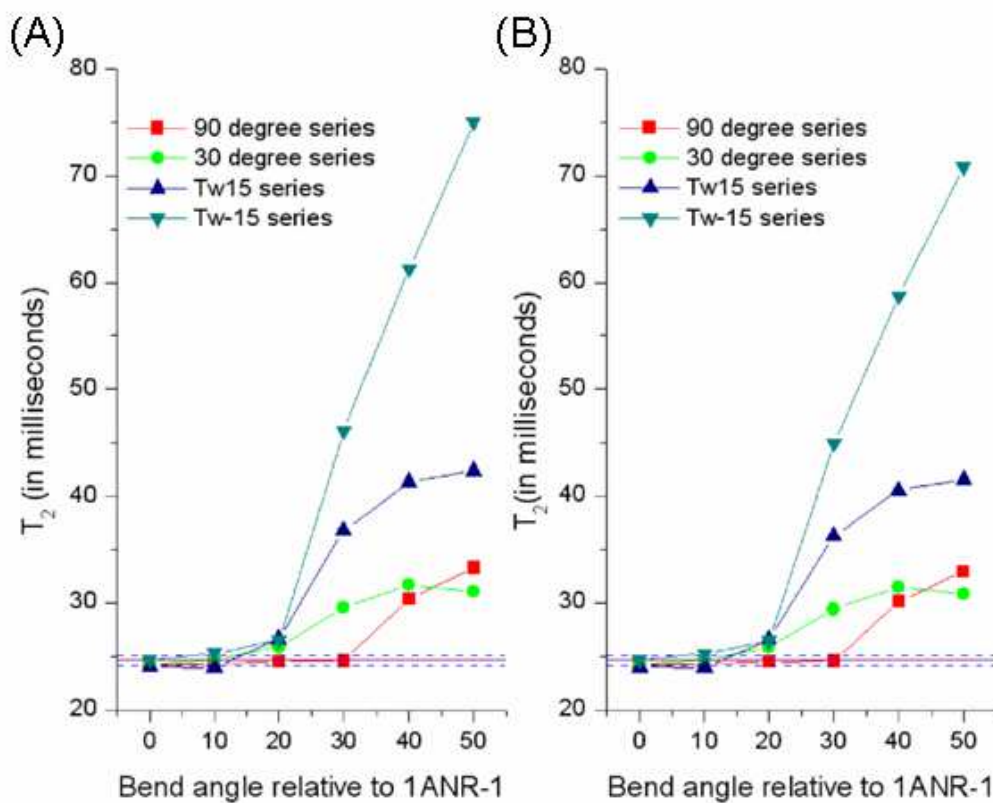


Figure IV.9. T_2 vs. bend angle for the four sets of structures for an exchange rate of: (A) $1 \times 10^{10} \text{ s}^{-1}$ and (B) $5 \times 10^9 \text{ s}^{-1}$. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation time is $T_2^{\text{Expt}} = 24.6 \pm 0.5 \text{ ms}$. The $\pm 5\%$ error region has been left out for clarity.

The T_2 graphs for rates of $1 \times 10^{10} \text{ s}^{-1}$ and $5 \times 10^9 \text{ s}^{-1}$ are shown in Figures 9A and 9B respectively.

Figure 9A shows that, unlike the T_1 values, the T_2 times have not changed their trend. A

comparison of the times at the two rates shows, however, that the difference in times is fairly

small. Upon testing the values of T_2 for the 50° models (which showed the most dramatic changes), we observe that an increase of the rate up to $1 \times 10^{12} \text{ s}^{-1}$ (data not shown) caused only small changes in the relaxation times relative to the values at $5 \times 10^9 \text{ s}^{-1}$. The T_2 times seem to have reached their fast time limits by the time the T_1 values turn around.

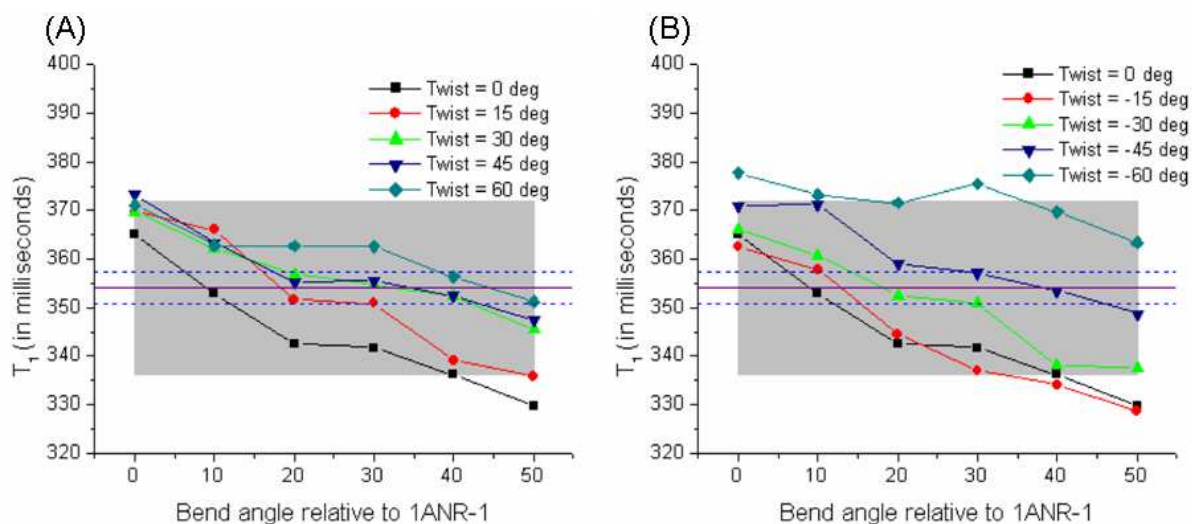


Figure IV.10. T_1 vs. bend angle, considering various degrees of twist about the upper helix, ranging from -60° to $+60^\circ$: (A) positive twist angles: $0^\circ \leq \text{Twist} \leq 60^\circ$; and (B) negative twist angles: $-60^\circ \leq \text{Twist} \leq 0^\circ$. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation time is $T_1^{\text{Expt.}} = 354.1 \pm 3.2 \text{ ms}$. Also shown in grey is the $\pm 5\%$ error region.

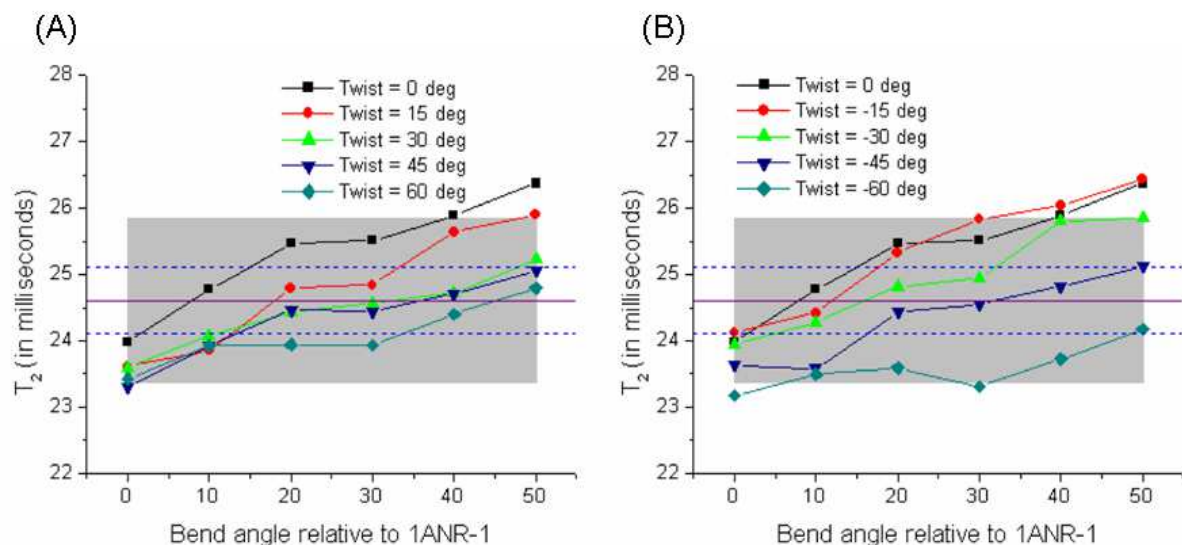


Figure IV.11. T_2 vs. bend angle, considering various degrees of twist about the upper helix, ranging from -60° to $+60^\circ$: (A) positive twist angles: $0^\circ \leq \text{Twist} \leq 60^\circ$; and (B) negative twist angles: $-60^\circ \leq \text{Twist} \leq 0^\circ$. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation time is $T_2^{\text{Expt.}} = 24.6 \pm 0.5$ ms. Also shown in grey is the $\pm 5\%$ error region.

d. Twist angle dependence

The “Tw15 series” and the “Tw-15 series” shown in Figure 2 represented two sets in a series of models that were constructed with various degrees of twist applied to the upper helix relative to its orientation in 1ANR-1. The twist angles that were applied range from -60° to $+60^\circ$ in 15° increments. Again, the initial structure is not the 1ANR-1 structure, but one with the upper helix rotated through the twist angle associated with that particular series, and all models, including this 0° bend angle model, are averaged with the 1ANR-1 structure to produce the results that follow.

Relaxation times have been calculated for the twist angle models using the solid-state exchange rates of $1.38 \times 10^6 \text{ s}^{-1}$. For clarity, the data for positive twist angles and negative twist angles have been separated into distinct graphs. The T_1 relaxation times are shown in Figure 10, with the positive twist angle data in Figure 10A and that for negative twists in Figure 10B. With few exceptions, changes in the twist angle relative to the 1ANR-1 are directly proportional to

changes in the T_1 time. Correspondingly, there is an increase in the best-fit bend angles with an increase in twist angle. For example, the 0° twist angle structures match the experimental data for bend angles less than 20° . However, by changing the twist in the upper helix to 45° , the range of best-fit bend angles shifts to $20^\circ < \theta_{\text{best-fit}} < 50^\circ$. Figure 11A and 11B show the corresponding T_2 simulations. There is an inverse relationship between twist angles and T_2 values, barring some instances which do not follow this simple trend. As with the T_1 values, this leads to the best-fit twist and bend angles being correlated.

The physical origin of this correlation must be understood by considering changes in the diffusion tensor and the orientation of the local C-D bond relative to the diffusion tensor frame. Since the helices of the molecule are not perfectly axially-symmetric, it is possible that changes in the diffusion tensor eigenvalues and bond orientation relative to the tensor eigenvectors arising from a twist in the upper helix for a given bend angle mimic those for a different set of twist and bend angles. Alternately, it is also possible that the relationship between the diffusion tensor eigenvalues and the bond orientation on the one hand, and the ensuing relaxation times on the other is simply not one-to-one. The relaxation times T_1 and T_2 are merely two numbers that are obtained analytically as a combination of several parameters, and what we observe in Figures 10 and 11 may be a manifestation of the degeneracy of the parametric landscape with respect to the relaxation times.

Using the solid-state values for the parameters, it can be seen from Figures 10A and 11A that the structure with a positive twist of 15° does not fit the T_1 experimental value, but almost fits the T_2 value. On the other hand, if we consider the -15° twist structure in Figures 10B and 11B, we see that the experimental T_1 value is very close to being fit and T_2 value fits well within the error bars. Since the solid-state experiments are unable to distinguish the sense of rotation,

solution relaxation times may help to break this degeneracy, if the values of other parameters are determined independently.

4. Discussion

The theoretical method and results presented in this manuscript aim to introduce a new method of applying molecular motional models with atomic-level detail derived from other techniques to solution NMR relaxation results, and include the possibility of motions that alter the diffusion tensor. We also have extended the TAR-specific results presented in the previous work,⁷⁵ by removing the key assumption of an infinitely slow exchange between conformers. With regards to the first goal, the applicability of the current approach extends beyond the solid-state/solution NMR application described here, and can be expanded to corroborate any type of motion inferred from other experiments, provided molecular structures are available. Moreover, it may be possible to simulate further time-dependent, diffusion-averaged solution-state observables that can be described in terms of correlation functions of Wigner rotation matrices, in addition to relaxation rates, simply by utilizing the transition probability (Equation 17) in the expression analogous to Equation 4.

It is worth restating the assumption, mentioned in the OM, that the exchange process considered here is such that there is no change in the orientation of the molecules during the instantaneous jump between conformers, i.e. the diffusion tensors of any pair of conformers are momentarily collinear during the jump. This assumption allows for the derivation of a relatively simple closed-form expression for the correlation function and is valid as long as the different exchanging conformers do not vary significantly. This condition of validity can be achieved in applications by considering a larger number of conformers along the trajectory between two very

disparate conformers, thereby better approximating a continuum transition. For a derivation of spectral densities for the general case of non-collinear diffusion tensors at the moment of exchange, the reader may refer to the work of Ryabov et al.¹¹⁰

In our recent work,⁷⁵ we referred to the ssNMR-based models for the residues U23 and U25 described in a previous article,¹⁰ models which include local motions that were close to the overall tumbling timescale. While an inclusion of simulations of these two bases would support the generality of the theory presented here, these two residues occur in the trinucleotide bulge region of TAR-RNA and any description of their dynamics would necessitate an understanding of changes in the bulge configuration. As mentioned in Section 3, the bulge is assumed to be rigid in our current structure manipulations. In order to reasonably simulate such motions, we would need to obtain a set of atomic-level orientations of the bulge residues and the concurrent helix configurations. Such studies require new approaches, one of which is currently being evaluated.

Along similar lines, we emphasize the requirement of detailed atomic-level structures. In the current manuscript we have used 1ANR-1 as the starting structure for all of our simulations. However, the use of any one of the 20 lowest energy structures reported under the PDB code 1ANR⁸ would result in significantly different constraints on parameters. We evaluated the impact of a different choice of initial structure by calculating the relaxation times for the first five lowest energy structures (1ANR 1-5) tumbling as single conformers, without internal motion. In the original reference⁸ energy is quantified by the energy of NOE constraint violations, and these first few structures do not differ greatly in energy. The maximum difference in T_1 was approximately 80 ms and in T_2 was about 5 ms (both between the results for 1ANR-2 and 1ANR-4). The discrepancies arise from the fact that the structures 1ANR 1-20 show

significant variations in relative helix placement, and relative base orientations, even within base pairs. These substantial differences motivate searches for well-established energy-minimum structures, especially given that computational capacity has improved vastly since the original publication of the structures 15 years ago. However, the methodology and the qualitative arguments made herein still hold for any given choice of initial structure, and given recent trends towards more accurate molecular structures this contributor to uncertainty is likely to be removed in the near future.

The work presented here includes a re-derivation of the results of Favro,⁸³ Huntress¹⁰³ and Freed¹⁰⁴ and attempts to incorporate the discrete-jump formalism described by Wong, Case and Szabo.⁸² We have applied this formalism to a molecule, HIV-I TAR RNA, where many of the structures deviate significantly from cylindrical idealizations. This application of the formalism, the second central objective of the current manuscript, has shown that it is possible to fit the solution experimental data given reasonable choices of the available parameters. We already considered the impact of changing the conformational bend angle, the atomic element radius (AER, representing the size of the hydrodynamic beads that are used to represent each atom) and the direction of helical bending in previous work. Additional parameters considered here are the rate of conformational exchange, and the twist angle about the upper helix.

Although discussed previously,⁷⁵ it bears reminding that the ssNMR models¹⁰ used as a basis for the simulations in this manuscript report only on the upper helix motions. No site labels in the lower helix were considered at the time so that information on the relative motions of the two helical domains could not be acquired. This motivated us to explore a wider range of inter-helical bend and twist angles than suggested by ssNMR models alone, to account for possible lower helical motions.

Based on our new analysis, it is worthwhile to revisit the comparison with the work of Zhang, Al-Hashimi and co-workers^{12,14} that were made previously.⁷⁵ An order tensor analysis in Zhang et al 2007¹⁴ reported an average upper helix bend of 25° relative to an elongated lower helix and an average lower helix bend of 54° relative to an elongated upper helix, with the difference between the two results being ascribed to differing twist motions of the two helices. As described before, ssNMR models¹⁰ corroborate these collective twisting and bending motions. The quoted results for the average orientations in Zhang et al 2007 are accompanied by a significantly lowered level of order for both helices, indicating very large amplitude helix motions (see also Zhang et al 2008⁸¹). A direct comparison of our results with the above average orientations is not possible as the order tensor elements probe the average inter-helical bend, while the bend angles considered herein are measured with respect to the lowest energy structure 1ANR-1. As an initial attempt at comparing the two sets of angles, we calculated the approximate symmetry axes of the upper and lower helices using 3DNA¹⁰⁶ and found an inter-helical bend of 75° for 1ANR-1. Given the equal populations of the two conformers considered here, this would imply a total helical bend relative to the 1ANR-1 structure of about 100° for an average of 25° (i.e. an amplitude of $2 \times (75^\circ - \text{Average angle})$) and about 40° for an average angle of 54°. We did not consider bend angles as large in magnitude as 100°, but it is conceivable that some combinations of the parameters, including the exchange rate, could allow this value to be fit. The amplitude of 40° can be fit to the experimental data if we assume the ssNMR exchange rate of $1.38 \times 10^6 \text{ s}^{-1}$ and twist angles of $30^\circ \leq \theta_{\text{twist}} \leq 60^\circ$ and θ_{twist} around -45°. Other AERs and directions of helical reorientation might also allow matches between the two sets of results.

Zhang et al 2007 also reports the results of a three-conformer ensemble study suggesting an overall helical bend of 94° , an upper helical twist of 110° and a lower helical twist of 53° . While we did not extend our analysis to bend angles of that magnitude, nor to upper helical twists greater in magnitude than 60° , it is conceivable, based on the trends in Figures 10 and 11 that the Zhang et al 2007¹⁴ values could fit our solution data.

However, the time scale of domain motions quoted in Zhang et al 2006¹² is 1.5-1.9 ns (rate $\sim 5\text{-}7 \times 10^8 \text{ s}^{-1}$) and that differs significantly from the ssNMR time scale of 725 ns (rate = $1.38 \times 10^6 \text{ s}^{-1}$). Figures 4 - 7 of the current manuscript show that using the higher rates from Zhang et al 2006¹² would preclude bend angles greater than about 30° for any of the directions of helical reorientation considered here. Thus, given particular choices of the parameters, the solution results may be reconciled with each other, although the discrepancy of the time scale in Zhang et al 2006¹² with the ssNMR time scale still remains. However, given the fact that the Zhang et al 2006 result was produced by a Model-free analysis, the rate of motions captured is necessarily faster than the overall tumbling rate, thereby precluding the inclusion of any slower conformational motions on the timescales considered here. Indeed, the RDC-based analysis in Zhang et al 2007 confirms the existence of larger domain motions (leading to a lower degree of order) potentially occurring on timescales up to milliseconds (to which an RDC would be sensitive). The ssNMR results are also able to capture motions on longer timescales, and previously¹⁰ a model-dependent approach was used to pin-point a particular exchange rate. Thus, the discrepancy mentioned above may be attributed to the greater sensitivity of the ssNMR approach to slower timescales of motion.

Multiple sets of parameters can produce the same values of the relaxation times: this degeneracy must be broken by an independent experimentally-driven assessment of the

parameters. For example, we may take the solid-state best-fit parameters as being relevant even under solution conditions, especially given that hydration-based studies show a relatively flat hydration dependence of the relaxation times at high hydration levels.⁸⁵ In this case, we can fix the conformational bend angle to 9°, the twist angle to 15° and the conformational exchange rate at $1.38 \times 10^6 \text{ s}^{-1}$. The evaluation of goodness-of-fit of the ensuing solution simulations has been made in Section 3. This leaves the AER and direction of helical bending as free parameters to be optimized. Our current value of an AER of 2.3 Å provides a tumbling correlation time $\tau_C^{rot} = (6D)^{-1}$ (where D is the trace of the rotational diffusion tensor) of 6.4 ns for the 1ANR-1 conformer alone, which accords well with the Model-free-derived value of 5.9 ns for free TAR as reported in Bardaro et al.¹¹ While the experimental time is averaged over all conformational states accessed by free TAR, the remarkable similarity in the time scale provides some independent corroboration of the choice of AER. Furthermore, detailed hydration studies of nucleic acids may provide more independent estimates of the impact of various hydration patterns, if we can solve for diffusion tensors using explicit water molecules. The direction of helical bending is physically determined by the backbone torsion degrees of freedom and the inter-residue interactions (including steric clashes), and these properties may be studied using molecular dynamics simulations and energy minimization studies. Since the bending of the upper helix is associated with changes in the bulge configuration, we have begun investigating the possibility of simulating both helix bending and the bulge residue orientations using the aforementioned techniques.

5. Results based on corrected equations

We have repeated several of the simulations for TAR-RNA (Figure 1) using the corrected equations for the time-dependent coefficients (Eqs. 31b and 32) and the correlation function (Eq. 37) and present these results in the following. The relaxation times are calculated in each case for an exchange between a structure modified in one or more parameters from the lowest energy model of TAR-RNA (PDB code 1ANR)⁸, labeled 1ANR-1, and 1ANR-1 itself. For comparison, the experimental relaxation times are $T_1 = 354 \pm 3$ ms and $T_2 = 24.6 \pm 0.5$ ms.¹¹

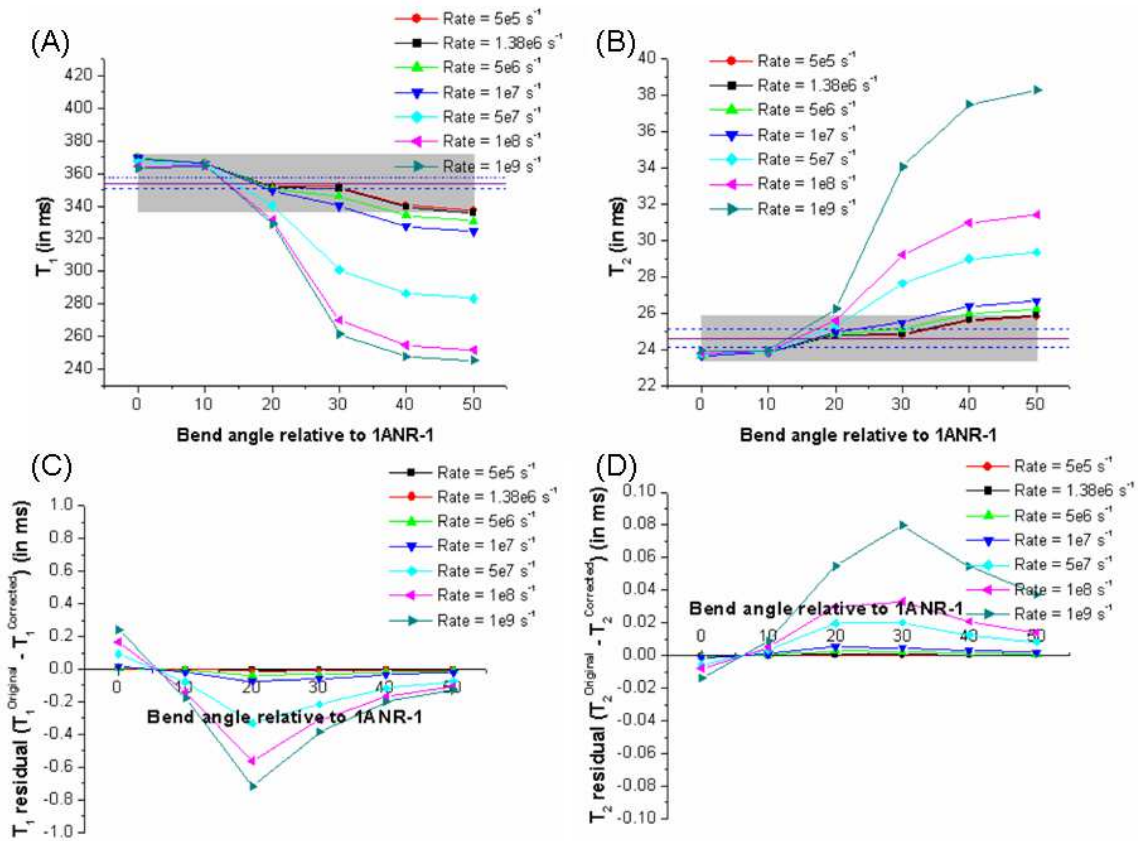


Figure IV.12. Relaxation times vs. bend angle for structures within the Tw15 series considering various exchange rates from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$: (A) T_1 times; (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Exptl.}} = 354 \pm 3$ ms, and $T_2^{\text{Exptl.}} = 24.6 \pm 0.5$ ms. Also, shown in grey is the $\pm 5\%$ error region. The two lower panels show the differences between the relaxation times calculated using the original theory and those calculated using the corrected expressions: (C) $T_1^{\text{Original}} - T_1^{\text{Correction}}$ and (D) $T_2^{\text{Original}} - T_2^{\text{Correction}}$.

In Figure 12, we have recalculated the solution relaxation times T_1 (Fig. 12A) and T_2 (Fig. 12B), for the set of structures termed as the “Tw15 series” (structures shown in Figure 2A), as a function of the conformational exchange rate. The rate is varied from $5 \times 10^5 \text{ s}^{-1}$ to $1 \times 10^9 \text{ s}^{-1}$. Shown in the lower two panels are the “residuals” i.e. the differences between the relaxation times published in the original article using the uncorrected equations and the new simulation results, $T_1^{\text{Original}} - T_1^{\text{Correction}}$ (Fig. 12C) and $T_2^{\text{Original}} - T_2^{\text{Correction}}$ (Fig. 12D). As can be seen, the difference in results is less than 0.8 ms in magnitude for the T_1 times and less than 0.08 ms in magnitude for the T_2 times, significantly smaller than the respective experimental errors. (Compare to Fig. 6.)

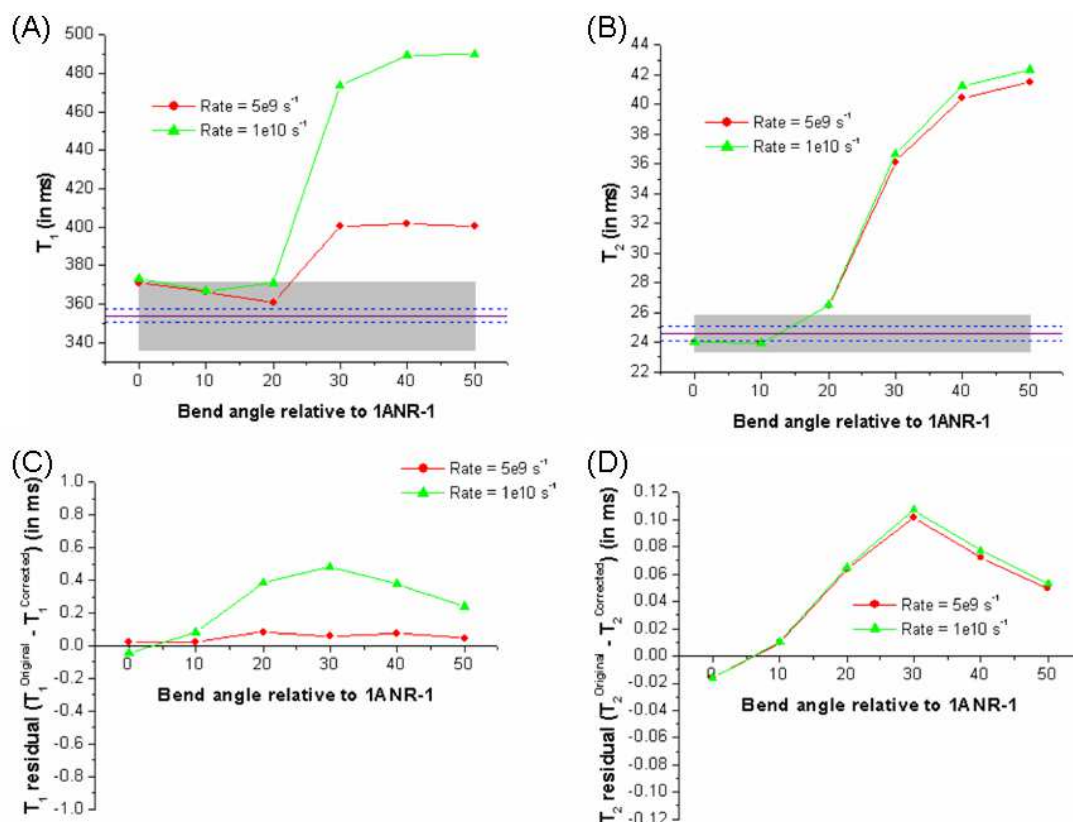


Figure IV.13. Relaxation times vs. bend angle for structures within the Tw15 series considering the two exchange rates of $5 \times 10^9 \text{ s}^{-1}$ and $1 \times 10^{10} \text{ s}^{-1}$: (A) T_1 times; (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3 \text{ ms}$, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5 \text{ ms}$. Also, shown in grey is the $\pm 5\%$ error region. The two lower panels show the differences between the relaxation times calculated using the original theory and those calculated using the corrected expressions: (C) $T_1^{\text{Original}} - T_1^{\text{Correction}}$ and (D) $T_2^{\text{Original}} - T_2^{\text{Correction}}$.

Figure 13 shows the relaxation times calculated for the Tw15 series with exchange rates of $5 \times 10^9 \text{ s}^{-1}$ and $1 \times 10^{10} \text{ s}^{-1}$. These have been shown separately, as in the uncorrected results, to show clearly the turnover in the T_1 trend as the rates increase. Whereas the T_1 values for rates considered in Fig. 12 steadily decrease with increases in the rate for the same bend angle, there is a reversal of this trend at a rate between $1 \times 10^9 \text{ s}^{-1}$ and $5 \times 10^9 \text{ s}^{-1}$, as can be seen from the increasing T_1 values for the rates in Fig. 13A. The T_2 values (Fig. 13B) do not show this turnover. Also shown are the T_1 residuals (Fig. 13C) and T_2 residuals (Fig. 13D). The magnitude of the residuals is once again seen to be smaller than the experimental errors (at most $\sim 0.5 \text{ ms}$ for

T_1 , and ~ 0.1 ms for T_2). This is strong confirmation of the earlier statement that the qualitative discussion in the previous Results section still remains valid, as these graphs depict simulations using very fast rates, and if any significant numerical discrepancy should have existed in the simulations, it would have manifested itself in these results. (Compare to Figs. 8 and 9.)

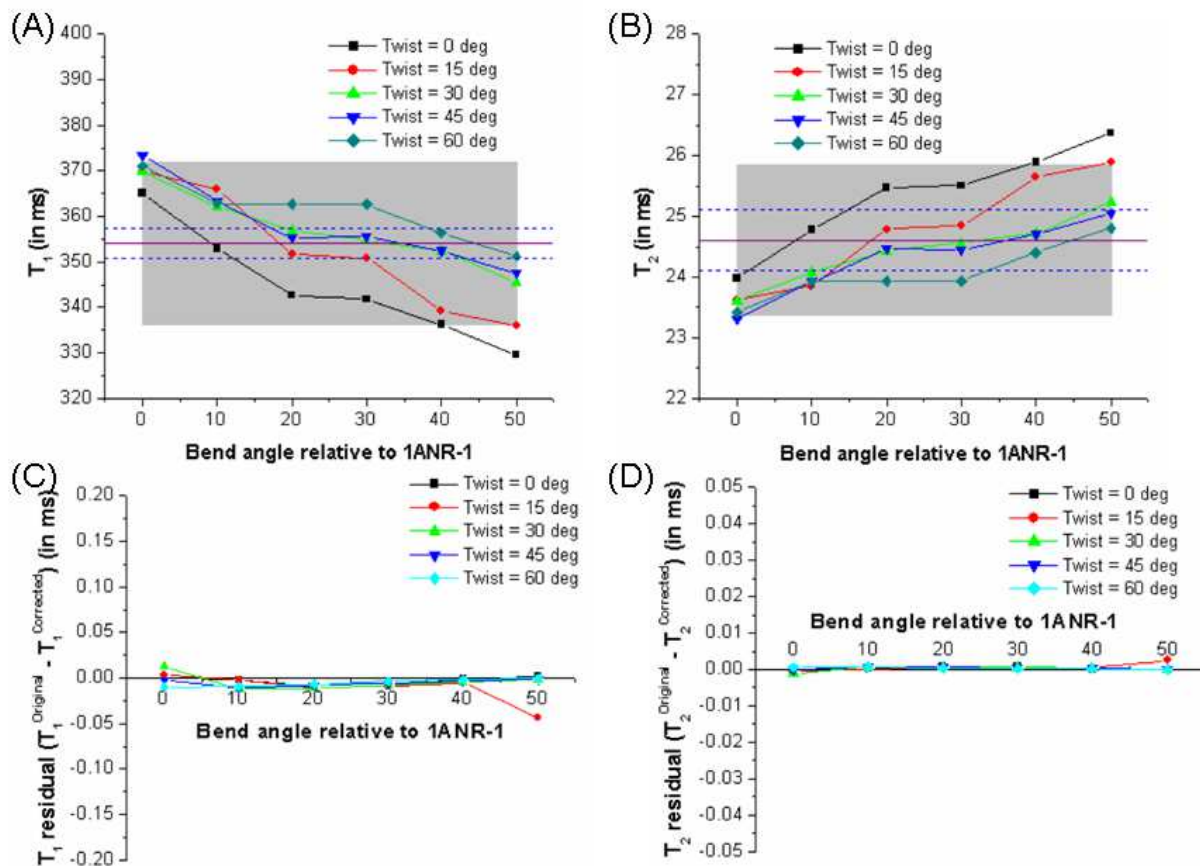


Figure IV.14. Relaxation times vs. bend angle for structures with various degrees of positive twist about the upper helix, ranging from 0° to $+60^\circ$: (A) T_1 times; (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Exptl}} = 354 \pm 3$ ms, and $T_2^{\text{Exptl}} = 24.6 \pm 0.5$ ms. Also, shown in grey is the $\pm 5\%$ error region. The two lower panels show the differences between the relaxation times calculated using the original theory and those calculated using the corrected expressions: (C) $T_1^{\text{Original}} - T_1^{\text{Correction}}$ and (D) $T_2^{\text{Original}} - T_2^{\text{Correction}}$.

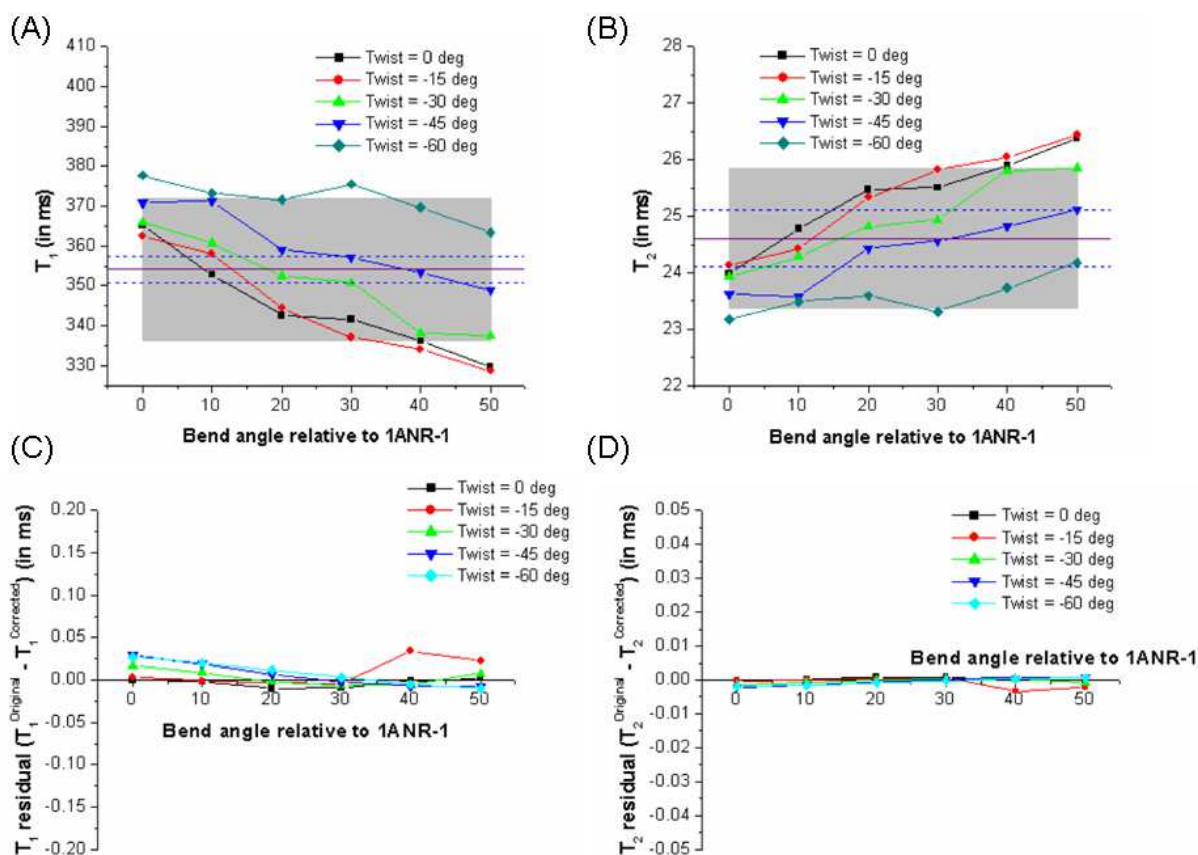


Figure IV.15. Relaxation times vs. bend angle for structures with various degrees of negative twist about the upper helix, ranging from -60° to 0° : (A) T_1 times; (B) T_2 times. Experimental values are marked by solid purple lines, and the error bars are indicated by dashed blue lines. The experimental relaxation times are: $T_1^{\text{Expt.}} = 354 \pm 3$ ms, and $T_2^{\text{Expt.}} = 24.6 \pm 0.5$ ms. Also, shown in grey is the $\pm 5\%$ error region. The two lower panels show the differences between the relaxation times calculated using the original theory and those calculated using the corrected expressions: (C) $T_1^{\text{Original}} - T_1^{\text{Correction}}$ and (D) $T_2^{\text{Original}} - T_2^{\text{Correction}}$.

Figures 14 and 15 show the T_1 (Figs. 14A and Fig. 15A) and T_2 (Figs. 14B and Fig. 15B) values for structures whose upper helices have been modified from 1ANR-1 to include a specified twist about the upper helical axis, in addition to the bend angles indicated on the x-axes. The upper helical axis is defined using the program 3DNA.¹⁰⁶ Figure 14 shows the relaxation times for positive twists (in a right-hand rule sense), along with the T_1 residuals (Fig. 14C) and the T_2 residuals (Fig. 14D). The panels in Figure 15 show analogous results for negative twists applied to the upper helix. The residuals are well below the experimental errors in

all cases shown, being at most ~ 0.03 ms for T_1 and ~ 0.003 ms for T_2 . (Compare to Figs. 10 and 11.)

Although we have not redone the simulations for all the various cases considered in the preceding figures (Figures 3 through 11), we believe that the range of parameters considered here do show sufficiently that numerically the original theory and the corrected expressions do not vary significantly enough to alter the qualitative arguments presented above. However, it is important to note that such may not be the case for a set of structures that may deviate substantially from cylindricity, or for an exchange between structures that may have very different diffusion tensors.

6. Conclusions

In this manuscript we have presented a method of testing proposed motional models of macromolecules in solution conditions by simulating solution NMR T_1 and T_2 relaxation times. Specifically, we have made use of available solid-state NMR-derived models to determine the types of motions experienced by the upper helix of the HIV-1 TAR RNA molecule. In addition to showing that the solid-state NMR (at a hydration of 16 waters per nucleotide) dynamic parameters provide reasonable fits to the solution relaxation data, we have also made an exploration of the dependence of the relaxation times on variations in the dynamic parameters. For future studies on this system or on RNA in general, the analyses conducted herein can serve as a means of reducing the large motional parameter space, and also provide qualitative insight regarding trends in parametric correlations. We believe that the method is generally applicable to a variety of macromolecules and can be extended easily to include new observables, thereby bridging gaps between multiple experimental techniques.

Appendix IV.A: Sample Mathematica code for the general rate theory

//Function definitions used for the general rate theory as described in Table IV.1

```
sig[{Dx_, Dy_, Dz_}] = Dx^2 + Dy^2 + Dz^2 - (Dx*Dy + Dy*Dz + Dz*Dx);
```

//Numerators and denominators of the diffusion tensor-dependent eigenvectors

```
num2[{Dx_, Dy_, Dz_}] = 0.5*((Dx + Dy - 2*Dz) + 2*Sqrt[sig[{Dx, Dy, Dz}]]);
```

```
num3[{Dx_, Dy_, Dz_}] = 0.5*((Dx + Dy - 2*Dz) - 2*Sqrt[sig[{Dx, Dy, Dz}]]);
```

```
den2[{Dx_, Dy_, Dz_}] = Sqrt[2*sig[{Dx, Dy, Dz}]] + Sqrt[sig[{Dx, Dy, Dz}]]*(Dx + Dy - 2*Dz);
```

```
den3[{Dx_, Dy_, Dz_}] = Sqrt[2*sig[{Dx, Dy, Dz}]] - Sqrt[sig[{Dx, Dy, Dz}]]*(Dx + Dy - 2*Dz);
```

//Matrix of eigenvectors

```
fla[{Dx_, Dy_, Dz_}] = {{-(2^(-0.5)), 0, 0, 0, 2^(-0.5)}, {0.25*Sqrt[6]*(Dx - Dy)/den2[{Dx, Dy, Dz}], 0, num2[{Dx, Dy, Dz}]/den2[{Dx, Dy, Dz}], 0, 0.25*Sqrt[6]*(Dx - Dy)/den2[{Dx, Dy, Dz}]}, {0.25*Sqrt[6]*(Dx - Dy)/den3[{Dx, Dy, Dz}], 0, num3[{Dx, Dy, Dz}]/den3[{Dx, Dy, Dz}], 0, 0.25*Sqrt[6]*(Dx - Dy)/den3[{Dx, Dy, Dz}]}, {0, 2^(-0.5), 0, 2^(-0.5), 0}, {0, -(2^(-0.5)), 0, 2^(-0.5), 0}, {2^(-0.5), 0, 2^(-0.5), 0, 0}};
```

//Vector of eigenvalues

```
g[{Dx_, Dy_, Dz_}] = {4*Dz + Dx + Dy, 2*(Dx + Dy + Dz) + 2*Sqrt[sig[{Dx, Dy, Dz}]], 2*(Dx + Dy + Dz) - 2*Sqrt[sig[{Dx, Dy, Dz}]], Dz + 4*Dx + Dy, Dz + Dx + 4*Dy};
```

//Spherical angle-dependent multipliers of the time-dependent coefficients in expression for the correlation function, Equation IV.37

```
coeff1[{thetaf_, phif_}, {thetai_, phii_}] = 0.75*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Sin[2*phif])*(Sin[2*phii]);
```

```
coeff4[{thetaf_, phif_}, {thetai_, phii_}] = 0.75*(Sin[2*thetaf])*(Sin[2*thetai])*(Sin[phif])*(Sin[phii]);
```

```
coeff5[{thetaf_, phif_}, {thetai_, phii_}] = 0.75*(Sin[2*thetaf])*(Sin[2*thetai])*(Cos[phif])*(Cos[phii]);
```

```
coeff2[{thetaf_, phif_}, {thetai_, phii_}, {Dxb_, Dyb_, Dzb_}, {Dxa_, Dya_, Dza_}] = 1.5*fla[{Dxb, Dyb, Dzb}][[2, 1]]*fla[{Dxa, Dya, Dza}][[2, 1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*fla[{Dxb, Dyb, Dzb}][[2, 1]]*fla[{Dxa, Dya, Dza}][[2, 3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) + (0.5*Sqrt[1.5])*fla[{Dxa, Dya, Dza}][[2, 1]]*fla[{Dxb, Dyb, Dzb}][[2, 3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) - 1) + 0.25*fla[{Dxb, Dyb, Dzb}][[2, 3]]*fla[{Dxa, Dya, Dza}][[2, 3]]*(3*(Cos[thetaf]^2) - 1)*(3*(Cos[thetai]^2) - 1);
```

```

c32[{thetaf_, phif_},{thetai_, phii_},{Dxb_, Dyb_, Dzb_}, {Dxa_, Dya_, Dza_}] =
1.5*fla[{Dxb, Dyb, Dzb}][[3, 1]]*fla[{Dxa, Dya, Dza}][[2,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*fla[{Dxb,
Dyb, Dzb}][[3, 1]]*fla[{Dxa, Dya, Dza}][[2,
3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) + (0.5*Sqrt[1.5])*fla[{Dxa, Dya,
Dza}][[2, 1]]*fla[{Dxb, Dyb, Dzb}][[3, 3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) -
1) + 0.25*fla[{Dxb, Dyb, Dzb}][[3, 3]]*fla[{Dxa, Dya, Dza}][[2, 3]]*(3*(Cos[thetaf]^2) -
1)*(3*(Cos[thetai]^2) - 1);

```

```

c23[{thetaf_, phif_},{thetai_, phii_},{Dxb_, Dyb_, Dzb_}, {Dxa_, Dya_, Dza_}] =
1.5*fla[{Dxb, Dyb, Dzb}][[2, 1]]*fla[{Dxa, Dya, Dza}][[3,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*fla[{Dxb,
Dyb, Dzb}][[2, 1]]*fla[{Dxa, Dya, Dza}][[3,
3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) + (0.5*Sqrt[1.5])*fla[{Dxa, Dya,
Dza}][[3, 1]]*fla[{Dxb, Dyb, Dzb}][[2, 3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) -
1) + 0.25*fla[{Dxb, Dyb, Dzb}][[2, 3]]*fla[{Dxa, Dya, Dza}][[3, 3]]*(3*(Cos[thetaf]^2) -
1)*(3*(Cos[thetai]^2) - 1);

```

```

coeff3[{thetaf_, phif_},{thetai_, phii_}, {Dxb_, Dyb_, Dzb_}, {Dxa_, Dya_, Dza_}] =
1.5*fla[{Dxb, Dyb, Dzb}][[3, 1]]*fla[{Dxa, Dya, Dza}][[3,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*fla[{Dxb,
Dyb, Dzb}][[3, 1]]*fla[{Dxa, Dya, Dza}][[3,
3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) + (0.5*Sqrt[1.5])*fla[{Dxa, Dya,
Dza}][[3, 1]]*fla[{Dxb, Dyb, Dzb}][[3, 3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) -
1) + 0.25*fla[{Dxb, Dyb, Dzb}][[3, 3]]*fla[{Dxa, Dya, Dza}][[3, 3]]*(3*(Cos[thetaf]^2) -
1)*(3*(Cos[thetai]^2) - 1);
*****

```

//Example of a two-structure exchange calculation

//Parameters of the jump matrix

//Equilibrium probabilities for the conformational exchange

```

peqA = 0.5;
peqB = 1 - peqA;

```

//Jump matrix elements for the conformational exchange

```

RAB = (1.38*10^6)*Sqrt[peqA/peqB];
RBA = (1.38*10^6)*Sqrt[peqB/peqA];
RAA = -RBA;
RBB = -RAB;

```

//Equilibrium probabilities for the base-libration

```

peq1 = 0.5;
peq2 = 1 - peq1;

```

```

//Jump matrix elements for the base-libration
k12 = (2.15*10^8)*Sqrt[peq1/peq2];

k21 = (2.15*10^8)*Sqrt[peq2/peq1];
k11 = -k21;
k22 = -k12;

//Diffusion tensor parameters for one of the two structures

(*Tw15 0 degrees*)
Diff = {1.964*10^7, 2.028*10^7, 3.440*10^7}; //Diffusion tensor eigenvalues

//Diffusion tensor eigenvectors
D2vec = {-0.5863, -0.3977, 0.7057};
D1vec = {-0.1444, 0.9085, 0.3921};
D3vec = {-0.7971, 0.1279, -0.5901};
D1vec /= Norm[D1vec];
D2vec /= Norm[D2vec];
D3vec /= Norm[D3vec];

//Assignment of diffusion tensor eigenvectors to Cartesian basis vectors
Dxvec = D2vec;
Dyvec = D1vec;
Dzvec = D3vec;

//Check to see if the diffusion tensor frame is a right-handed coordinate system (HYDRONMR
does NOT always produce a right-handed coordinate system)
Dzvec.Cross[Dxvec, Dyvec]

//Bond orientations for the two sites of the base-libration in the PDB coordinate system
site1p = {0.20858583638452116`, 0.9398019932105169`, -0.5125750310123471`};
site2p = {0.14654696550797644`, 1.0074430328966593`, -0.391217998523054`};

//Recalculating the sites in terms of the diffusion tensor principal axis system
site1 = {site1p.Dxvec, site1p.Dyvec, site1p.Dzvec};
site2 = {site2p.Dxvec, site2p.Dyvec, site2p.Dzvec};
site1 /= Norm[site1];
site2 /= Norm[site2];

//Calculation of the spherical angles for each of the site orientations
(*Conf 1 site 1*)
th11 = ArcCos[-site1[[3]]];
ph11 = ArcTan[site1[[2]]/site1[[1]]];
(*Conf 1 site 2*)
th12 = ArcCos[-site2[[3]]];
ph12 = ArcTan[site2[[2]]/site2[[1]]];

```

//Vector of diffusion eigenvalues ordered in terms of the Cartesian associations (i.e. as $\{D_x, D_y, D_z\}$)

Diff1={Diff[[2]], Diff[[1]],Diff[[3]] };

//Diffusion tensor parameters for the second of the two structures, always chosen for this work to be the lowest energy structure of TAR RNA, 1ANR-1

/(Same definitions as before)

(*ANR 0 degrees*)

Diff={2.057*10⁷,2.153*10⁷, 3.465*10⁷};

D2vec={-0.5329,-0.2868, 0.7961};

D1vec={-0.1060, 0.9560, 0.2735};

D3vec={-0.8395,0.0613, -0.5399};

D1vec /= Norm[D1vec];

D2vec /= Norm[D2vec];

D3vec /= Norm[D3vec];

Dxvec = D2vec;

Dyvec = D1vec;

Dzvec = D3vec;

Dzvec.Cross[Dxvec, Dyvec]

(* U38 *)

site1p={0.30277529984708695`,0.7683215988237551`, -0.7112201055815386`};

site2p={0.2418946715947777`,0.8676834436023567`, -0.613544953163507`};

(180/Pi)*ArcCos[site1p.site2p]

site1={site1p.Dxvec, site1p.Dyvec, site1p.Dzvec};

site2={site2p.Dxvec, site2p.Dyvec, site2p.Dzvec};

site1 /= Norm[site1];

site2 /= Norm[site2];

(*Conf 2 site 1*)

th21 = ArcCos[-site1[[3]]];

ph21= ArcTan[site1[[2]]/site1[[1]]];

(*Conf 2 site 2*)

th22 = ArcCos[-site2[[3]]];

ph22= ArcTan[site2[[2]]/site2[[1]]];

Diff2={Diff[[2]], Diff[[1]],Diff[[3]] };

//Calculating the time-dependent coefficients using the time-evolution matrices as defined in Equations 31b and 32

//Evolution matrix for eigenvalue 1

```

evolmat1 = {{-g[Diff1][[1]] + k11+ RAA,k12, RAB,0},{k21, -g[Diff1][[1]] + k22+ RAA,
0,RAB},{RBA,0, -g[Diff2][[1]] + k11+ RBB, k12},{0,RBA, k21,-g[Diff2][[1]] + k22+ RBB}};
//Overlap of the eigenfunctions for the two conformers for the two diffusion tensor-dependent
eigenvalues
sum22 = Sum[fla[Diff1][[2, k]]*fla[Diff2][[2, k]], {k,1,5}];
sum23 = Sum[fla[Diff1][[2, k]]*fla[Diff2][[3, k]], {k,1,5}];
sum32 = Sum[fla[Diff1][[3, k]]*fla[Diff2][[2, k]], {k,1,5}];
sum33 = Sum[fla[Diff1][[3, k]]*fla[Diff2][[3, k]], {k,1,5}];

evalmodff = g[Diff2][[2]]*(sum22^2) + g[Diff2][[3]]*(sum23^2);
evalmodfs = sum22*g[Diff2][[2]]*sum32 + sum23*g[Diff2][[3]]*sum33;
evalmodsf = evalmodfs;
evalmodss = g[Diff2][[2]]*(sum32^2) + g[Diff2][[3]]*(sum33^2);

//Combined evolution matrix for the diffusion tensor-dependent eigenvalues
evolmatcomb = {{-g[Diff1][[2]] + k11+ RAA,0,k12,0, RAB,0,0,0},{0, -g[Diff1][[3]] + k11+
RAA, 0,k12, 0,RAB,0,0},{k21, 0,-g[Diff1][[2]] + k22+ RAA,0,0,0, RAB,0}, {0, k21, 0,-
g[Diff1][[3]] + k22+ RAA,0,0,0, RAB},{RBA,0,0,0,-evalmodff + k11+ RBB,-evalmodfs,
k12,0},{0,RBA,0,0,-evalmodsf,-evalmodss + k11+ RBB,0,k12},{0,0,RBA,0,k21,0,-evalmodff +
k22+ RBB,-evalmodfs},{0,0,0,RBA,0,k21,-evalmodsf,-evalmodss + k22+ RBB}};

//Evolution matrix for eigenvalue 4
evolmat4 = {{-g[Diff1][[4]] + k11+ RAA,k12, RAB,0},{k21, -g[Diff1][[4]] + k22+ RAA,
0,RAB},{RBA,0, -g[Diff2][[4]] + k11+ RBB, k12},{0,RBA, k21,-g[Diff2][[4]] + k22+ RBB}};

//Evolution matrix for eigenvalue 5
evolmat5 = {{-g[Diff1][[5]] + k11+ RAA,k12, RAB,0},{k21, -g[Diff1][[5]] + k22+ RAA,
0,RAB},{RBA,0, -g[Diff2][[5]] + k11+ RBB, k12},{0,RBA, k21,-g[Diff2][[5]] + k22+ RBB}};

//Eigenvalues and eigenvectors of the evolution matrices
eval1 = -Eigenvalues[evolmat1];
evects1 = Eigenvectors[evolmat1];

evalcomb = -Eigenvalues[evolmatcomb];
evectscomb = Eigenvectors[evolmatcomb];

eval4 = -Eigenvalues[evolmat4];
evects4 = Eigenvectors[evolmat4];

eval5 = -Eigenvalues[evolmat5];
evects5 = Eigenvectors[evolmat5];

//Evaluating the time-dependent coefficients, as defined in Equations IV.33 and IV.34
T1 = Transpose[evects1];

```

```
lam1[t_] = { {Exp[-eval1[[1]]*t ],0,0,0},{0,Exp[-eval1[[2]]*t ],0,0},{0,0,Exp[-eval1[[3]]*t
],0},{0,0,0,Exp[-eval1[[4]]*t ]} };
c1[t_] = T1.lam1[t].Inverse[T1];
```

```
Tcomb = Transpose[evectscomb];
lamcomb[t_] = { {Exp[-evalcomb[[1]]*t ],0,0,0,0,0,0,0},{0,Exp[-evalcomb[[2]]*t ],0,0,0,0,0,0},
{0,0,Exp[-evalcomb[[3]]*t ],0,0,0,0,0},{0,0,0,Exp[-evalcomb[[4]]*t ],0,0,0,0,0},
{0,0,0,0,Exp[-evalcomb[[5]]*t ],0,0,0,0},{0,0,0,0,0,Exp[-evalcomb[[6]]*t
],0,0,0},{0,0,0,0,0,0,Exp[-evalcomb[[7]]*t ],0},{0,0,0,0,0,0,0,Exp[-evalcomb[[8]]*t ]} };
ccomb[t_] = Tcomb.lamcomb[t].Inverse[Tcomb];
```

```
T4 = Transpose[evects4];
lam4[t_] = { {Exp[-eval4[[1]]*t ],0,0,0},{0,Exp[-eval4[[2]]*t ],0,0},{0,0,Exp[-eval4[[3]]*t
],0},{0,0,0,Exp[-eval4[[4]]*t ]} };
c4[t_] = T4.lam4[t].Inverse[T4];
```

```
T5 = Transpose[evects5];
lam5[t_] = { {Exp[-eval5[[1]]*t ],0,0,0},{0,Exp[-eval5[[2]]*t ],0,0},{0,0,Exp[-eval5[[3]]*t
],0},{0,0,0,Exp[-eval5[[4]]*t ]} };
c5[t_] = T5.lam5[t].Inverse[T5];
```

```
*****
```

//Calculation of correlation function, spectral densities, relaxation times and NOE

```
solang11 = {th11, ph11};
solang12 = {th12, ph12};
solang21 = {th21, ph21};
solang22 = {th22, ph22};
solang = {solang11, solang12, solang21, solang22};
prconf = {peqA*peq1, peqA*peq2, peqB*peq1, peqB*peq2};
D4site = {Diff1, Diff1, Diff1, Diff1};
```

//Calculation of correlation function as defined in Equation III.37

```
corr[t_] = Sum[Sum[(c1[t][[j,k]]*coeff1[solang[[j]],solang[[k]] ] + ccomb[t][[2*j-1,2*k-
1]]*coeff2[solang[[j]],solang[[k]], D4site[[j]], D4site[[k]] ] + ccomb[t][[2*j,2*k-
1]]*c32[solang[[j]],solang[[k]], D4site[[j]], D4site[[k]] ] + ccomb[t][[2*j-
1,2*k]]*c23[solang[[j]],solang[[k]], D4site[[j]], D4site[[k]] ]+
ccomb[t][[2*j,2*k]]*coeff3[solang[[j]],solang[[k]], D4site[[j]], D4site[[k]] ] +
c4[t][[j,k]]*coeff4[solang[[j]],solang[[k]] ] + c5[t][[j,k]]*coeff5[solang[[j]],solang[[k]]
])*prconf[[k]],{k, 1, 4}], {j, 1, 4}];
```

//Spectral density

```
J0[om_] = Sqrt[Pi/2]*FourierCosTransform[corr[t], t, om];
```

```

//NMR parameters
γC = 6.728*10^7; //Gyromagnetic ratio of Carbon-13
γH = 2.675*10^8; //Gyromagnetic ratio of Hydrogen
rCH = 1.1*10^-10; //Carbon-hydrogen bond length for an aromatic carbon atom
δσ = 212*10^-6; //Chemical shift anisotropy (CSA) for a uridine base carbon atom
//Magnetic field strength
Bfield = 11.74;

//Larmor frequencies of Carbon-13 and Hydrogen
ωC = γC*Bfield
ωH = γH*Bfield

//Calculation of the spectral densities relevant to the solution relaxation times and NOEs as
defined by Equations IV.38-41
J0m = J0[ωC - ωH] ;
J0C = J0[ωC];
J0p = J0[ωC + ωH];
J00 = J0[0];
J0H = J0[ωH];
co2 = (2/15)*(δσ^2)* ωC^2;
d2 = (0.1*10^-14)*((1.05451*10^-34)*γH* γC /rCH^3)^2;

//Relaxation time and NOE calculations
T1 = 1000*(d2*(J0m + 3*J0C + 6*J0p) + co2*J0C )^-1//Chop
T2= 1000*(0.5*d2*(4*J00 + J0m + 3*J0C + 6*J0H + 6*J0p) + (1/6)*co2*(4*J00 + 3*J0C) )^-
1//Chop
NOE = 1 + (γH /γC)*d2*(6*J0p - J0m)*T1/1000//Chop
*****

```


Appendix IV.B: Generating modified atomic-coordinate files to simulate changes in helical parameters

The following changes to the coordinate file are carried out using the Chimera program:

Firstly, we need to translate and rotate the PDB coordinates to align the lower helix with the z-axis of a new coordinate system.

1. Shift the origin of the PDB file to the U40 backbone P atom.
2. Define the upper and lower helices using the program 3DNA (using the local dinucleotide helix vectors, not the normals).
3. Incorporate a unit vector coordinate system explicitly into the coordinate file, to aid with visualization.
4. Create a PDB file with only the target helical axis and superimpose on the original structure.
5. Create an additional PDB file with the original molecular coordinate system and superimpose on the two previous coordinate files—this serves again merely as a visualization aid, as all rotations are carried out with respect to this fixed coordinate frame.
6. Find the polar angle β (called so because it is the Euler angle of rotation as well) between the unit z-axis $(0, 0, 1)^T$ and the target lower helical axis, by calculation using the output of 3DNA.
7. Rotate the molecular coordinates (Tools → Movement → Transform molecular coordinates in the Chimera menu) such that the z-axis of the molecular axis system lines up with the target lower helical axis, and note the Euler angles used to bring about this alignment. Note: Chimera uses the $\{Z, X', Z''\}$ convention, where the first rotation is about the old Z axis, then about the new X' axis and finally about the new Z'' axis.
8. Go back to the original pre-rotated structure and carry out the inverse transformation relative to that in Step 7. This actually transforms the coordinates such that $(0, 0, 1)^T$ aligns with the lower helix (the inverse is required because Step 7 involves an active rotation of the coordinates, while what we require is a passive rotation of the frame).

When the molecular file is prepared in this manner (this new, rotated file with the lower helix along $(0, 0, 1)^T$ is designated as 1ANR-1 in all the calculations), the next steps are designed to rotate the upper helix through various angles of bend and twist relative to 1ANR-1.

9. Choose a direction of the rotation axis about which the upper helix is to be bent. The rotated structure 1ANR-1 from the steps above will in general have an x-axis that is arbitrarily oriented in a plane perpendicular to the lower helix. Therefore, the choice of the bending axis was made by visually inspecting the changes in the structure and finding one or more axes that resulted in a straightening of the upper helix relative to the lower helix. The choice of bending axis is defined in terms of the angle N relative to the 1ANR-1 x-axis.
10. To bend the upper helix through an angle β , apply the following rotation in Chimera: $R_{Z''}(N)R_{X'}(\beta)R_Z^{-1}(N) = R_{Z''}(N)R_{X'}(\beta)R_Z(-N)$, where the application requires simply inputting $\{N, \beta, -N\}$ into the three Euler angle slots.

11. Generate structures with bends of 10° through 50° in 10° increments.

The incorporation of the twist in the upper helix proceeds as follows:

12. Find the upper helical axis for each of the bend angle cases. Calculate the spherical angles of each axis relative to the 1ANR-1 coordinate system, designating the polar angle as θ and the azimuthal angle as ϕ .
13. In order to apply a twist of γ degrees to the upper helix, perform the rotation: $R_{Z''}^{-1}\left(\frac{\pi}{2} - \phi\right)R_{X''}^{-1}(\theta)R_{Z''}^{-1}(\gamma)R_{X''}(\theta)R_Z\left(\frac{\pi}{2} - \phi\right)$. A series of twist angles ranging from -60° to $+60^\circ$ with 15° degree increments are applied for each bend angle structure.
14. The newly rotated structure coordinate files are modified by replacing the lower helix atoms in them with the corresponding coordinates from 1ANR-1, to simulate bending and twisting in the upper helix alone.

All the structure files produced in this manner are inspected for severe steric clashes (i.e. overlaps of atoms from different parts of the molecule), which may occur given the brute force nature of the modifications made above, and the fact that the bulge is kept rigid in all cases. If such clashes are found to be substantial, the files may be discarded in favor of smaller bend angles or a different choice of bending axis.

Chapter V

Studying experimental structure-based trajectories for base-flipping in HhaI methyltransferase recognition DNA

(Work performed with Brian Trinh, former undergraduate student, University of Washington, Department of Chemistry)

Abstract

The HhaI-methyltransferase recognition DNA sequence, [d(G₁A₂T₃A₄G₅C₆G₇C₈T₉A₁₀T₁₁C₁₂)]₂, serves as the binding location for the HhaI protein, which catalyzes the methylation of the target cytosine C₆ at the C5 atom. This catalysis process involves the binding of the HhaI protein to the specific sequence 5'-[G₅C₆G₇C₈]-3', and the subsequent acceptance of the “flipped-out” cytosine into the protein’s binding pocket. A methyl group is then transferred from the donor molecule S-adenosyl-L-methionine (AdoMet) to the cytosine. The question of interest is whether the dynamics of base-flipping are innate to the DNA itself, or whether the protein is required to draw the cytosine out of the double helix. To study this question, we test a model of motion based on three atomic-level structures which represent the flipping trajectory of the target residue against the solid state NMR relaxation time and line shape data obtained for backbone, furanose ring and base sites along the cytosine.

1. Introduction

While DNA molecules are known to largely occur in a stable helical conformation on average, several motions have been reported to occur at all length scales: twisting and bending

motions in long molecules with flexibility at base pair-to-base pair junctions¹¹¹, as well as local base motions.^{101,112,113} (See also Alam and Drobny³⁵ for a review of DNA dynamics studies prior to 1991). One of the more dramatic instances of deviation from rigid helical behaviour was observed for the HhaI methyltransferase-recognition DNA sequence, $[d(G_1A_2T_3A_4\underline{G}_5\underline{C}_6\underline{G}_7\underline{C}_8T_9A_{10}T_{11}C_{12})]_2$.¹⁸ The underlined residues form the conserved recognition site for the HhaI methyltransferase, which acts as a catalyst in the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to the C5 atom of the target cytosine on the sequence. The DNA molecule has been observed to have a modified helix in the bound structure with one of the cytosines (C_6) being extruded out of the helix and being methylated in the process.¹⁸ The extruded C_6 is positioned within the active site of the protein, with the DNA as a whole being located between two domains of the protein. This methylation process is known to protect bacterial DNA from cleavage by restriction endonucleases, and is thought to play a role in eukaryotic gene transcription inhibition,¹¹⁴ X chromosome inactivation, DNA repair^{18,19} and potentially in cancerous cells.²⁰ This essential process in cells proceeds by means of the binding of the methyltransferase at the recognition site and the subsequent “base-flipping” of the target cytosine into the active site of the protein.¹⁸ It is of interest to us to understand whether this flipping process is active or passive, i.e. whether the cytosine is actively flipped out by the protein, or whether the protein opportunistically captures a naturally dynamic residue, similar to the TAR RNA conformational capture idea. This has been investigated extensively^{21,22} with ultimate hope of being able to prevent hypo- or hyper-methylation, if there is a proven medical need to do so.

NMR studies of the HhaI recognition sequence have yielded solid state line shape and relaxation data ($T_{1\rho}$)^{68,115} as well as solution relaxation times (T_1 , $T_{1\rho}$)⁶⁸ for furanose ring and

backbone sites. More recent studies have yielded solid state NMR line shape and relaxation time data for sites on the bases of several residues in the sequence (Kari Pederson, PhD thesis, University of Washington). Meints et al compare the data for the methylated and unmethylated target cytosine at the furanose and backbone sites, and find a change in dynamics between the two chemically different states of the cytosine only for the backbone sites.¹¹⁵ Miller et al carried out studies across various residues in the sequence, and find that the two cytosines in the core recognition sequence to have dramatically different solid state line shapes and $T_{1\rho}$ than their neighbours,⁶⁸ indicating that these residues stand out dynamically from the surrounding sequence. Much analysis has been devoted to the prediction of the solid state relaxation times and line shapes by simulating the furanose ring motions (specifically the C2'-H2'' bond motions) as being determined by continuous potential energy functions of the pseudo-rotation coordinate¹¹⁶: Meints et al propose a double-well potential, with each well representing one of the two primary configurations of the furanose ring, the C2'-endo and the C3'-endo conformations¹¹⁷; Echodu et al treat the dynamics as being centered around the B-DNA-preferred C2'-endo conformations and with deviations away from the potential minimum being determined by a harmonic potential well.²⁹ However, while active interest remains in the motions of individual sites along the nucleotide, the analysis of collective motions of all the components is yet to be considered.

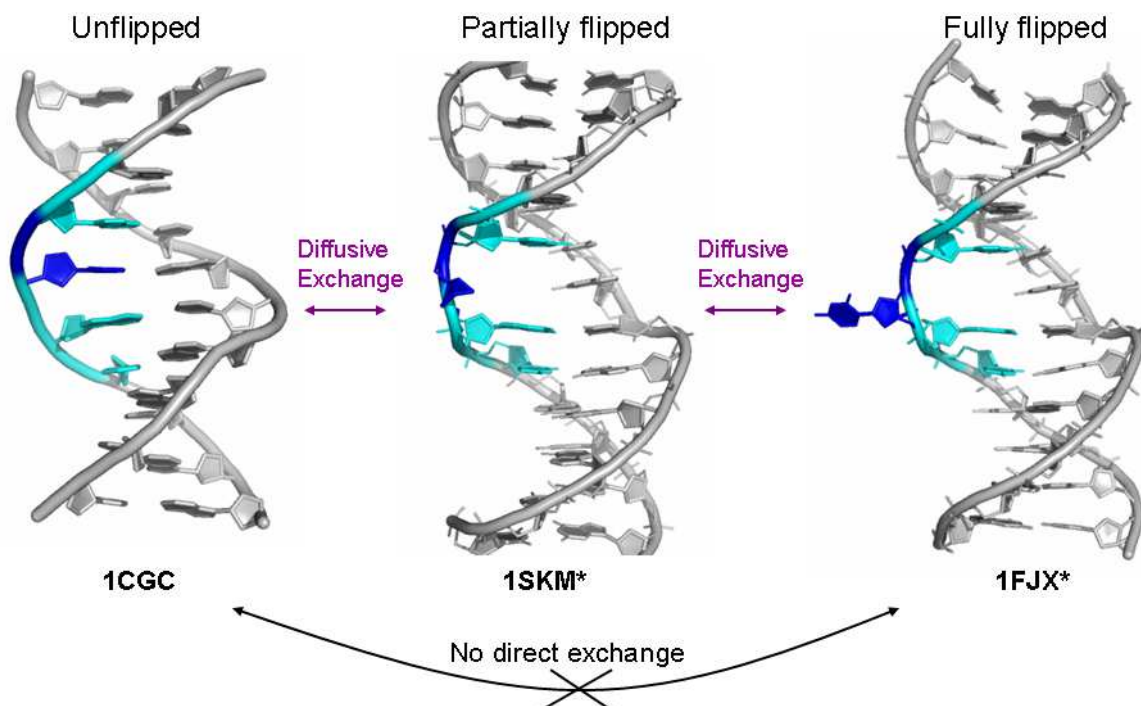


Figure V.1: The three structures used to simulate a trajectory of the flipping process.⁴⁹ The target cytosine of the GCGC recognition sequence is shown in dark blue, while the remaining nucleotides in the recognition sequence are rendered in cyan. The unflipped structure (PDB code: 1CGC) has only ten base-pairs while the partially flipped (PDB code: 1SKM) and fully flipped (PDB code: 1FJX) structures have 12 base-pairs each. As indicated, exchange is allowed only between adjacent structures on the trajectory, while the unflipped and fully flipped structures cannot undergo direct interconversion according to the chosen model. The asterisks next to 1SKM and 1FJX indicate that the DNA structure has been isolated from their PDB files, which also include the binding protein structures.

This was the context within which a potentially general method of parameter-fitting to NMR experimental data was explored: by using three fixed structures, or rotamers (shown in Figure 1) which shared the recognition sequence in common with the HhaI DNA and were deposited previously into the Protein Data Bank (PDB),⁴⁹ we tested the possibility that a single motion could explain observed $T_{1\rho}$ relaxation times by fitting the data for each of three deuterium-labeled sites on the target cytosine using two equilibrium probabilities (the 3rd is fixed by the requirement $\sum_{i=1}^{N=3} p_i = 1$) and a dynamic timescale. The labeled sites on the target cytosine were the backbone C5' (H5' and H5'' equally populated in the sample), the furanose C2' (at the

H2'' site) and the base C5 and C6 sites. These sites represent different environments with different degrees of access to the protein. By probing the cytosine at various locations, it is hoped that we can gain information on whether the different sections collude to produce a single flipping motion, or whether the backbone, sugar ring and base behave independently.

One of the key ideas was to establish an automated search process using a Markov-Chain Monte Carlo (MCMC) simulation (incorporating the Metropolis-Hastings algorithm and simulated annealing of the search condition), where the resulting parameter values were binned and displayed in histograms. Multiple sets of parameters near the peaks of the histograms were selected as potential best-fit parameters, input into a $T_{1\rho}$ calculation to confirm that they reproduced the experimental data within close proximity, and compared against the results for the other sites within the cytosine residue.

Given that the relaxation times are only sensitive to motions in the vicinity of the Larmor frequency of the deuteron (as well as $2\omega_D$, see Equation II.24), we must also consider the possibility that a flipping motion could occur on a slower time scale. Considering the availability of line shape data that are affected by motions on time scales of μs to ms , the subsequent goal is to carry out the fitting of the same dynamic parameters to the solid state NMR experimental line shapes. Michael Groves (former member of this research group) has created a simulated annealing suite of programs (unpublished as yet) that searches for the parameter set that minimizes the chi-squared between calculated and experimental line shapes, with the MXET1^{61,118} program being used within the suite to generate the line shape for each parameter set. By utilizing this automated fitting procedure, it is hoped that a common set of best-fit parameters may be obtained, and that this procedure of joint parameter matching may be

extended to larger numbers of variables. More specifically for the three-site jump motional model, the intentions behind the conjoined fitting are two-fold:

- a. Testing the validity of motional models:** If a common set of parameters is found, the case for the motional model under consideration is clearly strengthened. Other data, such as solution NMR relaxation times (see following subsection), could be simulated using this model, which could then be refined iteratively if necessary.
- b. Capturing motion on different timescales:** The relaxation time T_1 is sensitive to dynamics occurring on a scale up to tens of nanoseconds, while the line shape is sensitive to motions on the nanosecond to millisecond timescale. Any discrepancy between the parameter sets for each could represent the existence of additional motions on timescales longer than is accessible by relaxation times alone, or vice-versa. The combination of the two results could suggest the need for a diffusion model that incorporates multiple timescales, rather than one.

To date the procedure has only been able to find a joint set of fit parameters for the $T_{1\rho}$ and line shape data for the furanose ring using this particular model. This is not conclusive, however, given that there may be other areas of parameter space as yet unexplored that do allow for a single flip model.

2. Theory

The basic method used here is to construct a three-site jump process, using eigenvalue methods in order to derive an expression for the transition probability, which in turn helps establish the correlation function $C(t) = \left\langle D_{mn}^{(2)*}(\Omega_{PL}(0)) D_{m'n'}^{(2)}(\Omega_{PL}(t)) \right\rangle_{motion}$ (Equation II.21; the

definitions of Wigner rotation matrices in terms of Euler angles are provided in Equations II.5a and b). The first step in the procedure is to lay out the multiple frame transformations that interpolate between the principal axis system (P) of the EFG tensor in the neighbourhood of the deuteron and the lab frame (L). In short DNA molecules, such as the HhaI recognition DNA with 12 base pairs (bp), it is necessary to account for the diffusive motion of the double helix about the axis of symmetry, i.e. the helical axis, even in the solid state (see Alam and Drobny³⁵ and Gary Meints, Ph.D. Thesis, University of Washington). Thus, the first transformation is from the P frame to the helical axis frame H, which rotates at a characteristic diffusion rate about the long axis of the helix. Secondly, the frame H must be related to a static molecule-attached frame M. Clearly, the transformation between H and M can be effected by a single time-dependent polar angle, if we take their z-axes as coincident and along the molecular axis of symmetry. Finally, we transform between the static molecular frame M to the static lab frame. An important simplification is used at this point: all solid state data considered were treated as “powder-averaged”, i.e. there was assumed to be no preferential orientation of the molecules along particular crystal director axes. The molecules in the sample were assumed to be equally distributed over all solid angles. We combine the transformation set

$P \xrightarrow{\tilde{\Omega}_{PH}} H \xrightarrow{\tilde{\Omega}_{HM}} M \xrightarrow{\tilde{\Omega}_{ML}} L$ into the calculation of the correlation function in the following.

The correlation function can be separated out into the contributions from various frame transformations as follows:

$$\begin{aligned}
 C(t) &= \left\langle D_{mn}^{(2)*}(\tilde{\Omega}_{PL}(0)) D_{m'n'}^{(2)}(\tilde{\Omega}_{PL}(t)) \right\rangle \\
 &= \sum_{\substack{a,a'=-2 \\ b,b'}}^2 \left\langle D_{ma}^{(2)*}(\tilde{\Omega}_{PH}(0)) D_{m'a'}^{(2)}(\tilde{\Omega}_{PH}(t)) \right\rangle \left\langle D_{ab}^{(2)*}(\tilde{\Omega}_{HM}(0)) D_{a'b'}^{(2)}(\tilde{\Omega}_{HM}(t)) \right\rangle \left\langle D_{bn}^{(2)*}(\tilde{\Omega}_{ML}(0)) D_{b'n'}^{(2)}(\tilde{\Omega}_{ML}(t)) \right\rangle \quad (1)
 \end{aligned}$$

In Equation 1, the assumption has been made that the various frame transformations represent independent motions, allowing for the motional-averaging to be carried out over the separate parts. This is a fairly common assumption for motions that may either be on completely different time scales or that are dependent on different degrees of freedom whose variations have little physical impact on each other. Moreover, the Wigner functions in the third average in the final expression of equation 1, $\langle D_{bn}^{(2)*}(\vec{\Omega}_{ML}(0))D_{b'n'}^{(2)}(\vec{\Omega}_{ML}(t)) \rangle$, have no time dependence in a solid state sample. These functions will be powder-averaged as follows:

$$\begin{aligned}
\bar{C}(t) &= \frac{1}{8\pi^2} \int_{\vec{\Omega}_{ML}} d\vec{\Omega}_{ML} C(t) \\
&= \frac{1}{8\pi^2} \int_{\vec{\Omega}_{ML}} d\vec{\Omega}_{ML} \sum_{\substack{a,a'=-2 \\ b,b'}}^2 \langle D_{ma}^{(2)*}(\vec{\Omega}_{PH}(0))D_{m'a'}^{(2)}(\vec{\Omega}_{PH}(t)) \rangle \langle D_{ab}^{(2)*}(\vec{\Omega}_{HM}(0))D_{a'b'}^{(2)}(\vec{\Omega}_{HM}(t)) \rangle D_{bn}^{(2)*}(\vec{\Omega}_{ML})D_{b'n'}^{(2)}(\vec{\Omega}_{ML}) \\
&= \frac{1}{5} \sum_{\substack{a,a'=-2 \\ b}}^2 \langle D_{ma}^{(2)*}(\vec{\Omega}_{PH}(0))D_{m'a'}^{(2)}(\vec{\Omega}_{PH}(t)) \rangle \langle D_{ab}^{(2)*}(\vec{\Omega}_{HM}(0))D_{a'b'}^{(2)}(\vec{\Omega}_{HM}(t)) \rangle \delta_{nn'}\delta_{bb'} \\
(2)
\end{aligned}$$

where the orthogonality property of the Wigner matrices has been used:

$$\int d\vec{\Omega} D_{mk}^{(l)*}(\vec{\Omega})D_{m'k'}^{(l')}(\vec{\Omega}) = \frac{8\pi^2}{2l+1} \delta_{ll'}\delta_{mm'}\delta_{kk'} \quad (3)$$

The second average (corresponding to the transformation from the helical frame to the M frame) has already been computed²³ for the case of time-dependence only in the γ_{HM} Euler angle; the H and M frames are assumed to have their z-axes aligned with each other, and the H frame swivels around the common z-axes, resulting in a change in γ_{HM} only ($\alpha_{HM} = 0, \beta_{HM} = 0$):

$$\langle D_{ab}^{(2)*}(\vec{\Omega}_{HM}(0))D_{a'b'}^{(2)}(\vec{\Omega}_{HM}(t)) \rangle = \exp(-D_{Helix}a^2t)\delta_{ab}\delta_{a'b} \quad (4)$$

The helical diffusion rate D_{Helix} is taken in all simulations herein to be 1.1×10^4 Hz.^{21,35}

For the final average, $\langle D_{ma}^{(2)*}(\vec{\Omega}_{PH}(0)) D_{m'a}^{(2)}(\vec{\Omega}_{PH}(t)) \rangle$, we first write out the motional average in terms of the transition probability and the *a priori* probability (initial orientational probability distribution):

$$\langle D_{ma}^{(2)*}(\vec{\Omega}_{PH}(0)) D_{m'a}^{(2)}(\vec{\Omega}_{PH}(t)) \rangle = \int d\vec{\Omega}_0 P_{eq}(\vec{\Omega}_0) \int d\vec{\Omega}_t P(\vec{\Omega}_t, t | \vec{\Omega}_0) D_{ma}^{(2)*}(\vec{\Omega}_0) D_{m'a}^{(2)}(\vec{\Omega}_t) \quad (5)$$

where the notational simplifications of $\vec{\Omega}_0 \equiv \vec{\Omega}_{PH}(0)$ and $\vec{\Omega}_t \equiv \vec{\Omega}_{PH}(t)$ were made. $P(\vec{\Omega}_t, t | \vec{\Omega}_0)$ is called the transitional probability and represents the probability that the bond in question will be in an orientation of $\vec{\Omega}_t$ at time t, given its initial orientation of $\vec{\Omega}_0$. All information regarding the specific motional model being considered is contained in $P(\vec{\Omega}_t, t | \vec{\Omega}_0)$.

We now proceed to setting up a discrete site jump. For a discrete diffusive process, the site probability satisfies the equation

$$\frac{\partial}{\partial t} \vec{P} = \vec{R} \cdot \vec{P} \quad (6)$$

Here, the components of $\vec{P}$ are the probabilities of being in a site i at time t, and $\vec{R}$ is the jump matrix that “transfers” probability from one site to another at each point in time. Equation 6 is the discretized version of the full 3-dimensional Fokker-Planck equation in an Euler angle representation. The form of the jump matrix $\vec{R}$ is as follows¹¹⁹:

$$\begin{aligned} R_{ij} &= \frac{1}{\tau} \sqrt{\frac{P_{eq}(i)}{P_{eq}(j)}}; j = i \pm 1, \\ R_{ij} &= 0; \forall j \neq i, i \pm 1 \\ R_{ii} &= -\sum_{j \neq i} \frac{1}{\tau} \sqrt{\frac{P_{eq}(j)}{P_{eq}(i)}} \end{aligned} \quad (7)$$

where $P_{eq}(i)$ is the equilibrium probability that the bond will be in a given orientational site i . The elements $P_{eq}(i)$ are fit parameters in the MCMC search algorithm, under the constraint that $\sum_{i=1}^{N_{sites}} P_{eq}(i) = 1$. The form of the jump matrix in Expression 7 has been chosen based on two considerations: (a) only nearest neighbour jumps are permitted, as quantified by the restriction $j = i \pm 1$; and (b) the matrix elements satisfy the condition of detailed balance, $R_{ij} P_{eq}(j) = R_{ji} P_{eq}(i)$, which conserves probability for each pair of sites when the system is in equilibrium. For the specific motional model discussed in this chapter, consideration (a) forbids direct transitions between the unflipped (PDB: 1CGC) and fully-flipped (PDB: 1FJX) structures, as indicated in Figure 1. For one state to access the other, they must pass through the partially-flipped state (PDB: 1SKM). Physically, this is restricting the flipping process to the major groove only, as the partially-flipped structure takes this path. Also, the time scale τ in Expression 7 is one of the fit parameters sought in the MCMC protocol.

The method of solution of Equation 6 proceeds by solving for the eigenvalues and eigenvectors of the jump matrix $\vec{R}$. The transition probability is then given by:

$$P(\vec{\Omega}_i, t | \vec{\Omega}_j) = (\vec{T} e^{-\vec{\Lambda} t} \vec{T}^{-1})_{ij} \quad (8)$$

where $\vec{T}$ is the matrix with the eigenvectors of $\vec{R}$ as its columns, and $\vec{\Lambda}$ is the diagonal matrix of eigenvalues.

$$\text{The final step is to calculate the spectral density } J(\omega) = \text{Re} \left[\int_0^\infty \overline{C}(t) e^{i\omega t} dt \right] = \int_0^\infty \overline{C}(t) \cos(\omega t) dt$$

(Equation II.22), which combines the results of Equations 2, 4, 5 and 8 to yield:

$$J(\omega) = \frac{1}{5} \sum_{a=-2}^2 \sum_{k=1}^3 P_{eq}(\vec{\Omega}_j) D_{0a}^{(2)*}(\vec{\Omega}_j) D_{0a}^{(2)}(\vec{\Omega}_k) \sum_{l=1}^3 \left(\vec{T} \frac{D_{Helix} a^2 + \lambda_l}{(D_{Helix} a^2 + \lambda_l)^2 + \omega^2} \vec{T}^{-1} \right)_{kj} \quad (9)$$

A further simplification was made in arriving at Expression 9: the asymmetry parameter η was set to 0 for all deuterons attached to carbons. This results in the indices m and m' being set to 0 (see expressions for EFG tensor in the PAS frame, Equation II.9b).

Having derived an expression for the spectral density for a given choice of the dynamic parameters $\{p_1, p_2, \tau\}$, the relaxation time T_{1Z} is calculated as

$$\frac{1}{T_{1Z}} = \frac{3}{8} \left(2\pi \frac{e^2 q Q}{h} \right)^2 (J(\omega_D) + 4J(2\omega_D)) \quad (10)$$

The quadrupolar coupling magnitude is quantified by $\frac{e^2 q Q}{h} = 174$ kHz, and ω_D , the Larmor frequency of deuterium at a magnetic field of 11.74 T, $= 4.82 \times 10^8$ rad s⁻¹.

The detailed line shape calculation is incorporated into the program MXET1^{61,118} and will not be reiterated here. The calculation for the two motions discussed proceeds as described in Chapter II, Section 2b, where the frame C corresponds to the helical frame H.

3. Methods

a. Reconstruction of base and accounting for the contributions of multiple labels

The partially flipped structure used in the simulations (PDB: 1SKM) was “captured” in mid-flip by constraining the furanose ring to be in the south or C2'-endo conformation, which in turn was achieved by using an abasic south bicycle[3.1.0]hexane sugar analogue.⁴⁹ Therefore, given that the PDB file did not have coordinates for the base, we had to reconstruct the vector orientations of the base C5-H5 and C6-H6 bonds using ideal geometries and extrapolating

between the unflipped and fully-flipped cases: firstly, the glycosidic bond, C1'-N1, was calculated assuming a tetrahedral geometry for the aliphatic C1' bonds; secondly, the base was reconstructed using ideal sp^2 geometries in the plane containing the glycosidic bond and the O4'-C1' bond; and finally, the base was rotated through a torsion angle of $+19^\circ$ about the glycosidic bond as an approximate intermediate degree of torsional rotation relative to the two other structures. The resulting C5-H5 and C6-H6 bonds were used in the calculation of the Euler angle transformations as described in the next sub-section.

It is also worth explicitly stating that, given that the deuterium-labeling scheme for the backbone sites could not distinguish between the C5'-H5' and C5'-H5'' sites, it was assumed that they were both equally populated in the samples.¹¹⁵ The calculation of the relaxation times for backbone sites, therefore, was carried out by adding the equally weighted contributions from each of the two sites (each with a different set of Euler angles) to the spectral densities.

Moreover, since data was nearly identical for both the C5- and C6-deuterated base samples, the search algorithm sought to find parameters that simultaneously fit a single experimental relaxation time for both sites (again, with different sets of Euler angles) by modifying the test condition in the MCMC acceptance procedure to incorporate both calculated relaxation times. This assumes that the motions of the two base sites are identical, an assumption that remains valid as long as the base orientation remains fairly close to being perpendicular to the helical axis at all times. The furanose C2' was singly labeled at the H2'' site.

b. Evaluation of frame transformations

The input of information from the PDB structures into the algorithm elaborated above occurs through the Euler angles that describe the various frame transformations. The methods

used to extract this information from the PDB files will be briefly discussed in the following. Before that, however, it is important to mention that the three PDB files share the same recognition sequence, but have different flanking sequence. Originally, we attempted to address any differences that may arise from this by including an additional, static intermediate frame between H and the PAS (P) frame. This intermediate frame was established with reference purely to the recognition sequence. However, in the end this was deemed to be extraneous when tested against the relaxation time calculations and so was left out.

The helical frame H was defined by a combination of two different methods. Firstly, the C1' to C1' vectors were calculated for all nucleosides on one strand (excluding the terminal nucleosides, target nucleoside, and the nucleosides adjacent to the target cytosine), and for the complimentary strand as well, with the direction of vectors reversed. The resultant normalized average vector over all such C1' to C1' vectors yielded one evaluation of the helical axis. The terminal, target and adjacent-to-target nucleosides were left out to avoid non-helical discrepancies expected in these regions. Secondly, base-normal vectors for all the residues (with the exception of the ones mentioned above), using the cross products of the C6-N1 and C2-N1 vectors for the purines and the C4-N3 and C2-N3 vectors for the pyrimidines, were calculated and averaged to yield a second estimation of the helical axis. These two estimations were averaged to yield the final helical axis. This approximate symmetry axis of the molecule was chosen as the z-axis of the H frame. The y-axis of the H frame was taken to be the cross product of the helical axis and the C1'-to-C1' vector of the C-G base pair of the recognition sequence not adjacent to the target cytosine, to maintain consistency across the three structures with different flanking sequences.

The PAS (P) frames were taken to be axially symmetric (that is the quadrupolar asymmetry was assumed to be negligible), and so transformations from the H to P frames only required two Euler angles. To facilitate the calculations of these angles, we defined the P frame z-axis to be the carbon-deuterium bond of the site in questions and the P frame y-axis to be the normal to the plane containing the P frame z-axis and the H frame z-axis.

The Euler angle transformations from the molecular (M) frame to the H frame, and from the H frame to the P frame were then calculated by rotating the second frame relative to the fixed first frame: α is the angle between the fixed frame y-axis and the line of nodes, β is the angle between the fixed frame z-axis and the rotated frame z-axis, and γ is the angle between the line of nodes and the rotated frame y-axis.

For the relaxation time calculations, the M to H frame transformation was treated by a continuous diffusion process as a function of one of the Euler angles (either the α or the γ angle depending on which direction the transformation is made, for example, H to M or M to H). The line shape calculation using MXQET1 requires a discrete jump process and so the motion was treated by a six-site jump with 360° coverage, resulting in a site-separation of 60°. ^{21,22}

c. Automated calculation of relaxation times

The Euler angles calculated in the manner mentioned above are then inserted into Equation 9. Further evaluation of the expression for the spectral density requires a calculation of the eigenvalues and eigenvectors of the three-site jump matrix, which in turn is set up using the equilibrium probabilities of the sites and the time scale of the motion. In order to find the best-fit parameters for the relaxation data, we programmed Equations 7 through 10 into Mathematica¹²⁰, and placed the spectral density algorithm within a Markov Chain Monte Carlo (MCMC)

framework, where every iteration involved a change in one of the three parameters. The search algorithm was a Metropolis-Hastings methodology, together with an annealing mechanism, where the condition for acceptance of another set of parameters became more stringent with each step. The algorithm is shown briefly in Figure 2.

It should be noted that in order to cover several orders of magnitude in the search for the time scale τ , assuming that no prior conditions are placed on its value, we used the logarithm of τ in the method. Also, in order to ensure that the search algorithm sampled parameter space sufficiently without showing biases related to the initial conditions, the program was run with several different seed parameter sets. The probabilities were varied from value of 0.1 to 0.9 and initial $\log(\tau)$ values were cycled through the values $\{\log(10^{-2}), \log(10^{-3}), \log(10^{-4}), \log(10^{-5}), \log(10^{-6}), \log(10^{-7}), \log(10^{-8}), \log(10^{-9})\}$. The number of iterations per choice of initial condition was chosen to be 2000. Essentially, the method may be considered a hybrid grid search/MCMC algorithm.

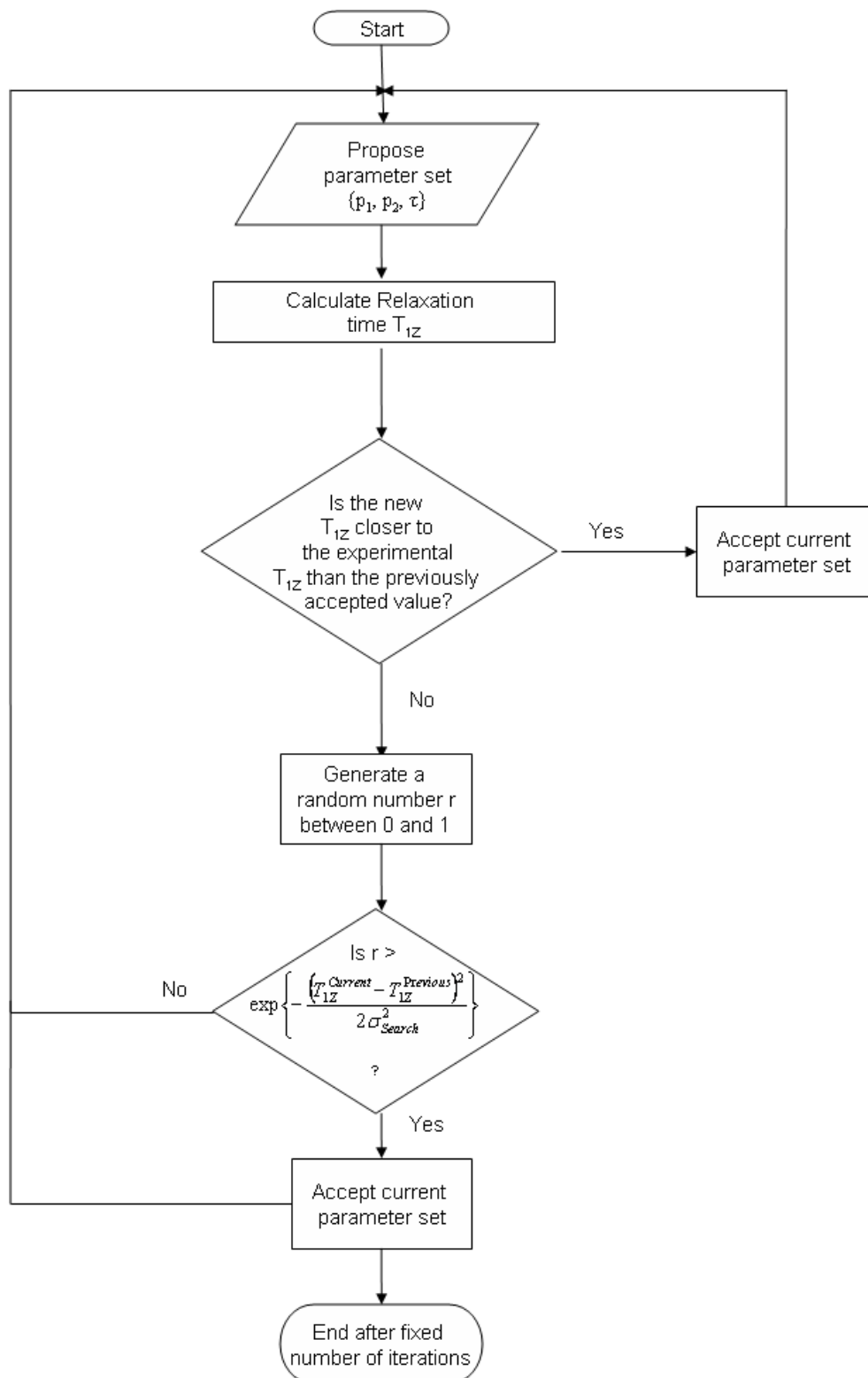


Figure V.2: Flowchart of the Metropolis-Hastings algorithm used to search through parameter space.

The resulting collection of accepted parameter sets are then binned in histograms, with the number of counts in a given region of parameter space being proportional to the probability that the “true value” of the parameter lies in that region. It is hoped that this method provides a more efficient means of exploring large regions of a potentially complicated parameter space.

Table V.1: Best-fit dynamic parameters for the three labeled sites on the target cytosine residue. Also shown are the T_{1Z} values calculated for these parameter choices and the corresponding experimental relaxation times (Meints, G. A. Ph.D. Thesis, University of Washington).

Labeled site	Fractional population of 1CGC (p_1)	Fractional population of 1SKM (p_2)	Fractional population of 1FJX (p_3)	Time scale of motion (τ) (in s^{-1})	Simulated T_{1Z} (in ms)	Experimental T_{1Z} (in ms)
Backbone C5'-H5'/H5''	0.19	0.51	0.30	7.05×10^{-8}	46.0	$46 \pm 4^*$
Furanose C2'-H2''	0.19	0.51	0.30	2.38×10^{-8}	29.0	$29 \pm 3^\dagger$
Base C5-H5 & C6-H6 (combined analysis)	0.19	0.51	0.30	1.72×10^{-7}	C5-H5: 134.3 C6-H6: 147.2	$141 \pm 2^*$

*Experiments done at a hydration of $W = 11.8 \pm 1$. † Experiments done at a hydration of $W = 11.5 \pm 1$. W is the number of waters per nucleotide.

4. Results

We present one set of results from carrying out fits to the T_{1Z} relaxation times in Table 1. Note that these particular best-fit equilibrium probabilities (i.e. fractional populations) are the same for all three labeled sites, implying that it is possible to find common weights for the

structures. It is important to understand, however, that there are several potential combinations of the parameters that fit the experimental T_{1Z} 's, as may be seen from examining the histograms for the backbone simulations (Figure 3), furanose simulations (Figure 4) and base simulations (Figure 5). The probabilities in Figures 3B, 4A, 4B, 5A and 5B are peaked around values close to those indicated in Table 1. While the backbone p_1 values in Figure 3A peak mainly at a slightly higher value than that shown in Table 1, there is sufficient width to the distribution to warrant the choice shown above. The histograms of $\log(\tau)$ do indicate several possible peaks in Figures 3C, 4C and 5C. In choosing a particular value of the time scale, however, one must take into account the possibility that the MCMC algorithm may retain “memory” of the initial condition. To test whether some of the peaks may arise as a result of insufficient exploration away from the starting point, we compared the peak times to the τ_{initial} 's chosen for the search: $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}\}$. The $\log(\tau)$ values for all peaks but one were within a value of 0.2 away of a certain value of $\log(\tau_{\text{initial}})$, where the central values of the bins were separated by 0.4. The one exceptional peak, which also had the most counts, was at a value slightly greater than 0.4 away from the nearest $\log(\tau_{\text{initial}})$. Indeed, when this peak was used to calculate the T_{1Z} value along with the probability peaks, it provided a close fit to the experimental data. Some exploration near the bin value was subsequently needed to fine-tune the fit.

While relaxation time-fitting is inherently sensitive to motional rates near the deuterium Larmor frequency (and $2\omega_D$), the algorithm clearly has the ability to discern differences within one or two orders of magnitude, which may result from genuinely different dynamics at the different sites.

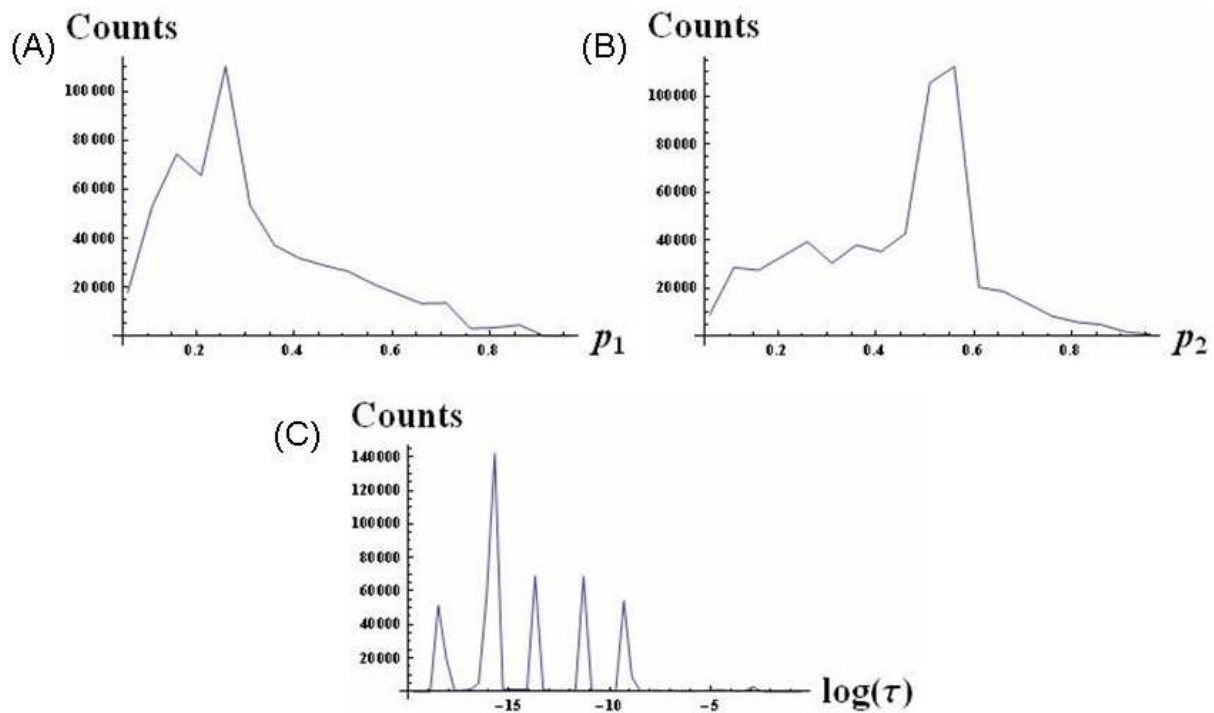


Figure V.3: Histograms resulting from a binning of the accepted MCMC dynamic parameter sets for the backbone-labeled sites (i.e. equal populations of C5'-D5' and C5'-D5'', where D represents the deuteration of the site) of the target cytosine: (A) Histogram for the fractional population of the unflipped structure (1CGC); (B) Histogram for the fractional population of the partially flipped structure (1SKM); and (C) Histogram of the logarithm of the time scale of motion.

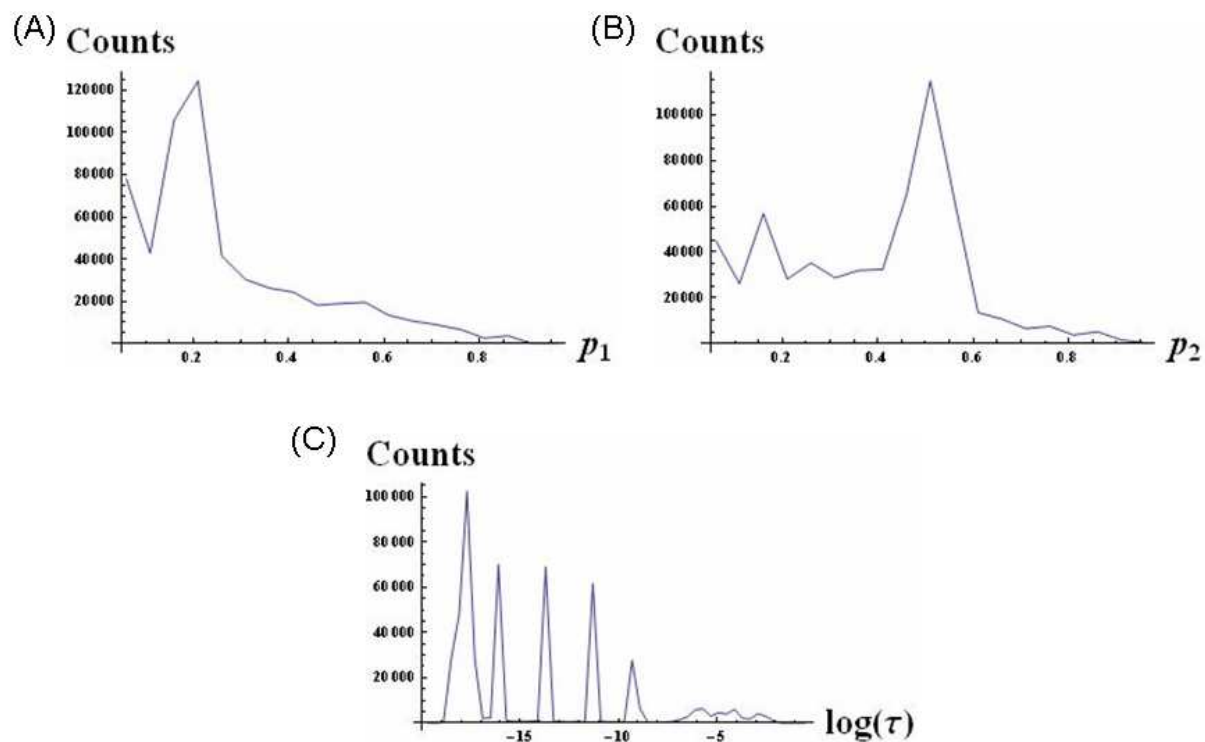


Figure V.4: Histograms resulting from a binning of the accepted MCMC dynamic parameter sets for the furanose ring-labeled sites (i.e. C2'-D2'', where D represents the deuteration of the site) of the target cytosine: (A) Histogram for the fractional population of the unflipped structure (1CGC); (B) Histogram for the fractional population of the partially flipped structure (1SKM); and (C) Histogram of the logarithm of the time scale of motion.

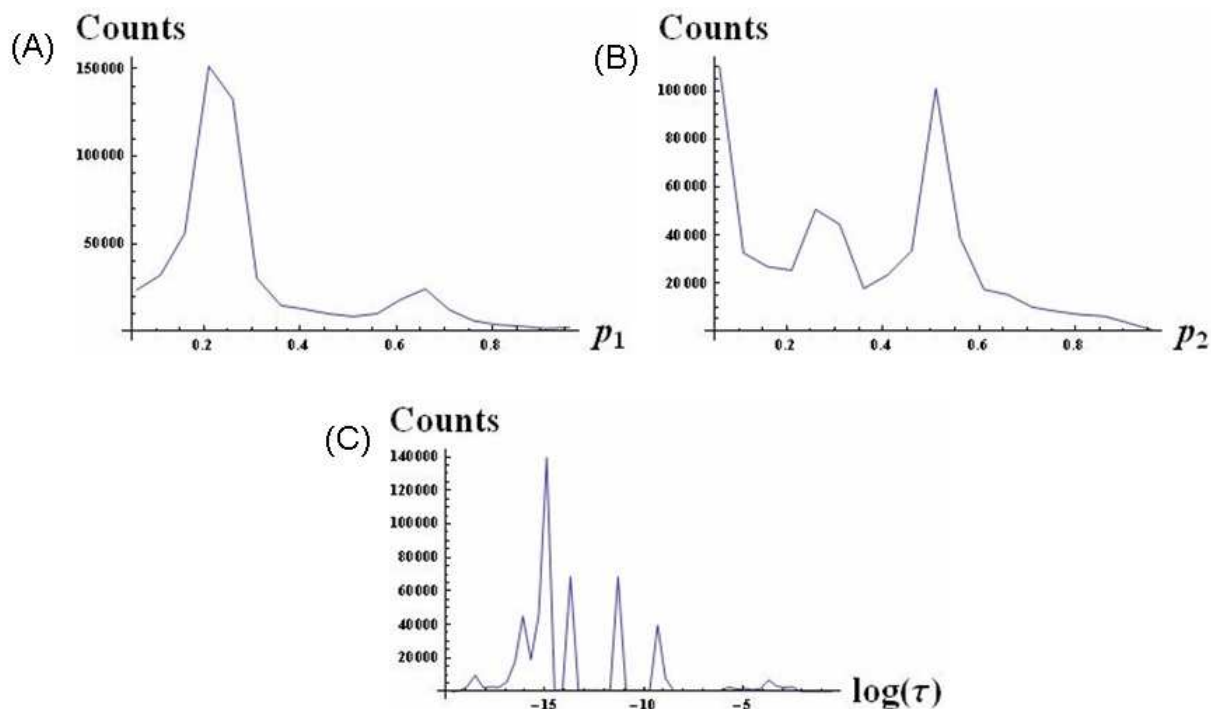


Figure V.5: Histograms resulting from a binning of the accepted MCMC dynamic parameter sets for the base-labeled sites (i.e. both C5-D5 and C6-D6, where D represents the deuteration of the site) of the target cytosine: (A) Histogram for the fractional population of the unflipped structure (1CGC); (B) Histogram for the fractional population of the partially flipped structure (1SKM); and (C) Histogram of the logarithm of the time scale of motion.

We subsequently attempted fits to the line shapes as well, hoping to either reaffirm or contradict the particular three-site jump models suggested by the relaxation times. To date we have not been able to produce reasonable fits to any of the sites, but we have seen that the reduced chi-square for the fits to the data using the best-fit values from Table 1 was extremely high implying that either there is a significantly different set of parameters that happens to fit both results, or more likely, that the complementary sensitivities of the two observables are detecting different motions. The attempts to fit the line shape data to these sites will continue.

5. Discussion

Considering the best-fit parameters displayed in Table 1, it is interesting that the same probabilities fit the data, but not the motion time scales. One potential explanation is that these three sites do represent highly sampled structures in the flipping trajectory, but different parts of the residue explore these sites semi-independently. Thus, the furanose ring and the backbone move between the three orientations at rates faster than the base, implying mobility at these sites that might be able to facilitate the final process of flipping the base, instead of all three having to change their conformations simultaneously. It is also possible that the two slightly different hydration levels (11.8 ± 1 waters per nucleotide for the backbone and base-labeled samples and 11.5 ± 1 waters per nucleotide for the furanose ring-labeled sample; values from Gary A. Meints, Ph.D. Thesis, University of Washington) would cause some discrepancies. However, this still does not explain the difference between the backbone and base results. The other, more mundane, explanation is that the similarity of the population fits to the relaxation times is purely coincidental, and that other, disparate combinations of parameters represent the actual physical situation. Finally, one must always pursue the notion that other motions and/or sites may be superimposed on top of this single flipping trajectory, or even that the relaxation times may not be sufficiently sensitive at the relevant time scale to capture the actual flipping process.

Additional dynamic contributions may include those of the base due to motion around the glycosidic bond or other degrees of freedom, unaccounted motions of the sugar atoms (for example, faster C2'-endo to C3'-endo inter-conversions, or motion within potential energy minima for one of the two conformations) or even the presence of other states with substantial populations along the overall flipping trajectory. These possibilities need to be considered by evaluating previous data and quantitative models derived from them.

With respect to the relaxation time-fitting alone, one must consider the possibility that the flipping process itself occurs on a time scale much slower than the characteristic time of T_1 . This would imply that while the lineshape would be sensitive to a slower flipping process (up to a millisecond time scale), the relaxation time measurements do not “see” the relevant motion. The fit to the relaxation data in this case could be purely accidental, given the simplicity of the three-site model and the number of free parameters in the problem. While at the same time emphasizing one of the potential pitfalls of a blind parameter-fitting procedure (i.e. fitting a model to data which cannot possibly be affected by the underlying model), such a result would tender the valuable conclusion that discrepancies in the T_1 values of the different sites must arise from differences in local motions.

It has been mentioned that, to date, attempts to fit the line shape have failed. This is not conclusive in any way, as the line shape fitting process involves additional free parameters (such as the Lorentzian and Gaussian broadening parameters for the inevitable water peak observed due to deuterium contamination of the solvent, as well as the width of the peak). It is possible that there is sufficient degeneracy in the fitting process, where the appropriate parameters have not been accessed by the search algorithm as yet. Further attempts will be made to this end. However, we must also consider the possibility that the line shapes cannot be fit satisfactorily by a simple three-site jump. Again, this is valuable information in itself, as the data would necessitate a more complex model. If it emerges that the relaxation times and line shapes have disparate domains of best-fit models, as was hinted at earlier by chi-square values of the line shape fits using relaxation time fit-parameters, then they might serve as useful indicators of different motions at their respective time scales. This would be a good first step towards building up a picture of the flipping.

Finally, it must be emphasized that all the preceding discussion was aimed at addressing the fundamental question: is the flipping motion inherent to the free DNA or is the presence of the protein required to force the residue out of the helix? In the end, a phenomenological model-based approach may have to reconcile itself to the possibility that the very assumption that led to its proposal may be wrong; the free DNA may not traverse the full flipping trajectory after all, and may not access all the three sites. The partially flipped and fully flipped structures were generated with the DNA in complex with the protein, and may not occur in its absence. An established lack of good fits to all sets of data may reflect this reality.

Having said that, we believe the value of the work presented here lies in the methodology. The fitting of models to a variety of data requires care and caution regarding the various assumptions, and, especially in the case of biological systems with many degrees of freedom, care and caution regarding the complex geometry at all scales. We have endeavoured to make a thorough examination of the process of modeling the dynamics of a seemingly simple scenario, and hope that the procedure can be emulated for other systems as well, with all the accompanying complications of interpretation being addressed fully.

6. Conclusions

The work presented herein represents a scheme of model-testing where the motional trajectory of residue is constructed from suggested rotamers of the given molecule and subsequently tested against solid state NMR relaxation data. While it has been shown that the automated parameter fitting process does indeed result in parameter sets that reproduce the relaxation data, these choices of probabilities and time scales do not seem to be able to fit the solid state NMR line shape data, signaling the need for either more expansive searches through

the parametric landscape, or a more complex model of motion. Such a model may incorporate the simple three-site jump investigated here or even supercede it altogether by providing a richer trajectory that explains all the data. We believe, however, that we have designed a methodology that helps to provide verification of entire residue-motional models at the atomic level by combining fits to data from multiple sites along the residue into a unified picture.

Appendix V.A: Sample Mathematica code for best-fit parameter search for the backbone sites

(Font change allows for special Mathematica symbols to be easily incorporated)

//Function and parameter definitions

//Definition of the jump matrix for the three-site jump and its eigenvalues and eigenvectors

```
M[{k12_, k21_, k23_, k32_}] = {{-k21, k12, 0}, {k21, -k12 - k32, k23}, {0, k32, -k23}};
Values[{k12_, k21_, k23_, k32_}] = Eigenvalues[M[{k12, k21, k23, k32}]];
λ1[{k12_, k21_, k23_, k32_}] = Values[{k12, k21, k23, k32}][[1]];
λ2[{k12_, k21_, k23_, k32_}] = Values[{k12, k21, k23, k32}][[2]];
λ3[{k12_, k21_, k23_, k32_}] = Values[{k12, k21, k23, k32}][[3]];
Vectors[{k12_, k21_, k23_, k32_}] = Eigenvectors[M[{k12, k21, k23, k32}]]];
v1p[{k12_, k21_, k23_, k32_}] = Vectors[{k12, k21, k23, k32}][[1]];
v1[{k12_, k21_, k23_, k32_}] = v1p[{k12, k21, k23, k32}]/Norm[v1p[{k12, k21, k23, k32}]];
v2p[{k12_, k21_, k23_, k32_}] = Vectors[{k12, k21, k23, k32}][[2]];
v2[{k12_, k21_, k23_, k32_}] = v2p[{k12, k21, k23, k32}]/Norm[v2p[{k12, k21, k23, k32}]];
v3p[{k12_, k21_, k23_, k32_}] = Vectors[{k12, k21, k23, k32}][[3]];
v3[{k12_, k21_, k23_, k32_}] = v3p[{k12, k21, k23, k32}]/Norm[v3p[{k12, k21, k23, k32}]];
Tm[{k12_, k21_, k23_, k32_}] = {{v1[{k12, k21, k23, k32}][[1]], v2[{k12, k21, k23, k32}][[1]], v3[{k12, k21, k23, k32}][[1]]}, {v1[{k12, k21, k23, k32}][[2]], v2[{k12, k21, k23, k32}][[2]], v3[{k12, k21, k23, k32}][[2]]}, {v1[{k12, k21, k23, k32}][[3]], v2[{k12, k21, k23, k32}][[3]], v3[{k12, k21, k23, k32}][[3]]}};
λmat[{k12_, k21_, k23_, k32_}] = {{e-λ1[{k12, k21, k23, k32}]t, 0, 0}, {0, e-λ2[{k12, k21, k23, k32}]t, 0}, {0, 0, e-λ3[{k12, k21, k23, k32}]t}};
Inv[{k12_, k21_, k23_, k32_}] = Inverse[Tm[{k12, k21, k23, k32}]];
lvec[{k12_, k21_, k23_, k32_}] = {-λ1[{k12, k21, k23, k32}], -λ2[{k12, k21, k23, k32}], -λ3[{k12, k21, k23, k32}]}];
```

//Definition of β-dependence of the Wigner rotation matrix

```
d[β_] = {{((1 + Cos[β])/2)2, -((1 + Cos[β])/2) Sin[β], √(3/8) Sin[β]2, -((1 - Cos[β])/2) Sin[β]}, {(1 - Cos[β])/2 Sin[β], ((1 - Cos[β])/2)2, {(1 + Cos[β])/2 Sin[β], Cos[β]2 - (1 - Cos[β])/2, -√(3/8) Sin[2β], (1 + Cos[β])/2 - Cos[β]2, -((1 - Cos[β])/2) Sin[β]}, {√(3/8) Sin[β]2, √(3/8) Sin[2β], (3 Cos[β]2 - 1)/2, -√(3/8) Sin[2β], √(3/8) Sin[β]2}, {(1 - Cos[β])/2 Sin[β], (1 + Cos[β])/2 -
```

```

Cos[β]2,  $\sqrt{\frac{3}{8}}$  Sin[2β], Cos[β]2 - (1 - Cos[β])/2, -
((1 + Cos[β])/2) Sin[β], {(1 - Cos[β])/2)2, (1 - Cos[β])/2 Sin[β],  $\sqrt{\frac{3}{8}}$ 
Sin[β]2, (1 + Cos[β])/2 Sin[β], ((1 + Cos[β])/2)2}};

//Euler angles for each of the three structures for the two backbone sites
αcgc1 = -2.77753;
αskm1 = 0.7884694923884119`;
αfjx1 = 2.1825019941021275`;
βcgc1 = 2.71275;
βskm1 = 1.496575883609125`;
βfjx1 = 1.731052948112305`;

αcgc2 = 0.561229;
αskm2 = -1.112474570811166`;
αfjx2 = 0.31242019138561344`;
βcgc2 = 1.61544;
βskm2 = 1.9878124168417663`;
βfjx2 = 1.4814398220020453`;
A1 = {αcgc1, αskm1, αfjx1};
A2 = {αcgc2, αskm2, αfjx2};
B1 = {βcgc1, βskm1, βfjx1};
B2 = {βcgc2, βskm2, βfjx2};

P[{p1_, p2_, p3_}] = {p1, p2, p3};

//Populations of the two backbone sites in the sample
pmix1 = 0.5;
pmix2 = 1 - pmix1;

pop = {pmix1, pmix2};

Diff = 1.1*104; //Diffusion rate for the helical rotation
dflip1 = d[-B1];
dflip2 = d[-B2];

ωD = 76.76*2*Pi*106; //Deuterium Larmor frequency
ωQsq = 0.5*0.75*(174000*2*Pi)2; //Quadrupolar coupling coefficient in Equation
V.10

//Standard deviations (widths) of the parameter search Gaussians
sdpr = 0.05;
sdtau = 0.4;
sd = {sdpr, sdpr, sdtau, sdtau};

```

```

//Tolerance for the comparison to the experimental relaxation time
distsd = 1; (*in milliseconds*)
count = 1;

//Simulated annealing parameter (a multiplier that reduces the standard deviations of the search
Gaussians whenever a new parameter set is accepted)
simann = 0.99995;

flag = 1;
min = 1000;
imin = 0;

//Experimental relaxation time
Tlexpt = 46; (*in milliseconds*)

//Spectral density function as defined in Equation V.9 (with weighted averaging over the two
backbone sites in this example)
specdens[  $\omega$ _, {pvar1_, pvar2_, pvar3_}
, {kvar12_, kvar21_, kvar23_, kvar32_}] := Module[{Tmflip =
N[Tm[{kvar12, kvar21, kvar23, kvar32}]],
Invflip = N[Inv[{kvar12, kvar21, kvar23, kvar32}]],
lamflip = lvec[{kvar12, kvar21, kvar23, kvar32}],
Pflip = {pvar1, pvar2, pvar3}],

1/5*Sum[Pflip[[j]]*

$$\sum_{k=1}^3 \left[ \sum_{l=1}^3 Tmflip[[k, l]] * Invflip[[l, j]] * \right.$$


$$\left. \left( \sum_{a=1}^5 \left( \frac{(Diff * (-a + 3)^2 + lamflip[[1]])}{(Diff * (-a + 3)^2 + lamflip[[1]]^2 + \omega^2)} * \right. \right. \right.$$


$$\left. \left( pmix1 * Exp[n * (-a + 3) * (R1[[k]] - R1[[j]])] * dflip1[[3, a, j]] * dflip1[[3, a, k]] + \right. \right.$$


$$\left. \left. pmix2 * Exp[n * (-a + 3) * (R2[[k]] - R2[[j]])] * dflip2[[3, a, j]] * dflip2[[3, a, k]] \right) \right] \right], \{
j, 3}]

];
*****

//MCMC algorithm

For[rate=1, rate<9, rate++;

//Number of trials
ntrials = 2000;

//Upper and lower bounds for the search process for each choice of rate exponent
uptau = Log[10-(rate-1)];
lotau = Log[10-(rate+2)];

For[index1=0, index1<8, index1++;
For[index2=0, index2<9-index1, index2++;$$

```

```

//Initial choice of parameters (p1, p2, τ)
pres = {index1/10, index2/10, Log[1*10-rate]};
next = pres;
p1 = pres[[1]];
p2 = pres[[2]];
p3 = 1 - p1 - p2;
tau12 = Exp[ pres[[3]] ];

tau23=tau12;

//Rates as defined in Equation V.7
k12 = Sqrt[p1/p2]*(tau12)^-1;
k21 = Sqrt[p2/p1]*(tau12)^-1;
k23 = Sqrt[p2/p3]*(tau23)^-1;
k32 = Sqrt[p3/p2]*(tau23)^-1;

Pvec = {p1, p2, p3};
Kvec = {k12, k21, k23, k32};

//Spectral densities
specdens1=Re[specdens[ ωD, {p1, p2, p3}, {k12, k21, k23,
k32}]];
specdens2=Re[specdens[2*ωD, Pvec, Kvec ]];

//Relaxation time calculated for the current parameter set
T1pres = 1000*(ωQsq*(specdens1 + 4*specdens2))^-1;

//Deviation of current relaxation time from the experimental time
sqdevpres = (T1pres - T1expt)^2;

//Opening the accepted parameter files for each set of initial parameter choices
outputfile=StringJoin["C:\\Documents and Settings\\Drobny
Group\\Desktop\\DNA flip
folder\\BackboneNoIntQuad1_74sdt0_4\\three_site_flip_t",ToString[
rate], "_p", ToString[index1], ToString[index2], ToString[10-index1-
index2], ".txt"];
output = OpenWrite[outputfile];

For[i = 0, i < ntrials, i++;
flag = 1;
//Keep searching for next parameter set until one is found within the upper and lower bounds
While[flag == 1,
r = RandomReal[NormalDistribution[pres[[count]]],
sd[[count]] ] ];

next[[count]] = r;
sum = next[[1]] + next[[2]];
If[(count == 1) || (count == 2), If[(r<0)|| (r>1)|| (sum >
1), flag = 1, flag = 0], If[(r < lotau)|| (r>uptau), flag = 1, flag
= 0] (*outer If*));
];

```

```

    (*****)
//Same definitions of terms for the next parameter set as for the current parameter set above)
    next[[4]] = next[[3]];
    p1 = next[[1]];
    p2 = next[[2]];
    p3 = 1 - p1 - p2;

    tau12 = Exp[next[[3]] ];
    tau23 = tau12;

    k12 = Sqrt[p1/p2]*(tau12)^-1;
    k21 = Sqrt[p2/p1]*(tau12)^-1;
    k23 = Sqrt[p2/p3]*(tau23)^-1;
    k32 = Sqrt[p3/p2]*(tau23)^-1;
    Pvec = {p1, p2, p3};
    Kvec = {k12, k21, k23, k32};

    specd1=Re[specdens[  $\omega$ D, Pvec,Kvec ]];
    specd2= Re[specdens[2* $\omega$ D, Pvec, Kvec ]];

    T1next = 1000*( $\omega$ Qsq*(specd1 + 4*specd2))^-1;
    sqdevnext = (T1next - T1expt)^2;

    If[Mod[i,50] == 0,Print[i]; Print[sqdevpres, "
",sqdevnext]; Print[specd1, specd2]; Print[imin, "
", min]];
    (*****)

//Comparing the squared deviation away from the experimental relaxation time for the next
parameter relative to the current parameter set
    lratio =-(sqdevnext - sqdevpres)/(2*distsd^2);

//If the next set results in an improvement, this new set is accepted and becomes the current set
for the next iteration
    If[lratio ≥ 0, pres = next; sqdevpres = sqdevnext; sd *=
simann,
    trial = RandomReal[];
//If not, accept even if the difference is greater than the log of a randomly chosen number
between 0 and 1
    If[lratio ≥ Log[trial], pres = next; sqdevpres = sqdevnext;
sd*= simann]
];

fileout = pres;
AppendTo[fileout, sqdevpres];
AppendTo[fileout, i];
If[sqdevpres < min, min = sqdevpres; imin = i];

Write[output, fileout];

```



```
count++;
If[count == 4, count = 1];

]
  Print[imin];
  Print[min];
  Close[outputfile];
]
]
]
*****
```

Chapter VI

Constructing an RNA motional trajectory through energy-minimized structures

Abstract

Understanding the conformational variability of RNA molecules often requires the characterization of the coupled time-evolution of several interconnected flexible domains. While phenomenological modeling of the dynamics of individual sites using discrete-jump processes and continuous dynamics in simple potentials has yielded success, the elucidation of collective behaviours of molecules forces the models to become untenably complex. Moreover, in the current absence of proven molecular dynamics (MD) simulation packages for nucleic acids, the problem of obtaining atomic-level detail for the motion of multiple residues in a molecule stands out all the more starkly. In our current work, we propose a new methodology that utilizes energy-minimized structures generated by the program “Fragment Assembly of RNA with Full-Atom Refinement (FARFAR)” as representatives of the conformational exploration of the HIV-1 TAR RNA, and simulate solution T_1 and $T_{1\rho}$ relaxation times and NOEs for uridine residues in three distinct domains of the molecule using a single motional model.

1. Introduction

A complete description of the dynamics of a biological molecule requires a characterization of the various conformational states available to the molecule, both long- and short-lived. Experimentally capturing these transitory states of the molecule is complicated by the fact that the rate of transition may be too fast to allow for a comprehensive catalogue of all states,¹²¹ in addition to the inherent difficulty of “freezing-out” a particular state.^{122,123} Therefore,

an elucidation of dynamics often requires concurrent analytic and computational modeling. One such approach is to start with a phenomenological exchange model of a few conformational states extrapolated from NMR or X-ray experimental constraints, and fit a limited set of free parameters to experimental data that are sensitive to the dynamics such as NMR relaxation times and lineshapes.^{10,29,75,124} This procedure involves a method of guess-and-check where physical constraints on possible motions of the labeled residue(s) guide the model-building process. However, even this semi-analytic approach can quickly become complicated when further degrees of freedom and new free parameters are required by the model to fit the data.

Alternately, a purely computational approach to the description of molecular states would necessitate the elaboration of an accurate potential energy function (PEF), followed by either molecular dynamics (MD) or energy-minimization calculations. Molecular dynamics simulations are able to generate dynamic trajectories of the molecule and, in principle, represent the desired exploration of molecular parameter space, if sufficient numbers of trajectories are available (for examples, see work based on AMBER^{42-45,50} and CHARMM packages^{46,47,49}; Lindorff-Larsen et al⁴⁸). However, while simulations of protein dynamics have seen many successes recently, the extrapolation of dynamics in nucleic acids to time scales on the order of microseconds, where many conformational changes are expected to take place has proven to be difficult. Another alternative is energy-minimization. Energy-minimization techniques also rely on carefully setting up a PEF, and involve the subsequent alteration of the relative conformations of parts of the molecule in an iterative manner so as to find the global energy minima of the molecule.⁵²⁻⁵⁵

The basis for our dynamics simulation was the set of 500 lowest energy models generated by the group of Rhiju Das using the program FARFAR⁵⁴ for the 29-nucleotide HIV-1 TAR (Trans-Activation Response) RNA. Interest in the TAR RNA as an exemplary system arises

from the fundamental nature of the structural motif, where two helices are connected by a single-stranded bulge on one end and a backbone “hinge” at the other. Other RNA systems exhibit similar structures, such as the K-turns⁹⁹, or slight complications thereof, such as in the RRE (Rev-Response Element)^{125,126}. It is therefore worthwhile to characterize the dynamics of such a basic motif, especially in terms of the “large-scale” motions (as opposed to local single-residue motions) of one helix relative to another. Earlier work in our group used solid-state NMR data to suggest certain motions of the U38 residue in the upper helix,¹⁰ and recent work has characterized the motions of U40 and U42 in the lower helix {Huang et al 2012, *JACS*, in revisions}. Both models suggest that the helices are moving relative to each other at slow rates relative to the diffusion rates ($\sim 10^5 \text{ s}^{-1} - 10^6 \text{ s}^{-1}$). These results from solid-state data have motivated us to look at the distributions of the orientations of the upper helix relative to the lower helix among the set of lowest energy structures, where the relative orientations are characterized by the Euler angles of an upper helix-attached frame relative to a lower helix-attached one. The hope is that this reduction to a set of three angles sufficiently characterizes the dynamical range of this particular structural motif, while simplifying a potentially large-dimensional problem to a set of 3 modes. We have also included local base “librations”, i.e. rotations of the bases around a base normal (representing vibrations of the base around the equilibrium base-paired orientation) or rotations of the base around the glycosidic bond (for the single stranded bases) in the simulations.

NMR studies by Zhang, Hansen, Al-Hashimi and co-workers on TAR RNA have exploited the sensitivity of RDCs¹⁴ and of partial magnetic alignment⁸¹ to overall molecular conformation, in addition to using direct relaxation measurements^{12,13} in order to discern the

amplitudes of large-scale domain motions. Their conclusions will serve as a useful check for the results presented herein.

Another point worthy of mention is the difficulty in establishing conformers for single-stranded regions of RNA. A residue embedded inside a helical stem finds itself in a relatively restrictive free energy environment and physically reasonable arguments may be made to build possible dynamic models. However, the same does not hold for a residue in a single-stranded region, such as the bulge region of the TAR RNA. In fact, it was in response to the limitation of our ability to describe the changes in bulge conformation with changes in the upper helix orientation in our previous work^{75,124} that we sought the energy-minimized structures. Our hope is that the structures capture the relatively under-constrained motion of the bulge residues by including interactions with the solvent, and with adjacent and non-adjacent residues.

Using energy-minimized structures as the basis for a conformational exchange simulation of dynamics hinges on the assumption of the ergodic hypothesis: if the biomolecule under consideration has reached dynamic equilibrium, then an ensemble average of an observable over the states of the system is equivalent to the time-average of the observable. Thus, if this ergodic behaviour holds, then a sufficiently representative characterization of the energy landscape of the phase space will allow a calculation of the requisite time-averages of observables. The proposed methodology to be elaborated here additionally assumes that the conformational states obtained from the energy-minimization process covers such a representative fraction of the true phase space available to the molecule. In practice, the method uses the relaxation data itself as a means of checking the veracity of this assumption: we will be attempting to fit NMR data for multiple residues in a sample, and will set aside certain data sets as corroboration of the model fits made to the remainder.

However, the difficulty in merely acquiring a set of energy-minimized structures is that, while they mark the locations of energy minima, they do not provide any information about the free energy terrain of the multi-dimensional landscape that occurs between any pair of minima. Further developments in the energy minimization process could potentially provide valuable entropy information, which essentially is a measure of the shape of the free energy potential at a given location. At present, however, in the absence of such information, a dynamics calculation necessarily relies on the fitting of several phenomenological rate parameters to the data, a process that could very quickly become computationally laborious as well as inconclusive (when the number of free parameters significantly exceeds the available data) if the number of structures becomes large. Faced with this situation, it is to our advantage to find a means of pre-screening the structures. Several methods have been considered with respect to such a screening process, but the protocol that was adjudged best for our purposes is one where we utilize information gleaned from a visual inspection of the distributions of certain parameters of the system. Other candidate methods, and the issues associated with them, will be described briefly in the Discussion section in the hope that they may be applicable in the future with further investigations.

2. Methods

Structure generation was carried out using the program FARFAR and the simulation protocol as well as tests against other nucleic acid molecules are described in the original reference (Das et al⁵⁴). The torsion angles that were altered in the generation of the structures used here were those of the residues in the bulge and the bulge-adjacent base pairs of the two helical domains. The 500 lowest energy structures represented a distribution in energy of about

13 Rosetta units, where 1 Rosetta unit is approximately 1 $k_B T$ (Rhiju Das, private communication). In the first iteration of the procedure presented here, we did not utilize this energy distribution information to bias the probabilities, especially given a lack of explicit information regarding the entropies, and therefore of the free energies, of each of the states.

As mentioned earlier, solid-state¹⁰{Huang et al 2012, *JACS*, in revisions} and solution¹¹ experiments were performed on several residues in the 29 nucleotide HIV-1 TAR RNA. While the solution data includes T_1 , T_2 relaxation times and NOEs for ^{13}C -labeled residues at the C6 (pyrimidine) and C8 (purine) base sites and at the C1' sugar sites of many residues, the solid-state data so far has focused on ^2H -labeling the uridine residues in the sample at the C5 and C6 base proton sites: U40 and U42 in the lower helical stem, U23 and U25 in the bulge, and U38 in the upper helical stem. In the following, we therefore focus on the uridine residues in the sample. The method itself is perfectly general and may be extended to any nucleotide and any site in a sample.

The method is as follows:

- a) Define the upper and lower helical axes for all structures using the program 3DNA¹⁰⁶. The upper helical axis is taken to be the local helical axis of the A27-U38::G28-C37 dinucleotide step, where the base-pairs flanking non-helical regions were excluded due to possible distortions from helical structure. The lower helix is calculated as the average over the G17-C45::G18-C44, G18-C44::C19-G43 and C19-G43::A20-U42 dinucleotide steps. In both cases, we avoided including those residues whose parameters were varied as a part of the energy minimization process so that we consistently used the same frame to represent the lower helix and treated the upper helix as purely a rigid object moving as a result of

changes in and around the bulge. The impact that this choice of helix definitions has on the results may be explored in the future.

- b) Define the lower helix coordinate frame (LHF) by choosing the helical axis as the z-axis and the perpendicular from the z-axis to the G43 C8 atom as the y-axis.
- c) Calculate the α angles of each of the structures as the angle between projection of the upper helical axis onto the xy-plane of the LHF and the x-axis of the LHF.
- d) Calculate the β angles as the angle between the upper helical axis and lower helical axis.
- e) The γ angle is defined by the orientation of the U38 C6H6 bond about the upper helical axis. Extracting this information from the structures requires first removing the α and β dependences by rotating the original U38 C6H6 vector $\vec{v}_{C6H6}$ about the fixed LHF

axes: $\vec{v}'_{C6H6} = \vec{R}_{y_{LHF}}(-\beta)\vec{R}_{z_{LHF}}(-\alpha)\vec{v}_{C6H6}$. The resultant vectors are distributed around the LHF z-axis as a function of their γ angles.

The Euler angles described above are related to the domain motions as shown in Figure 1.

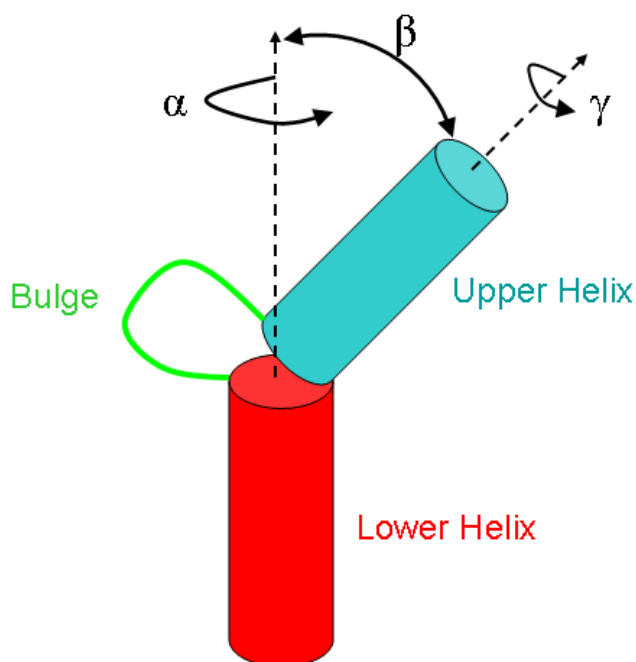


Figure VI.1: Relation between the Euler angles determined from the atomic coordinates and the associated domain motions.

- f) Visually inspect the distributions of structures as a function of each of the angles and bin the structures. If it turned out that there were significant clusterings of structures at particular values of the Euler angles, we could proportionally weight the probabilities of such angular positions. The hope is that inspecting such distributions would ultimately provide sufficient information regarding the entropy as a function of the coarse-grained parameter set (in this case the Euler angles of the upper helix relative to the lower helix).
- g) Propose a jump model using selected structures. The selection involves sufficiently sampling the three Euler angle degrees of freedom. A general $N \times N$ jump matrix for a set of N structures can get computationally expensive quickly, and it is helpful to try and minimally represent the diffusive trajectory. To that end, we have attempted to bin the three angular distributions, and defined an m_i -site jump for each degree of freedom i . We then assumed that the jumps in the three Euler angles are uncorrelated and occur only between

nearest neighbours $\{\alpha_i, \beta_i, \gamma_i\} \rightarrow \{\alpha_j, \beta_j, \gamma_j\}, j = i-1, i, i+1$. The uncorrelated assumption was made based on visually inspecting the angular distributions. The β -angle-color-coded plots seem to show that there is no bias in either the α or γ angle based on the value of β angle. Additionally, binning the α angle allowed a preliminary check on the correlation with the γ ; while there does seem to be a slight inter-dependence, the number of structures begins to become sparse when binned both in α and β making it difficult to fully characterize this correlation. Therefore, as a first try, we assumed no correlation, an assumption that may change in future iterations of the method. To elucidate the above statements, we provide some of the parameters for one of the first iterations. The β angle was first sorted into 8 bins each of width 10° , covering the range from 0° to 80° . This sorting included 472 of the 500 structures. The structures in each of these bins were further sorted into 13 bins of width 20° as a function of the α angles from 30° to 290° (this reduced the number of included structures to 467). The rotation in the α angle was treated as a three-site jump, with a conformer near each end of the angular arc swept out by the structures (as seen from Figure 2 below), and one structure in the middle. For each of the $\{\alpha, \beta\}$ bins (7×3 bins in total), we chose one or two structures to represent the spread the γ angle. As mentioned earlier, the number of structures in each of the $\{\alpha, \beta\}$ bins could start to become sparse, especially towards the ends of the trajectories. To properly represent the full sweep in the α angle, we were forced to choose the first bins at each end where there was at least one structure, therefore resulting in only one γ structure in two instances. The net result was a jump-process involving 40 structures. The exchange matrix for such a setup was a sparsely populated 40×40 matrix. Visual inspection of Figures 1-3 also motivated the use of uniform equilibrium probabilities. Thus, for a structure in a bin with two γ sites the equilibrium

probability was $= (\text{Number of } \beta \text{ sites} \times \text{Number of } \alpha \text{ sites} \times \text{Number of } \gamma \text{ sites})^{-1} = (7 \times 3 \times 2)^{-1} = 1/42$. We are currently also considering the possibility of using the RDC measurements described above to better inform our choice of the probability distribution.

- h) Include the local motions of the bases into the jump simulation. Solid-state data have shown that in addition to the larger helical domain motions, the bases also undergo small amplitude, faster motions. Previous work in our group¹⁰ has indicated that the helical residue U38 experiences a base-libration of amplitude $\pm 4^\circ$ at a rate of $2.15 \times 10^8 \text{ s}^{-1}$, while recent work {Huang et al 2012, *JACS*, in revisions} ascribes a libration of $\pm 7.5^\circ$ at a rate of $1.2 \times 10^8 \text{ s}^{-1}$ to the U40 and U42 bases in the lower helix. The single-stranded bulge residue U23 undergoes, in addition to a slower hop, a $\pm 11^\circ$ amplitude rotation about the glycosidic bond at a rate of 10^{10} s^{-1} .¹²⁷ In all these cases, the two-sites in the librations or the glycosidic rotations have been weighted equally in the data simulations. This would imply an additional factor of $\frac{1}{2}$ for the equilibrium probability of each site in the jump process, corresponding to the probability of being in any one of two possible local orientations. In order to incorporate these additional motions, which are assumed to have a negligible impact on the overall molecular configuration given the small amplitudes, we simply double the number of exchanging conformers in the simulation by including a single lowest-energy structure twice with only the local base orientations changed in each. If a local motion is deemed to be significantly faster than all other relevant rates in the system, the pre-averaged orientation of the base site may be used instead, thereby halving the dimensions of the jump matrix. In general however, in the absence of such simplifications, the above method must be used.

i) Calculate the solution relaxation times given the jump-matrix chosen above. To carry out this calculation we use the results of our recently published paper¹²⁴ (the theory presented in Chapter IV) where we allow for an exchange between a discrete set of conformers with different diffusion tensors in addition to the overall rotational diffusion undergone by any single conformer. The inputs to this process are the three Euler angle exchange rates $\{r_\alpha, r_\beta, r_\gamma\}$ and the outputs include the T_1 and T_2 relaxation times, as well as the NOEs. Using the expressions directly from our previous work can be computationally expensive if the number of conformers grows large. Therefore, in order to calculate the spectral densities needed in the calculation of the relaxation times and NOEs, we use a Gauss-Laguerre quadrature of the appropriate cosine Fourier transform of the correlation functions. The Gauss-Laguerre integration method calculates the integral over a function approximately as a sum over discrete values of the integrand evaluated at the roots of Laguerre polynomials and weighted by particular values. We chose to use 100-150 roots in our simulations, with an increase in the number of evaluation points naturally resulting in an increase in computational time. In addition, the solution to the time evolution equation for the transition probability (probability that the molecule will be in a given Euler angle orientation $\vec{\Omega}_t$ relative to the lab frame at a time t given its orientation at $t = 0, \vec{\Omega}_0$) requires the input of the rotational diffusion tensor eigenvalues for each conformer. We obtained the rotational diffusion tensor eigenvalues and eigenvectors for each structure using the program HYDRONMR⁹². The input parameters to HYDRONMR include the atomic element radius (AER), which is the radius of the beads used to represent the van der Waals' radii and hydration shells of the non-hydrogenic atoms of the structure. This value was chosen to be 2.3 Å as in previous simulations^{75,124} and the choice is more extensively

evaluated in those papers. Finally, it must be borne in mind that the diffusion-and-exchange expressions presented in our previous work¹²⁴ assume that the diffusion tensors of any pair of exchanging partners are momentarily collinear at the instant of exchange. This assumption will fail if the molecule radically changes shape between any two partners, as the instantaneous jump in angular orientation of the diffusion tensor will not be accounted for by our expressions. The discrepancy is mitigated by discretizing the Euler angular space into a sufficient number of sites and by allowing jumps only between nearest neighbours whose diffusion tensors are not drastically different. In our current simulations, we do not evaluate this discrepancy explicitly.

- j) Finally, it is possible to iteratively fit the experimental solution NMR data by incorporating the relaxation time calculation into a Markov-chain Monte Carlo (MCMC) procedure. The fit parameters we have considered so far have been the Euler angle jump rates (it is possible to include even equilibrium probabilities in the same procedure, but this would significantly increase the number of free parameters in the problem). We have also attempted to independently corroborate the jump model itself by fitting to only a fraction of the data, and checking to see if the remainder of the data can be fit using the same parameter set. Specifically, we have considered the relaxation times of all uridine residues in TAR, U40 and U42 in the lower helix, U23 and U25 in the bulge, and U38 in the upper helix, but fit the rates only to the U38 T_1 , T_2 and NOE values.

3. Results

We first present the results of plotting the distributions of the structures with respect to the Euler α and β angles in Figures 2A and 2B. The two figures represent the same distribution

when viewed from different perspectives in order to highlight the two Euler angle dependences.

A few points are worthy of notice in Figure 2.

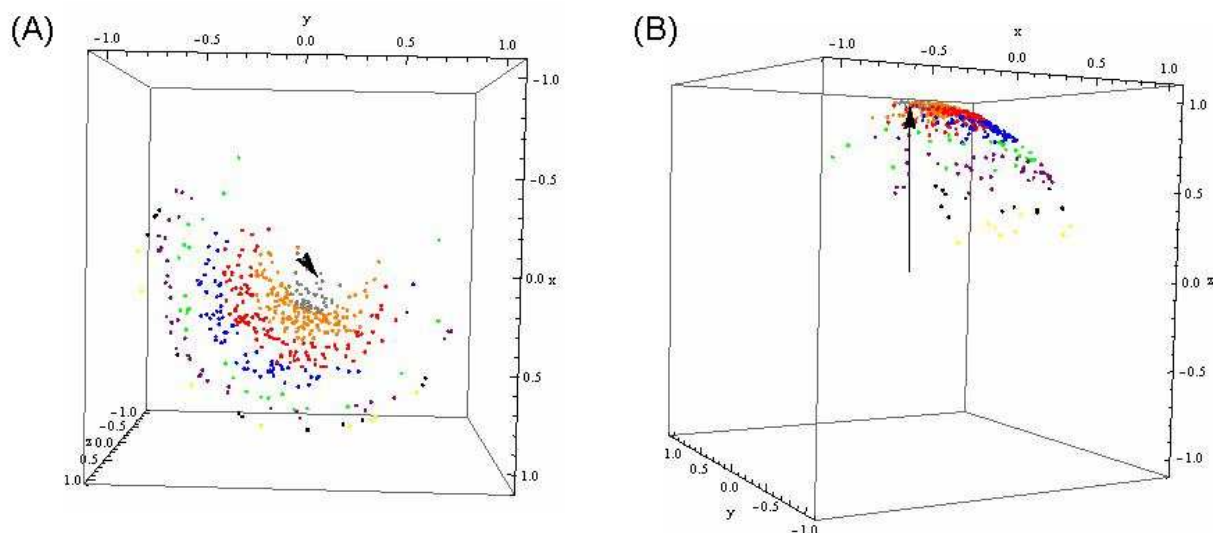


Figure VI.2: The distributions of upper helix unit vectors for the structures relative to the lower helix (orientation of the lower helix indicated by the arrow). The upper helix vectors are indicated by the end points of the unit vectors, with the origin located at the center of the 3D graph. The colors represent the different bins for the β angle between the upper helix and the lower helix, in 10° increments. The two panels are: (A) α angle dependence, as viewed from the top down along the lower helix; and (B) β angle dependence, as viewed from the side profile.

In Figure 2A, the α angle sweep does not cover the full 360° , implying that the upper helix seems to be restricted to move in a limited cone about the lower helical axis, or conversely, that the lower helix has a restricted freedom of twist relative to the upper helix. Momentarily ignoring the considerations of how well the energy-minimized structures represent the available phase space and of the sequence dependence, these distributions may be providing information on the how much motion is possible from this particular structural motif.

The same type of information may be drawn from the restricted nutation of the upper helix shown in Figure 2B. This may just be a simple representation of the steric hindrance of the upper helix by the lower helix.

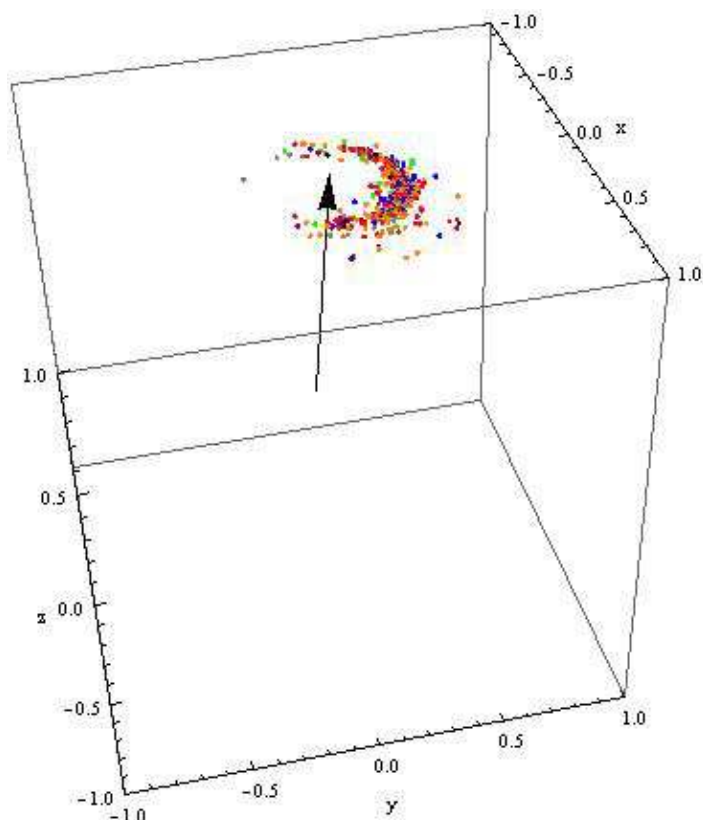


Figure VI.3: The distributions of U38 C6-H6 unit vectors in a frame where the α and β dependences of the upper helix were transformed away. The C6-H6 vectors are indicated by the end points of the unit vectors, with the origin located at the center of the 3D graph. The colors represent the different bins for the β angle between the upper helix and the lower helix, in 10° increments.

In Figure 3 we present the γ angle distributions of the U38 C6-H6 bond, after transforming to a frame where α and β angles of the upper helix are 0. After the transformation, the remaining angular sweep can be attributed purely to the γ angle, with the lower helix still represented by the arrow in the figure. Note that the polar angle relative to the lower helix is fairly stable across all structures, merely reflecting the restriction of the U38 C6-H6 orientation relative to the upper helix to a sharply defined value.

We also studied the distributions of the C6-H6 bond vectors of the bulge uridine residues, U23 and U25, for patterns and correlations with the Euler angle distributions described above. Searching for these patterns was one of the primary goals of using energy-minimized structures,

due to the difficulty in characterizing single-stranded regions. The distributions are shown in Figure 4.

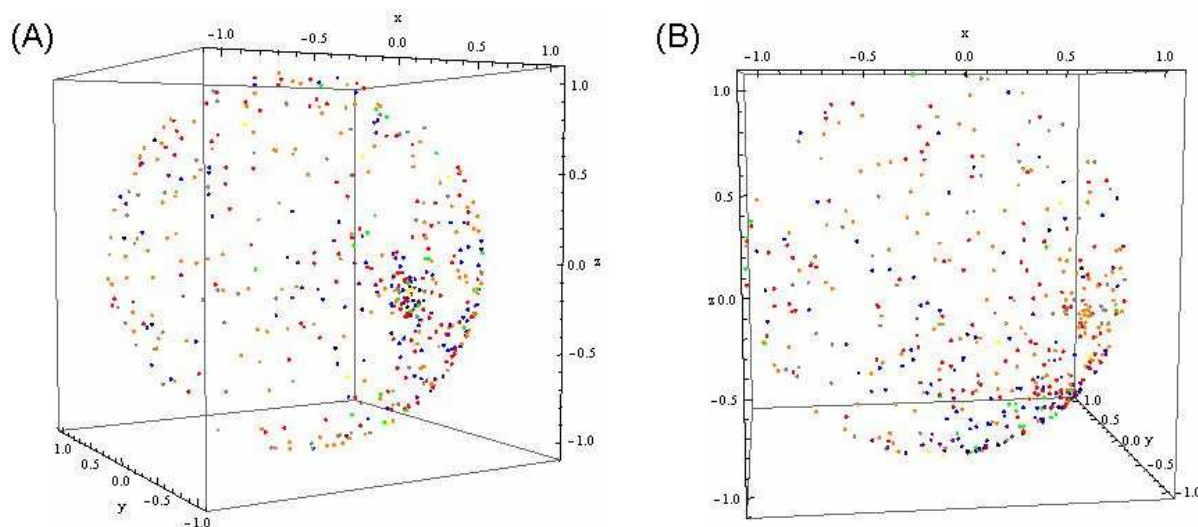


Figure VI.4: The distributions of bulge uridine C6-H6 bond unit vector on the unit sphere, relative to a lower-helix-affixed frame: (A) U23 C6-H6 bond distributions; and (B) U25 C6-H6 bond distributions. The origin is at the center of the 3D graph. The colors represent the different bins for the β angle between the upper helix and the lower helix, in 10° increments.

The U23 (Figure 4A) and U25 (Figure 4B) bond vector distributions seem to be scattered all over the unit sphere. This seems to indicate strongly that these residues are very under-constrained in their conformations. There is a small degree of ordering in both with small concentrations of structures at certain orientations, but, in general, the bonds seem to be distributed almost isotropically. However, it is important to note that these distributions are relative to a coordinate frame affixed to the lower helix, most of whose degrees of freedom were not altered in the course of generating the structures. It is still possible that there is a correlation between orientations of these bonds and those of other residues in the bulge or helices. To that end, some preliminary investigations are being made as to the relative orientations of these residues with the A22 (lower helix) and A27 (upper helix) bases.

As mentioned above, we used this information to posit a particular jump model, where 7 of the 8 β angle bins were used to represent nutation jumps, and 3 α bins (two end points and one mid-point) were selected for each such β choice. Finally, within each such α bin, one or two structures were chosen based on their γ angle values, where if available, the two structures were at the extrema of γ angles within that bin.

Table VI.1: Relaxation times and NOEs from fit to the solution relaxation data, for the case where the local librations and glycosidic rotations are not included. The relaxation data for U25 is taken to be similar to that of C24, as such data was not available from the solution reference (Bardaro et al¹¹).

Residue	T₁ fit (in ms)	T₁ expt. (in ms)^a	T₂ fit (in ms)	T₂ expt. (in ms)^a	NOE fit (unitless)	NOE expt. (unitless)^a
U38 (data used for fit)	354.0	354.1	24.8	24.6	1.12	1.14
U23	364.0	328.3	24.0	29.1	1.12	1.32
U25	366.3	N/A; ~321 (from C24)	23.9	N/A; ~35 (from C24)	1.12	N/A; ~1.35 (from C24)
U40	355.9	363.7	24.6	25.2	1.12	1.15
U42	352.3	360.5	24.9	24.8	1.12	1.15

^aData taken from Bardaro et al¹¹.

We first ran the program without the internal motions included, to see if the structures could be simulated purely by their large-scale domain motions. As mentioned in the Methods section, the data was fit only for U38, as a corroborative attempt to see if the chosen model succeeds in modeling the data for the other residues. The best match to the experimental data so far was obtained for the rates of each type of motion

was $\{r_\alpha, r_\beta, r_\gamma\} = \{8.64 \times 10^5 \text{ s}^{-1}, 9.80 \times 10^5 \text{ s}^{-1}, 8.54 \times 10^5 \text{ s}^{-1}\}$. The resulting relaxation times and NOEs are given in Table 1. For reference, the experimental error bars are 3.2 ms for T₁ measurements, and 0.5 ms for T₂ measurements.

The results shown in Table 1 indicate that the relaxation times are fairly close to the experimental times for the helical residues. This is not particularly surprising given that the characteristic diffusion rates are similar for most of the conformers and the rigid rotation of any single conformer on its own produces similar relaxation times. At the same time, the observed discrepancies in the U23 and U25 values, as well as those in the U40 and U42 residues, could be explained by the differences in the local motions of the residues, such as glycosidic rotations or base-librations. The next step was therefore to incorporate such motions into the simulation and to see if indeed the solid-state data-derived values hold up in solution as well.

Table VI.2: Relaxation times and NOEs from fit to the solution relaxation data, for the case where the local librations and glycosidic rotations are included. The relaxation data for U25 is taken to be similar to that of C24, as such data was not available from the solution reference (Bardaro et al¹¹).

Residue	T ₁ fit (in ms)	T ₁ expt. (in ms) ^a	T ₂ fit (in ms)	T ₂ expt. (in ms) ^a	NOE fit (unitless)	NOE expt. (unitless) ^a
U38 (data used for fit)	346.2	354.1	24.9	24.6	1.12	1.14
U23	386.5	328.3	25.8	29.1	1.18	1.32
U25	390.9	N/A; ~321 (from C24)	26.4	N/A; ~35 (from C24)	1.19	N/A; ~1.35 (from C24)
U40	375.3	363.7	26.1	25.2	1.17	1.15
U42	368.5	360.5	26.0	24.8	1.16	1.15

^aData taken from Bardaro et al¹¹.

The results of the fitting procedure after the local motions were included are shown in Table 2. It is important to emphasize, however, that these results are very preliminary. Due to the computationally intensive nature of the program, the algorithm had not run through a significant number iterations at the time these values were extracted, and the resulting best-fit rate values of $\{r_\alpha, r_\beta, r_\gamma\} = \{1.00 \times 10^5 \text{ s}^{-1}, 7.7 \times 10^4 \text{ s}^{-1}, 4.7 \times 10^4 \text{ s}^{-1}\}$ barely had an opportunity to explore far

from the seed rate values of $1.00 \times 10^5 \text{ s}^{-1}$ (for all three rates). Nonetheless, it may help to remark on these preliminary results.

The helical residues show larger deviations from experiment than the values shown in Table 1 for the case of no local motions. Possible reasons for this in addition to the ones stated above (results too premature and deficiencies of the structures) are that the probabilities for the various orientations may not be uniform, and that the solid state NMR-derived local motion parameters are not unique, with some spread in the best-fit values to the solid state data.

The U23 and U25 simulations exhibit greater deviations in both relaxation times and the NOEs. Given the relatively unconstrained motions expected of the bulge residues, it is likely that this is signaling the need for a different exchange rate among some or all of the conformations of these residues. When the orientations of the U23 and U25 were plotted for the structures, it was noticed that they both explored a large portion of the unit sphere (see Figures 4 and 5, for the distributions of U23 and U25 C6H6 unit vectors, respectively) except for a few non-committal clusterings. Therefore, it would be difficult and even incorrect to surmise a simpler jump model set until the distribution can be further explored. If further studies do not reveal any significant correlation between the Euler angle jumps of the helices and the orientations of the bulge residues, then we might be able to treat the motion as an independent, nearly free diffusion process. In any case, the data seems to indicate that more rapid jumps are required for the bulge residues as evinced by the larger NOEs, which are sensitive to faster rates. As stated above, future work will attempt to correlate the U23 orientation to frames attached to both A22 and A27 to check for solid-state model-motivated correlations.

4. Discussion

The methodology presented here is a work in progress. However, it is hoped that the basic framework is sound, and that this protocol would be a small step forward in the simulation of large scale motions in biological molecules in general, and in nucleic acids in particular. The algorithm could potentially allow modelers to simultaneously simulate the dynamics of multiple residues in various structural contexts, such as helical domains, loop regions and single-stranded regions. This would represent utilization of the site-specificity of NMR data to describe full-molecule restructurings. Moreover, the use of energy-minimized structures may enable modelers to bypass the difficulties of long-time molecular dynamics.

Many reasonable modifications to the simulations presented here may be proposed and easily implemented. Firstly, one may choose to vary the equilibrium probability distributions of the structures based on their inter-domain orientations, to see if there are preferred conformations. Secondly, the number of points along the trajectory may be increased, albeit at the cost of computational time. Finally, a larger number of energy-minimized structures can be incorporated into the analysis and the trajectories observed here can be corroborated or deemed incorrect.

Prior to settling on the method described herein, we considered other methods of distilling motional information from the energy-minimized structures. A description of these is presented below both for completeness, and in case some future value may be derived from the work.

1. **Solid-state NMR REDOR constraints:** REDOR (Rotational Echo Double Resonance) is a means of measuring the dipolar coupling strength between hetero-nuclei, whereupon the average distance between the two spins may be inferred. Two such measurements were

made on solid-state TAR samples: (a) between a ^{19}F at the 2' position of U23 and a ^{31}P label at the 5' position of A27, yielding an average distance of $10.3 \text{ \AA}^{127}$; and (b) between a ^{19}F label at the 5 position of U42 and ^{31}P label at the 5' position of C39, yielding an average distance of $10.9 \text{ \AA}^{128}$. These values would potentially allow us to filter out some of the structures from the entire collection. While these provide very stringent constraints (the error bars on both the measurements are on the order $0.6 \text{ \AA}$), it must be borne in mind that the samples in both cases were lyophilized, i.e. results were for the unhydrated case. We loosened the restrictions on the distances by allowing for the molecule to toggle between the unbound distances and those in the TAR-Tat peptide bound complex, where measurements yielded values of (a) $6.6 \text{ \AA}$ and (b) $6.8 \text{ \AA}$, respectively. This choice of allowed parameter space followed from arguments that the free TAR molecule dynamically exchanges between the unbound conformation and a binding-receptive conformation (the so-called “conformational capture” mechanism)^{14,15}. However, even this was considered to be potentially too restrictive given that the samples were lyophilized and did not represent sufficiently the full range of motion that would be available to the molecule in solution.

2. **Solid-state jump models:** Previous work on solid-state samples in our group has suggested various dynamic models for specific residues¹⁰ in TAR. Given the physical intuition gained from solid-state conditions, we have carried out simulations to see if the solution NMR data could also be fit by the same models incorporated into solution structures, with overall rotational diffusion added^{75,124}. The work in these two references was performed with structures that were modified from the lowest energy NMR structure (PDB: 1ANR⁸) by the alteration of multiple parameters, including direction and amount of bending of the upper helix relative to the lower helix, twist of the upper helix about its own axis of symmetry,

and hydrodynamic parameters. However, the modification process was rudimentary in that the bulge was kept fixed in a single conformation and some steric clashes with the upper helix were allowed as the bend and twist were varied. In the current work, it is our intention to utilize the energy minimized structures to properly account for the alteration of bulge conformation coincident with changes in the relative orientation of the upper helix and lower helix. Therefore, in order to test the validity of solid-state models for bulge residues using the 500 lowest energy structures, we applied the physical arguments gleaned from the solid-state modeling process to filter out structures. More specifically, we considered the U23 bulge residue in conjunction with the U38 upper helical residue as follows: Olsen et al found a best-fit model for U23 (deuterium labeled at the C6-H6 site) where the residue experienced a slower 24° jump between two-sites at a rate of $6.7 \times 10^7 \text{ s}^{-1}$, in addition to a faster (10^{10} s^{-1}) rotation of $\pm 11^\circ$ around the glycosidic bond, and the two sites of the slower “hop” were interpreted as (a) a stacking conformation with A22 in the lower helix; and (b) a base-triple formed with the A27-U38 base pair. We utilized this information by sifting through the structures for the two cases where (a) the relative orientation of the U23 base normal and the A22 base normal was less than 50° and the centroid-to-centroid distances between the two is $\leq 4.5 \text{ \AA}$, to represent the base-stacking scenario; and (b) the relative orientation of the U23 base normal and the A27 base normal was within 50° , with the distance between the U23 O4 and A27 N6 atoms, as well as between the U23 N3 and A27 N7 atoms being allowed to vary by up to $8 \text{ \AA}$ more than the hydrogen bonding distance of $3 \text{ \AA}$ established for a Hoogsteen-edge pairing,¹²⁹ to represent the U23-A27-U38 base triple formation. Furthermore, as the solid-state models predicted a slower two-site jump ($1.38 \times 10^6 \text{ s}^{-1}$) for the U38 residue in the upper helix, in order to simultaneously represent

the motion of both residues, we attempted to find two pairs of stacking-base-triple conformations for U23 each of the conformations of U38 (which were separated by a 9° bend and 15° twist in the upper helix orientation). While we had to stretch the limits of the definitions of stacking and base-triple interactions in order to find four such structures, in the end this method was shelved on account of the same concerns as for the REDOR case: in a solid-state sample, the packing of the molecules in close proximity may have an impact on both the range and rate of the dynamics, especially for large-scale motions. This by no means invalidates the use of solid-state NMR to infer motions, as local, small-scale motions occurring in rate regimes invisible to solution NMR may be unhindered by inter-molecular interactions and so show up faithfully in solid-state NMR data. However, the proposed motions here involve the bulk motion of a large portion of the molecule and so could be impacted by nearest neighbour interactions. Until the fidelity of large-scale motional parameters derived from solid-state NMR can be proven in a solution NMR context, it is difficult to warrant their use as filters.

3. **Torsion-angle binning and coincidence:** In our continuing efforts to simplify the dynamics fitting problem, we tried to search for coincidences in sets of torsion angles for U23, C24 and U25 (the bulge residues) among the structures so as to identify possible preferred conformations of these residues. However, since the 7-torsion angle backbone “suite”¹³⁰ only accounts for the local orientation of the backbone, in order to establish repeated global conformations of a residue relative to the rest of the molecule, we also considered the non-hydrogen atom root-mean-squared deviations (RMSD) between each pair of structures for the entire bulge. The criteria have usually been set at a match of 5 or more out of the 7 backbone torsion angles, and an RMSD between bulges of ≤ 2.5 Å.

Attempts at establishing such fixed conformational states did not produce clusters for C24 or U25. Interestingly, there does seem to be a grouping of structures that exhibit similarity in 5 or even 6 out of the 7 torsion angles for U23. While establishing a single such conformation is interesting in light of the stacking and base-triple interpretations based on solid-state data, we are also curious to understand the correlation of these structures with the motion of the upper helix relative to the lower helix. This is related to the correlation studies of U23 and U25 described under the Results section.

4. **Solution Residual Dipolar Coupling (RDC) constraints:** A molecule in a sufficiently dilute solution will undergo unhindered, isotropic overall rotational diffusion. From the perspective of an NMR experiment, this means that the interactions between spins or between a spin and its environment are averaged on a time scale corresponding to the characteristic time of diffusion. Thus, any experiment that seeks to determine the strength of a particular interaction would need to be sensitive to time scales shorter than the diffusion time scale. Alternately, it is possible to partially hinder the rotation of the molecule by preferentially aligning it, using either its own magnetic susceptibility^{81,131} or an aligning medium, and then measure the residual interaction of interest that results from incomplete averaging over orientations. Aligning media used for this purpose include liquid crystalline media¹³², Pf1 filamentous bacteriophage¹³³, stressed polyacrylamide gels¹³⁴ and PEG/Hexanol mixtures¹³⁵ among others. In order to support the choice of a subset of structures to be used for the multi-site jump simulations, we are using experimental RDCs measured {Bardaro et al, unpublished data} in a variety of aligning media in conjunction with the predictive program PALES¹³⁶. Currently, we are only considering the data from PEG/Hexanol as this uncharged media aligns the charged RNA only through steric

hindrance of the overall rotation. This simpler case does not require a proper characterization of the RNA charge density, as would be required for charged alignment media (a case which PALES has been extended to solve). Extension to the charged media data may be considered in the future. In what follows I provide a quick summary of the procedure being used: The Saupe matrix or alignment tensor, $A_{ij} = \frac{3}{2} \langle \cos \xi_i \cos \xi_j \rangle - \frac{1}{2} \delta_{ij}$, is a function of the orientations of the molecular coordinate frame relative to the magnetic field axis, where the $\cos \xi_i$'s are the direction cosines of the magnetic field axis in the molecular frame, averaged over the allowed solid angle space for the molecule in the presence of the aligning medium. The purely steric version of PALES calculates the alignment tensor of the molecule by the sampling of multiple allowed orientations in the presence of a flat (infinitely large) obstruction. The alignment tensor eigenvalues $\{A_{xx}, A_{yy}, A_{zz}\}$ are then used in conjunction with directional information for a particular bond relative to the alignment tensor principal axis frame (PAS). It is important to note that the bond in this case is assumed to be static relative to the molecular frame. Bond motion will be considered subsequently as an average over several static conformations. The resulting RDC is given by the

$$\text{expression } D_{ij}(\theta, \phi) = -\frac{\gamma_i \gamma_j \mu_0 h}{8\pi^3 r_{ij,eff}^3} (A_{xx} \sin^2 \theta \cos^2 \phi + A_{yy} \sin^2 \theta \sin^2 \phi + A_{zz} \cos^2 \theta),$$

with γ_i representing the gyromagnetic ratio of the atomic species i . The RDCs can be computed by this method for multiple bonds in several residues. Finally, to include the possibility of dynamics, we use the weighted average of the RDCs of each conformer

$$RDC_{Total} = \sum_{i=1}^{N_{Conformers}} p_i RDC_i \text{ where the individual RDCs are measured on a time scale much}$$

longer than any dynamic time scales in the problem (which motivates the use of the equilibrium probabilities of the conformers in the above formula). In this way, it is possible to check to see if a given jump model is consistent with observed RDCs to within the experimental error. Moreover, the problem may be inverted to find the best-fit equilibrium probabilities given experimental RDCs, assuming the choice of conformers and their RDCs are believed to be physically representative. Both these methods may as yet prove to be useful and will be considered in the future.

5. Conclusions

We have presented a potential means of utilizing energy-minimized structures of RNA as representatives of the conformational landscape of the molecule, and have tested the validity of such a model against solution relaxation times and NOEs. The results have proven promising for fitting the data for helical residues in both stems, but the model does need further modification in order to describe the motion of the bulge residues. The answer may lie in the inclusion of additional, independent motions for the bulge residues. The procedure attempts to incorporate the motions of various parts of the molecule into a unified trajectory, with the hope of building an intuition of the conformational range of the structural motif using the properties of the energy-minimized structures in concert with experimental data.

Appendix VI.A: Mathematica code for the jump rate parameter search

algorithm using the energy-minimized structures

//Parameters of the jump matrix for each of the five sites (all have equally weighted base-libration or glycosidic rotations)

p1U = 0.5;
p2U = 1 - 0.5;
rU38 = 2.15*10⁸;
JumpU38 = 4;

rU40 = 1.2*10⁸;
JumpU40 = 7.5;

rU42 = 1.2*10⁸;
JumpU42 = 7.5;

rU25 = 10¹⁰;
JumpU25 = 11;

rU23 = 10¹⁰;
JumpU23 = 11;

//Gauss-Laguerre integration:

//Number of roots calculated

n=250;

//Array of roots

t=Array[N[Root[LaguerreL[n,x],#]]&,n]

//Coefficients in the summation

newmod = Table[(Exp[t[[k]]]*t[[k]])/(((n+1)*LaguerreL[(n+1),t[[k]]])²), {k,1,n}]

//Size of the time step used in the discretized integration

delT = 10⁻⁹;

//NMR parameters

γC = 6.728*10⁷; *//Gyromagnetic ratio of Carbon-13*

γH = 2.675*10⁸; *//Gyromagnetic ratio of Hydrogen*

rCH = 1.1*10⁻¹⁰; *//Carbon-hydrogen bond length for an aromatic carbon atom*

δσ = 212*10⁻⁶; *//Chemical shift anisotropy (CSA) for a uridine base carbon atom*

//Magnetic field strength

Bfield = 11.74;

//Larmor frequencies of Carbon-13 and Hydrogen

$\omega_C = \gamma_C \cdot B_{\text{field}}$

$\omega_H = \gamma_H \cdot B_{\text{field}}$

$mval = \omega_C - \omega_H;$

$pval = \omega_C + \omega_H;$

$co2 = (2/15) \cdot (\delta\sigma^2) \cdot \omega_C^2;$

$d2 = (0.1 \cdot 10^{-14}) \cdot ((1.05451 \cdot 10^{-34}) \cdot \gamma_H \cdot \gamma_C / r_{CH}^3)^2;$

//Same functions as in Appendix IV.A

$\text{sig}\{Dx_ , Dy_ , Dz_ \} = Dx^2 + Dy^2 + Dz^2 - (Dx \cdot Dy + Dy \cdot Dz + Dz \cdot Dx);$

$\text{num2}\{Dx_ , Dy_ , Dz_ \} = 0.5 \cdot ((Dx + Dy - 2 \cdot Dz) + 2 \cdot \text{Sqrt}[\text{sig}\{Dx, Dy, Dz\}]);$

$\text{num3}\{Dx_ , Dy_ , Dz_ \} = 0.5 \cdot ((Dx + Dy - 2 \cdot Dz) - 2 \cdot \text{Sqrt}[\text{sig}\{Dx, Dy, Dz\}]);$

$\text{den2}\{Dx_ , Dy_ , Dz_ \} = \text{Sqrt}[2 \cdot \text{sig}\{Dx, Dy, Dz\}] + \text{Sqrt}[\text{sig}\{Dx, Dy, Dz\}] \cdot (Dx + Dy - 2 \cdot Dz);$

$\text{den3}\{Dx_ , Dy_ , Dz_ \} = \text{Sqrt}[2 \cdot \text{sig}\{Dx, Dy, Dz\}] - \text{Sqrt}[\text{sig}\{Dx, Dy, Dz\}] \cdot (Dx + Dy - 2 \cdot Dz);$

$\text{fla}\{Dx_ , Dy_ , Dz_ \} = \{ \{ -(2^{-0.5}), 0, 0, 0, 2^{-0.5} \}, \{ 0.25 \cdot \text{Sqrt}[6] \cdot (Dx - Dy) / \text{den2}\{Dx, Dy, Dz\}, 0, \text{num2}\{Dx, Dy, Dz\} / \text{den2}\{Dx, Dy, Dz\}, 0, 0.25 \cdot \text{Sqrt}[6] \cdot (Dx - Dy) / \text{den2}\{Dx, Dy, Dz\} \}, \{ 0.25 \cdot \text{Sqrt}[6] \cdot (Dx - Dy) / \text{den3}\{Dx, Dy, Dz\}, 0, \text{num3}\{Dx, Dy, Dz\} / \text{den3}\{Dx, Dy, Dz\}, 0, 0.25 \cdot \text{Sqrt}[6] \cdot (Dx - Dy) / \text{den3}\{Dx, Dy, Dz\} \}, \{ 0, 2^{-0.5}, 0, 2^{-0.5}, 0 \}, \{ 0, -(2^{-0.5}), 0, 2^{-0.5}, 0 \} \};$

$\text{g}\{Dx_ , Dy_ , Dz_ \} = \{ 4 \cdot Dz + Dx + Dy, 2 \cdot (Dx + Dy + Dz) + 2 \cdot \text{Sqrt}[\text{sig}\{Dx, Dy, Dz\}], 2 \cdot (Dx + Dy + Dz) - 2 \cdot \text{Sqrt}[\text{sig}\{Dx, Dy, Dz\}], Dz + 4 \cdot Dx + Dy, Dz + Dx + 4 \cdot Dy \};$

//Reading in the U38 C6-H6 bond orientations for all 500 energy-minimized structures from a file; the same file input is repeated for all five sites, but they are not shown here for brevity

U38bond = OpenRead["C:\\Documents and Settings\\Drobny\\Desktop\\LSM Transfer files\\U38C6H6.txt"];

U38all = {};

num = 500;

For[i = 0, i < num, i++;

If[Mod[i,100]==0, Print[i]];

U38C6H6 = Read[U38bond];

AppendTo[U38all, U38C6H6];

]

Close[U38bond];

pequil = {};

flaAll = {};

gAll = {};

DiffAll = {};

U38angAll = {};

U25angAll = {};

```

U23angAll = {};
U40angAll = {};
U42angAll = {};

```

*//Labels (1 through 500, as defined by FARFAR) of the 40 structures chosen for the simulation:
there are 7 β angle bins, 3 α bins and either one or two γ bins*

```

structures = { {{339},{200, 5},{46,73}},{{467, 278},{316,338},{7,198}},{{75,
136},{329,48},{183,
178}},{{151,343},{201,352},{284,130}},{{166,458},{469,202},{396}},{{341,386},{344,335}
,{438,172}},{{156,281},{499,354},{168,347}}} };

```

```

dimstr = Dimensions[structures]

```

```

numstr = 0;

```

//Automated counting of the number of structures in the chosen set

```

For[i = 0, i < dimstr[[1]], i++;
  For[j = 0, j < dimstr[[2]], j++;
    numstr += Length[structures[[i,j]]];
  ];
]
numstr

```

//The following are a set of structures that connect each structure in the set of 40 with all nearest neighbour structures: so, for example, fnum β [[2]] is the structure that is connected with structure 2 in the list by a β jump in the forward direction along the list, while bnum α [[35]] is the structure connected to structure 35 in the backward direction along the list. These vectors are set up in order to help build the jump matrix between states for a given structure list.

```

fnum $\beta$  = Table[0, {k, numstr}];
bnum $\beta$  = Table[0, {k, numstr}];
fnum $\gamma$  = Table[0, {k, numstr}];
bnum $\gamma$  = Table[0, {k, numstr}];
fnum $\alpha$  = Table[0, {k, numstr}];
bnum $\alpha$  = Table[0, {k, numstr}];
count = 0;
For[i = 0, i < dimstr[[1]], i++;
  For[j = 0, j < dimstr[[2]], j++;
    For[k = 0, k < Length[structures[[i,j]]], k++;
      count++;
      If[i+1 ≤ dimstr[[1]],
        next = structures[[i+1,j]];
        If[Length[next] == Length[structures[[i,j]]], fnum $\beta$ [[count]] = next[[k]],
          If[Length[next] > Length[structures[[i,j]]], fnum $\beta$ [[count]] = next[[1]],
            If[k > Length[next], fnum $\beta$ [[count]] = 0, fnum $\beta$ [[count]] = next[[k]] ]
        ];
      ];
    ];
  ];
  If[i-1 ≥ 1,

```

```

prev = structures[[i-1,j]];
If[Length[prev] == Length[structures[[i,j]]],bnum $\beta$ [[count]] = prev[[k]],
If[Length[prev] > Length[structures[[i,j]]], bnum $\beta$ [[count]] = prev[[1]],
If[k>Length[prev],bnum $\beta$ [[count]] = 0,bnum $\beta$ [[count]] = prev[[k]] ]
];
];
If[j+1 ≤ dimstr[[2]],
next = structures[[i,j+1]];
If[Length[next] == Length[structures[[i,j]]],fnum $\alpha$ [[count]] = next[[k]],
If[Length[next] > Length[structures[[i,j]]], fnum $\alpha$ [[count]] = next[[1]],
If[k>Length[next],fnum $\alpha$ [[count]] = 0,fnum $\alpha$ [[count]] = next[[k]] ]
];
];
If[j-1 ≥ 1,
prev = structures[[i,j-1]];
If[Length[prev] == Length[structures[[i,j]]],bnum $\alpha$ [[count]] = prev[[k]],
If[Length[prev] > Length[structures[[i,j]]], bnum $\alpha$ [[count]] = prev[[1]],
If[k>Length[prev],bnum $\alpha$ [[count]] = 0,bnum $\alpha$ [[count]] = prev[[k]] ]
];
];
If[k+1 ≤ Length[structures[[i,j]]],
fnum $\gamma$ [[count]] = structures[[i,j,k+1]];
];
If[k-1 ≥ 1,
bnum $\gamma$ [[count]] = structures[[i,j,k-1]];
];
];
];
]
fnum $\beta$ 
bnum $\beta$ 
fnum $\alpha$ 
bnum $\alpha$ 
fnum $\gamma$ 
bnum $\gamma$ 
fnum $\gamma$ [[6]]

```

//Constructing the list of equilibrium probabilities; in this case, the probabilities are set to be uniform in the sense of: probability = (Number of β bins × Number of α bins in the given β bin × Number of γ bins in the given α bin)⁻¹
pequil = {};

```
For[i = 0, i < dimstr[[1]], i++,
```

```

For[j = 0, j < dimstr[[2]], j++;
  For[k = 0, k < Length[structures[[i,j]], k++;
    pr = p1U*(dimstr[[1]]*dimstr[[2]]*Length[structures[[i,j]])^-1;
    AppendTo[pequil, pr];
    pr = p2U*(dimstr[[1]]*dimstr[[2]]*Length[structures[[i,j]])^-1;
    AppendTo[pequil, pr];
  ];
];
]
Sum[pequil[[i]], {i, 2*numstr}];
//Final adjustments to the structure vectors described above, to allow them to help build the jump matrix
For[no=0, no<numstr, no++;
  count = 0;
  ff $\beta$  = 0;
  fb $\beta$  = 0;
  ff $\alpha$  = 0;
  fb $\alpha$  = 0;
  ff $\gamma$  = 0;
  fb $\gamma$  = 0 ;
  For[i = 0, i < dimstr[[1]], i++;
    For[j = 0, j < dimstr[[2]], j++;
      For[k = 0, k < Length[structures[[i,j]], k++;
        count++;

        If[(fnum $\beta$ [[no]]==structures[[i,j,k]])&&(ff $\beta$  == 0), fnum $\beta$ [[no]] = count; ff $\beta$ ++];
        If[(bnum $\beta$ [[no]]==structures[[i,j,k]])&&(fb $\beta$  == 0), bnum $\beta$ [[no]] = count; fb $\beta$ ++];
        If[(fnum $\alpha$ [[no]]==structures[[i,j,k]])&&(ff $\alpha$  == 0), fnum $\alpha$ [[no]] = count; ff $\alpha$ ++];
        If[(bnum $\alpha$ [[no]]==structures[[i,j,k]])&&(fb $\alpha$  == 0), bnum $\alpha$ [[no]] = count; fb $\alpha$ ++];
        If[(fnum $\gamma$ [[no]]==structures[[i,j,k]])&&(ff $\gamma$  == 0), fnum $\gamma$ [[no]] = count; ff $\gamma$ ++];
        If[(bnum $\gamma$ [[no]]==structures[[i,j,k]])&&(fb $\gamma$  == 0), bnum $\gamma$ [[no]] = count; fb $\gamma$ ++];
      ];
    ];
  ];
]
fnum $\beta$ 
bnum $\beta$ 
fnum $\alpha$ 
bnum $\alpha$ 
fnum $\gamma$ 
bnum $\gamma$ 
fnum $\gamma$ [[6]]
*****

```

//Example of the calculations carried out using the atomic coordinate PDB file for a single structure

(*Structure 339*)

sval = 339; *//Structure label*

fname = "C:\Documents and Settings\Drobny\Desktop\LSM Transfer files\FARFAR_";

fname = fname <> ToString[sval] <> "_A2_3.res";

inphnmr = OpenRead[fname];

Find[inphnmr, "Anisotropic rotational diffusion"];

section = Read[inphnmr];

difflist = {};

For[i = 0, i < 15, i++;

AppendTo[difflist, Read[inphnmr, Number]];

]

Close[inphnmr];

//Reading in the diffusion eigenvalues and eigenvectors and assigning them to their Cartesian counterparts

Diff = {difflist[[2]], difflist[[7]], difflist[[12]]};

D2vec = {-difflist[[8]], -difflist[[9]], -difflist[[10]]}

D1vec = {-difflist[[3]], -difflist[[4]], -difflist[[5]]}

D3vec = {difflist[[13]], difflist[[14]], difflist[[15]]}

D1vec /= Norm[D1vec];

D2vec /= Norm[D2vec];

D3vec /= Norm[D3vec];

Dxvec = D2vec;

Dyvec = D1vec;

Dzvec = D3vec;

Dzvec.Cross[Dxvec, Dyvec]

Diffinp = {Diff[[2]], Diff[[1]], Diff[[3]]};

//Creating lists of the eigenvalues and eigenvectors needed by the general rate theory for all the structures (repeated here to represent the two internal motion sites)

AppendTo[DiffAll, Diffinp];

AppendTo[flaAll, fla[Diffinp]];

AppendTo[gAll, g[Diffinp]];

AppendTo[DiffAll, Diffinp];

AppendTo[flaAll, fla[Diffinp]];

AppendTo[gAll, g[Diffinp]];

//Calculations for the U38 site

(* U38 *)

sitep=U38all[[sval]];

//Instead of importing pre-calculated internal motion orientations, the program calculates the orientations in situ, using the U38 C6-H6 bond from the PDB file as the intermediate orientation between the two jump-sites. As can be seen, for the helical residues, the internal motion of the base is treated as a “base-libration”, i.e. a rotation about the base normal

```
site1p = RotationMatrix[JumpU38 Degree, U38normal[[sval]]].sitep;
site2p = RotationMatrix[-JumpU38 Degree, U38normal[[sval]]].sitep;
```

```
site1p /= Norm[site1p];
site1 = {site1p.Dxvec, site1p.Dyvec, site1p.Dzvec};
site1 /= Norm[site1];
//Calculation of spherical angles for each orientation
```

```
(*Site 1*)
U38th = ArcCos[-site1[[3]] ];
U38ph= ArcTan[site1[[2]]/site1[[1]] ];
(180/Pi)*U38th
(180/Pi)*U38ph
AppendTo[U38angAll, {U38th, U38ph}];
```

```
site2p /= Norm[site2p];
site2 = {site2p.Dxvec, site2p.Dyvec, site2p.Dzvec};
site2 /= Norm[site2];
(*Site 1*)
U38th = ArcCos[-site2[[3]] ];
U38ph= ArcTan[site2[[2]]/site2[[1]] ];
(180/Pi)*U38th
(180/Pi)*U38ph
AppendTo[U38angAll, {U38th, U38ph}];
```

//Same as for U38, with the exception that the internal motion for U25 and U23 is a rotation of the base around the glycosidic bond

```
(*      U25      *)
siteU25p=U25all[[sval]];
siteU251p = RotationMatrix[JumpU25 Degree, U25glycosidic[[sval]]].siteU25p;
siteU252p = RotationMatrix[-JumpU25 Degree, U25glycosidic[[sval]]].siteU25p;
```

```
siteU251p /= Norm[siteU251p];
siteU251 = {siteU251p.Dxvec, siteU251p.Dyvec, siteU251p.Dzvec};
siteU251 /= Norm[siteU251];
(*Site 1*)
U25th = ArcCos[-siteU251[[3]] ];
U25ph= ArcTan[siteU251[[2]]/siteU251[[1]] ];
(180/Pi)*U25th
(180/Pi)*U25ph
AppendTo[U25angAll, {U25th, U25ph}];
siteU252p /= Norm[siteU252p];
siteU252 = {siteU252p.Dxvec, siteU252p.Dyvec, siteU252p.Dzvec};
```

```

siteU252 /= Norm[siteU252];
(*Site 1*)
U25th = ArcCos[-siteU252[[3]]];
U25ph= ArcTan[siteU252[[2]]/siteU252[[1]]];
(180/Pi)*U25th
(180/Pi)*U25ph
AppendTo[U25angAll, {U25th, U25ph}];

```

//The calculation is repeated for the three other sites, but these are not shown for the sake of brevity

//As in Appendix IV.A, the spherical angle-dependent multipliers of the time-dependent coefficients are defined as functions of the spherical angles

```

coeff1[{thetaf_, phif_},{thetai_, phii_}] =
0.75*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Sin[2*phif])*(Sin[2*phii]);
coeff4[{thetaf_, phif_},{thetai_, phii_}] =
0.75*(Sin[2*thetaf])*(Sin[2*thetai])*(Sin[phif])*(Sin[phii]);
coeff5[{thetaf_, phif_},{thetai_, phii_}] =
0.75*(Sin[2*thetaf])*(Sin[2*thetai])*(Cos[phif])*(Cos[phii]);
coeff2[{thetaf_, phif_},{thetai_, phii_,inb_, ina_] = 1.5*flaAll[[inb,2, 1]]*flaAll[[ina,2,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*flaAll[[inb,2,
1]]*flaAll[[ina,2, 3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) +
(0.5*Sqrt[1.5])*flaAll[[ina,2, 1]]*flaAll[[inb,2,
3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) - 1) + 0.25*flaAll[[inb,2, 3]]*flaAll[[ina,2,
3]]*(3*(Cos[thetaf]^2) - 1)*(3*(Cos[thetai]^2) - 1);
c32[{thetaf_, phif_},{thetai_, phii_,inb_, ina_] = 1.5*flaAll[[inb,3, 1]]*flaAll[[ina,2,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*flaAll[[inb,3,
1]]*flaAll[[ina,2, 3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) +
(0.5*Sqrt[1.5])*flaAll[[ina,2, 1]]*flaAll[[inb,3,
3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) - 1) + 0.25*flaAll[[inb,3, 3]]*flaAll[[ina,2,
3]]*(3*(Cos[thetaf]^2) - 1)*(3*(Cos[thetai]^2) - 1);
c23[{thetaf_, phif_},{thetai_, phii_,inb_, ina_] = 1.5*flaAll[[inb,2, 1]]*flaAll[[ina,3,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*flaAll[[inb,2,
1]]*flaAll[[ina,3, 3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) +
(0.5*Sqrt[1.5])*flaAll[[ina,3, 1]]*flaAll[[inb,2,
3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) - 1) + 0.25*flaAll[[inb,2, 3]]*flaAll[[ina,3,
3]]*(3*(Cos[thetaf]^2) - 1)*(3*(Cos[thetai]^2) - 1);
coeff3[{thetaf_, phif_},{thetai_, phii_, inb_, ina_] = 1.5*flaAll[[inb,3, 1]]*flaAll[[ina,3,
1]]*(Sin[thetaf]^2)*(Sin[thetai]^2)*(Cos[2*phif])*(Cos[2*phii]) + (0.5*Sqrt[1.5])*flaAll[[inb,3,
1]]*flaAll[[ina,3, 3]]*(Sin[thetaf]^2)*(Cos[2*phif])*(3*(Cos[thetai]^2) - 1) +
(0.5*Sqrt[1.5])*flaAll[[ina,3, 1]]*flaAll[[inb,3,
3]]*(Sin[thetai]^2)*(Cos[2*phii])*(3*(Cos[thetaf]^2) - 1) + 0.25*flaAll[[inb,3, 3]]*flaAll[[ina,3,
3]]*(3*(Cos[thetaf]^2) - 1)*(3*(Cos[thetai]^2) - 1);

```

//In the following, lists of these multipliers are formed for each site; only the case of U38 is shown here, with the understanding that the same calculations are repeated for the other sites as well

```
(*U38: *)
U38co1list = {};
U38c22list = {};
U38c23list = {};
U38c32list = {};
U38c33list = {};
U38co4list = {};
U38co5list = {};
For[i = 0, i < Length[DiffAll], i++;
  c1temp = {};
  c22temp = {};
  c32temp = {};
  c23temp = {};
  c33temp = {};
  c4temp = {};
  c5temp = {};
  For[j = 0, j < Length[DiffAll], j++;
    AppendTo[c1temp, coeff1[U38angAll[[i]], U38angAll[[j]] ] ];
    AppendTo[c22temp, coeff2[U38angAll[[i]], U38angAll[[j]], 1,1 ] ];
    AppendTo[c32temp, c32[U38angAll[[i]], U38angAll[[j]], 1,1 ] ];
    AppendTo[c23temp, c23[U38angAll[[i]], U38angAll[[j]], 1,1 ] ];
    AppendTo[c33temp, coeff3[U38angAll[[i]], U38angAll[[j]], 1,1 ] ];
    AppendTo[c4temp, coeff4[U38angAll[[i]], U38angAll[[j]] ] ];
    AppendTo[c5temp, coeff5[U38angAll[[i]], U38angAll[[j]] ] ];
  ];
  AppendTo[U38co1list,c1temp];
  AppendTo[U38c22list,c22temp];
  AppendTo[U38c32list,c32temp];
  AppendTo[U38c23list,c23temp];
  AppendTo[U38c33list,c33temp];
  AppendTo[U38co4list,c4temp];
  AppendTo[U38co5list,c5temp];
]
*****
//Defining matrices of the eigenvalues of the rotational diffusion process (for ease of calculation
of the relaxation times in the final algorithm)
rlam1 = Table[0, {k, Length[DiffAll]}, {l,Length[DiffAll]}];
For[i = 0, i < Length[DiffAll], i++;
  rlam1[[i,i]] = -gAll[[i,1]];
]
rlam4 = Table[0, {k, Length[DiffAll]}, {l,Length[DiffAll]}];
For[i = 0, i < Length[DiffAll], i++;
  rlam4[[i,i]] = -gAll[[i,4]];
]
```

```

]
rlam5 = Table[0, {k, Length[DiffAll]}, {1, Length[DiffAll]}];
For[i = 0, i < Length[DiffAll], i++;
  rlam5[[i,i]] = -gAll[[i,5]];
]
rlamcomb = Table[0, {k, 2*Length[DiffAll]}, {1, 2*Length[DiffAll]}];
For[i = 0, i < Length[DiffAll], i++;
  sum22 = Sum[flaAll[[1,2, k]]*flaAll[[i,2, k]], {k,1,5}];
  sum23 = Sum[flaAll[[1,2, k]]*flaAll[[i,3, k]], {k,1,5}];
  sum32 = Sum[flaAll[[1,3, k]]*flaAll[[i,2, k]], {k,1,5}];
  sum33 = Sum[flaAll[[1,3, k]]*flaAll[[i,3, k]], {k,1,5}];
  evalmodff = gAll[[i,2]]*(sum22^2) + gAll[[i,3]]*(sum23^2);
  evalmodfs = sum22*gAll[[i,2]]*sum32 + sum23*gAll[[i,3]]*sum33;
  evalmodsf = evalmodfs;
  evalmodss = gAll[[i,2]]*(sum32^2) + gAll[[i,3]]*(sum33^2);
  rlamcomb[[2*i-1, 2*i-1]] = -evalmodff;
  rlamcomb[[2*i-1, 2*i]] = -evalmodfs;
  rlamcomb[[2*i, 2*i-1]] = -evalmodsf;
  rlamcomb[[2*i, 2*i]] = -evalmodss;
];

```

//Defining the jump matrix for the entire 40×40 lattice of structures

```

ratemat[{rβ_, rα_, rγ_}, rbase_] := Module[{umat = Table[0, {k, Length[DiffAll]}, {1,
Length[DiffAll]}]
},
For[j = -1, j < Length[DiffAll]-3, j+=2;
  umat[[j, j+1]] = rbase*Sqrt[pequil[[j]]/pequil[[j+1]] ];
  umat[[j, j]] -= rbase*Sqrt[pequil[[j+1]]/pequil[[j]] ];
  umat[[j+1, j]] = rbase*Sqrt[pequil[[j+1]]/pequil[[j]] ];
  umat[[j+1, j+1]] -= rbase*Sqrt[pequil[[j]]/pequil[[j+1]] ];

  indβ = 2*fnumβ[(j+1)/2]-1;
  indα = 2*fnumα[(j+1)/2]-1;
  indγ = 2*fnumγ[(j+1)/2]-1;

  If[indβ ≠ 0,
    umat[[j, indβ]] = rβ*Sqrt[pequil[[j]]/pequil[[indβ]] ];
    umat[[j, j]] -= rβ*Sqrt[pequil[[indβ]]/pequil[[j]] ];
  ];
  If[indα ≠ 0,
    umat[[j, indα]] = rα*Sqrt[pequil[[j]]/pequil[[indα]] ];
    umat[[j, j]] -= rα*Sqrt[pequil[[indα]]/pequil[[j]] ];
  ];
  If[indγ ≠ 0,
    umat[[j, indγ]] = rγ*Sqrt[pequil[[j]]/pequil[[indγ]] ];
    umat[[j, j]] -= rγ*Sqrt[pequil[[indγ]]/pequil[[j]] ];
  ];

```

```

];
ind $\beta$  = 2*bn $\beta$ [(j+1)/2]-1;
ind $\alpha$  = 2*bn $\alpha$ [(j+1)/2]-1;
ind $\gamma$  = 2*bn $\gamma$ [(j+1)/2]-1;
If[ind $\beta$   $\neq$  0,
  umat[[j, ind $\beta$ ]] = r $\beta$ *Sqrt[pequil[[j]]/pequil[[ind $\beta$ ]] ];
  umat[[j, j]] -= r $\beta$ *Sqrt[pequil[[ind $\beta$ ]]/pequil[[j]] ];
];
If[ind $\alpha$   $\neq$  0,
  umat[[j, ind $\alpha$ ]] = r $\alpha$ *Sqrt[pequil[[j]]/pequil[[ind $\alpha$ ]] ];
  umat[[j, j]] -= r $\alpha$ *Sqrt[pequil[[ind $\alpha$ ]]/pequil[[j]] ];
];
If[ind $\gamma$   $\neq$  0,
  umat[[j, ind $\gamma$ ]] = r $\gamma$ *Sqrt[pequil[[j]]/pequil[[ind $\gamma$ ]] ];
  umat[[j, j]] -= r $\gamma$ *Sqrt[pequil[[ind $\gamma$ ]]/pequil[[j]] ];
];

];
For[j = 0, j < Length[DiffAll]-2, j+=2;
  ind $\beta$  = 2*f $\beta$ [(j/2)];
  ind $\alpha$  = 2*f $\alpha$ [(j/2)];
  ind $\gamma$  = 2*f $\gamma$ [(j/2)];

  If[ind $\beta$   $\neq$  0,
    umat[[j, ind $\beta$ ]] = r $\beta$ *Sqrt[pequil[[j]]/pequil[[ind $\beta$ ]] ];
    umat[[j, j]] -= r $\beta$ *Sqrt[pequil[[ind $\beta$ ]]/pequil[[j]] ];
  ];
  If[ind $\alpha$   $\neq$  0,
    umat[[j, ind $\alpha$ ]] = r $\alpha$ *Sqrt[pequil[[j]]/pequil[[ind $\alpha$ ]] ];
    umat[[j, j]] -= r $\alpha$ *Sqrt[pequil[[ind $\alpha$ ]]/pequil[[j]] ];
  ];
  If[ind $\gamma$   $\neq$  0,
    umat[[j, ind $\gamma$ ]] = r $\gamma$ *Sqrt[pequil[[j]]/pequil[[ind $\gamma$ ]] ];
    umat[[j, j]] -= r $\gamma$ *Sqrt[pequil[[ind $\gamma$ ]]/pequil[[j]] ];
  ];
  ind $\beta$  = 2*bn $\beta$ [(j/2)];
  ind $\alpha$  = 2*bn $\alpha$ [(j/2)];
  ind $\gamma$  = 2*bn $\gamma$ [(j/2)];
  If[ind $\beta$   $\neq$  0,
    umat[[j, ind $\beta$ ]] = r $\beta$ *Sqrt[pequil[[j]]/pequil[[ind $\beta$ ]] ];
    umat[[j, j]] -= r $\beta$ *Sqrt[pequil[[ind $\beta$ ]]/pequil[[j]] ];
  ];
  If[ind $\alpha$   $\neq$  0,
    umat[[j, ind $\alpha$ ]] = r $\alpha$ *Sqrt[pequil[[j]]/pequil[[ind $\alpha$ ]] ];
    umat[[j, j]] -= r $\alpha$ *Sqrt[pequil[[ind $\alpha$ ]]/pequil[[j]] ];
  ];
];

```

```

If[ind $\gamma$   $\neq$  0,
umat[[j, ind $\gamma$ ]] = r $\gamma$ *Sqrt[pequil[[j]]/pequil[[ind $\gamma$ ]] ];
umat[[j, j]] -= r $\gamma$ *Sqrt[pequil[[ind $\gamma$ ]]/pequil[[j]] ];
];

];
umat
]
*****

```

//MCMC parameter search algorithm: search is for the rates of the β , α , γ jumps (in logarithm space)

```

prevparams = Log[{1.38*10^6, 1.38*10^6, 1.38*10^6}]; //Initial parameter set
sd = {0.5, 0.5, 0.5}; //Standard deviations of parameter-search Gaussians
uprate = 24; //Upper search limit of log of the rates
lorate = 10; //Lower search limit of log of the rates
ntrials = 150; //Number of MCMC iterations
//Experimental data for U38: the data for U38 alone is used for goodness-of-fit comparisons
T1expt = 354;
T2expt = 24.6;
NOEexpt = 1.14;

```

//Tolerances for the comparison against the experimental data (i.e. the standard deviations of the Gaussian weighted comparisons of the parametric relaxation and NOE data to the experimental data)

```

sdT = 0.5;
sdNOE = 0.1;

```

//Simulated annealing parameter

```

simann = 0.9995;
count = 1;

```

//Size of the time step used in the Gauss-Laguerre integration process

```

delT = 1*10^-9;

```

//Reduced number of roots (relative to the total of 250 calculated above) used in the actual integration (can vary if needed)

```

nroots = 150;

```

//File to write the selected parameter sets

```

outputfile = "C:\\Documents and Settings\\Drobny\\Desktop\\LSM Transfer
files\\LSM_MCMCfit.txt";
output = OpenWrite[outputfile];

```

```

//Beginning of the MCMC iteration loop
For[trial = 0, trial < ntrials, trial++;
  Print[trial];
  flag = 1;
  If[trial == 1, currparams = prevparams,

//Search for the next parameter set, proceed only if within limits set above
  While[flag == 1,
    r = RandomReal[NormalDistribution[prevparams[[count]], sd[[count]] ] ];
    If[(r < lorate)|| (r > uprate), flag = 1, flag = 0; currparams[[count]] = r];
  ];
  count ++;
  If[count == 4, count = 1];
  locrate = {rU38, rU23, rU25, rU40, rU42};
  corrvec = {};

//Calculation of the rate matrix and correlation function for the given parameter set:

//Iteration over the three types of bases: U38; U23 and U25 which are taken to have the same
internal motional parameters; and, U40 and U42, which are also taken to have the same internal
motional parameters
  For[nbases = 0, nbases < 3, nbases++;
    //Rate matrix calculation
    rmatrix = ratemat[Exp[currparams], locrate[[nbases]]];

    rmatcomb = Table[0, {k, 2*Length[DiffAll]}, {l, 2*Length[DiffAll]}};
    Dimensions[rmatcomb];
    For[i = 0, i < Length[DiffAll], i++;
      For[j = 0, j < Length[DiffAll], j++;
        rmatcomb[[2*i-1, 2*j-1]] = rmatrix[[i, j]];
        rmatcomb[[2*i, 2*j]] = rmatrix[[i, j]];
      ];
    ];

//Spherical angle multipliers for the given type of base
  If[nbases == 1,
    co1list1 = U38co1list;
    c22list1 = U38c22list;
    c32list1 = U38c32list;
    c23list1 = U38c23list;
    c33list1 = U38c33list;
    co4list1 = U38co4list;
    co5list1 = U38co5list;
    co1list2 = {};

```

```

c22list2 = {};
c32list2 = {};
c23list2 = {};
c33list2 = {};
co4list2 = {};
co5list2 = {},
If[nbases==2,
  co1list1 = U23co1list;
  c22list1= U23c22list;
  c32list1 = U23c32list;
  c23list1 = U23c23list;
  c33list1 = U23c33list;
  co4list1 = U23co4list;
  co5list1 = U23co5list;

  co1list2 = U25co1list;
  c22list2 = U25c22list;
  c32list2 = U25c32list;
  c23list2 = U25c23list;
  c33list2 = U25c33list;
  co4list2 = U25co4list;
  co5list2 = U25co5list,
If[nbases==3,
  co1list1 = U40co1list;
  c22list1 = U40c22list;
  c32list1 = U40c32list;
  c23list1 = U40c23list;
  c33list1 = U40c33list;
  co4list1 = U40co4list;
  co5list1 = U40co5list;

  co1list2 = U42co1list;
  c22list2 = U42c22list;
  c32list2 = U42c32list;
  c23list2 = U42c23list;
  c33list2 = U42c33list;
  co4list2 = U42co4list;
  co5list2 = U42co5list;

]

];
//Calculation of the correlation function, which proceeds similar to the previously commented
code for the general rate theory
evolmat1 = rlam1 + rmatrix;

```



```

evolmatcomb = rlamcomb + rmatcomb;
evolmat4 = rlam4 + rmatrix;
evolmat5 = rlam5 + rmatrix;
Clear[rmatrix];
eval1 = -Eigenvalues[evolmat1];
evecs1 = Eigenvectors[evolmat1];
T1mat = Transpose[evecs1];
Clear[evecs1];

Inv = Inverse[T1mat];

c1[tin_] := Module[{c1mat = Table[0, {k, Length[DiffAll]}, {l, Length[DiffAll]}]},
  For[i = 0, i < Length[DiffAll], i++,
    For[j = 0, j < Length[DiffAll], j++,
      c1mat[[i, j]] = Sum[T1mat[[i, k]]*Exp[-eval1[[k]]*tin]*Inv[[k, j]], {k, Length[DiffAll]}];
    ];
  ];
c1mat
];
corr1[tin_] := Module[{c1mat = c1[tin]}, corrbase1 =
Sum[Sum[(c1mat[[j, k]]*co1list1[[j, k]])*pequil[[k]], {k, 1, Length[DiffAll]}], {j, 1,
Length[DiffAll]}];
  If[Length[co1list2] > 0,
    corrbase2 = Sum[Sum[(c1mat[[j, k]]*co1list2[[j, k]])*pequil[[k]], {k, 1, Length[DiffAll]}], {j,
1, Length[DiffAll]}], corrbase2 = 0];
  {corrbase1, corrbase2}
];
//Appending the values of the correlation function evaluated for each base at a particular time
step to a list
corrtemp1 = {};
corrtemp2 = {};
corrduo = {};
For[nr = 0, nr < nroots, nr++,
  AppendTo[corrduo, corr1[delT*t[[nr]]]];
];
For[nr = 0, nr < nroots, nr++,
  AppendTo[corrtemp1, corrduo[[nr, 1]]];
  AppendTo[corrtemp2, corrduo[[nr, 2]]];
];
Clear[c1, corr1, eval1];
Print["Eval 1"];
evalcomb = -Eigenvalues[evolmatcomb];
evecscomb = Eigenvectors[evolmatcomb];
Tcombmat = Transpose[evecscomb];
Clear[evecscomb];

```

```

Inv = Inverse[Tcombmata];

corrcomb[tin_] := Module[{ccombmat = Table[0, {k, 2*Length[DiffAll]}, {l,
2*Length[DiffAll]}]},
  For[i = 0, i < 2*Length[DiffAll], i++;
    For[j = 0, j < 2*Length[DiffAll], j++;
      ccombmat[[i, j]] = Sum[Tcombmata[[i, k]]*Exp[-evalcomb[[k]]*tin]*Inv[[k, j]], {k,
2*Length[DiffAll]}];
    ];
  ];
  corrbas1 = Sum[Sum[(ccombmat[[2*j-1, 2*k-1]]*c22list1[[j, k]] + ccombmat[[2*j, 2*k-
1]]*c32list1[[j, k]] + ccombmat[[2*j-1, 2*k]]*c23list1[[j, k]]
+ ccombmat[[2*j, 2*k]]*c33list1[[j, k]])*pequil[[k]], {k, 1, Length[DiffAll]}], {j, 1,
Length[DiffAll]}];

  If[Length[c22list2] > 0,
    corrbas2 = Sum[Sum[(ccombmat[[2*j-1, 2*k-1]]*c22list2[[j, k]] + ccombmat[[2*j, 2*k-
1]]*c32list2[[j, k]] + ccombmat[[2*j-1, 2*k]]*c23list2[[j, k]]
+ ccombmat[[2*j, 2*k]]*c33list2[[j, k]])*pequil[[k]], {k, 1, Length[DiffAll]}], {j, 1,
Length[DiffAll]}], corrbas2 = 0];
  {corrbas1, corrbas2}
];
corrduo = {};
For[nr = 0, nr < nroots, nr++;
  AppendTo[corrduo, corrcomb[delT*t[[nr]]] ];
];
For[nr = 0, nr < nroots, nr++;
  corrtmp1[[nr]] += corrduo[[nr, 1]] ;
  corrtmp2[[nr]] += corrduo[[nr, 2]] ;
];

Clear[ccomb, corrcomb, evalcomb];
Print["Evals 2 & 3"];
eval4 = -Eigenvalues[evolmat4];
evecs4 = Eigenvectors[evolmat4];
T4mat = Transpose[evecs4];
Clear[evecs4];

Inv = Inverse[T4mat];

c4[tin_] := Module[{c4mat = Table[0, {k, Length[DiffAll]}, {l, Length[DiffAll]}]},
  For[i = 0, i < Length[DiffAll], i++;
    For[j = 0, j < Length[DiffAll], j++;
      c4mat[[i, j]] = Sum[T4mat[[i, k]]*Exp[-eval4[[k]]*tin]*Inv[[k, j]], {k, Length[DiffAll]}];
    ];
  ];
];

```

```

c4mat
];
corr4[tin_]:=Module[{c4mat = c4[tin]},corrbase1 =
Sum[Sum[(c4mat[[j,k]]*co4list1[[j,k]])*pequil[[k]],{k, 1, Length[DiffAll]}], {j, 1,
Length[DiffAll]}];
If[Length[co4list2] >0,
corrbase2 = Sum[Sum[(c4mat[[j,k]]*co4list2[[j,k]])*pequil[[k]],{k, 1, Length[DiffAll]}], {j,
1, Length[DiffAll]}],corrbase2 = 0];
{corrbase1, corrbase2}
];

corrduo = {};
For[nr = 0, nr < nroots, nr++;
AppendTo[corrduo, corr4[delT*t[[nr]]] ];
];

For[nr = 0, nr < nroots, nr++;
corrtemp1[[nr]] += corrduo[[nr,1]] ;
corrtemp2[[nr]] += corrduo[[nr,2]] ;
];
Clear[c4,corr4, eval4];
Print["Eval 4"];
eval5 = -Eigenvalues[evolmat5];
evecs5= Eigenvectors[evolmat5];
T5mat = Transpose[evecs5];
Clear[evecs5];

Inv = Inverse[T5mat];

c5[tin_] := Module[{c5mat = Table[0, {k,Length[DiffAll]},{l, Length[DiffAll]}]},
For[i = 0, i < Length[DiffAll], i++;
For[j=0, j<Length[DiffAll],j++;
c5mat[[i,j]] = Sum[T5mat[[i,k]]*Exp[-eval5[[k]]*tin]*Inv[[k,j]],{k, Length[DiffAll]}];
];
];
c5mat
];
corr5[tin_]:=Module[{c5mat = c5[tin]},corrbase1 =
Sum[Sum[(c5mat[[j,k]]*co5list1[[j,k]])*pequil[[k]],{k, 1, Length[DiffAll]}], {j, 1,
Length[DiffAll]}];
If[Length[co5list2] >0,
corrbase2 = Sum[Sum[(c5mat[[j,k]]*co5list2[[j,k]])*pequil[[k]],{k, 1, Length[DiffAll]}], {j,
1, Length[DiffAll]}],corrbase2 = 0];
{corrbase1, corrbase2}
];

```

```

corrduo = {};
For[nr = 0, nr < nroots, nr++;
  AppendTo[corrduo, corr5[delT*t[[nr]]] ];
];

For[nr = 0, nr < nroots, nr++;
  corrtemp1[[nr]] += corrduo[[nr,1]] ;
  corrtemp2[[nr]] += corrduo[[nr,2]] ;
];
Clear[c5,corr5, eval5];
Print["Eval 5"];
Print[nbases];

AppendTo[corrvec,corrtemp1];
If[nbases ≠ 1, AppendTo[corrvec,corrtemp2]];
Clear[evolmat1, evolmatcomb, evolmat4, evolmat5,rmatrix, rmatcomb,T1mat, Tcombm,
T4mat, T5mat, eval1, evalcomb, eval4, eval5, c1,
ccomb,c4,c5,lam1,lamcomb,lam4,lam5,corr1,corrcomb, corr4,corr5, corrtemp];
];
//End of loop over type of base

//Calculation of the relevant spectral densities and relaxation times for each of the sites (only
the U38 and U23 cases are shown here for brevity
Allparams = {};
T1temp = {};
T2temp = {};
NOEtemp = {};
(* U38 *)
Print["U38 relaxation times:"];
J0m = delT*Sum[newmod[[i]]*corrvec[[1,i]]*Cos[mval*delT*t[[i]]], {i,1,nroots}];
J0C = delT*Sum[newmod[[i]]*corrvec[[1,i]]*Cos[ωC*delT*t[[i]]], {i,1,nroots}];
J0p = delT*Sum[newmod[[i]]*corrvec[[1,i]]*Cos[pval*delT*t[[i]]], {i,1,nroots}];
J00 = delT*Sum[newmod[[i]]*corrvec[[1,i]], {i,1,nroots}];
J0H = delT*Sum[newmod[[i]]*corrvec[[1,i]]*Cos[ωH*delT*t[[i]]], {i,1,nroots}];

T1 = 1000*(d2*(J0m + 3*J0C + 6*J0p) + co2*J0C )^-1//Chop;
T2= 1000*(0.5*d2*(4*J00 + J0m + 3*J0C + 6*J0H + 6*J0p) + (1/6)*co2*(4*J00 + 3*J0C)
)^-1//Chop;
NOE = 1 + (γH/γC)*d2*(6*J0p - J0m)*T1/1000//Chop;

Print[currparams];
Print[T1];
Print[T2];

```

```

Print[NOE];

AppendTo[T1temp, T1];
AppendTo[T2temp, T2];
AppendTo[NOEtemp, NOE];

(* U23 *)
Print["U23 relaxation times:"];
J0m = delT*Sum[newmod[[i]]*corrvec[[2,i]]*Cos[mval*delT*t[[i]]], {i,1,nroots}];
J0C = delT*Sum[newmod[[i]]*corrvec[[2,i]]*Cos[ωC*delT*t[[i]]], {i,1,nroots}];
J0p = delT*Sum[newmod[[i]]*corrvec[[2,i]]*Cos[pval*delT*t[[i]]], {i,1,nroots}];
J00 = delT*Sum[newmod[[i]]*corrvec[[2,i]], {i,1,nroots}];
J0H = delT*Sum[newmod[[i]]*corrvec[[2,i]]*Cos[ωH*delT*t[[i]]], {i,1,nroots}];

T1 = 1000*(d2*(J0m + 3*J0C + 6*J0p) + co2*J0C )^-1//Chop;
T2 = 1000*(0.5*d2*(4*J00 + J0m + 3*J0C + 6*J0H + 6*J0p) + (1/6)*co2*(4*J00 + 3*J0C)
)^-1//Chop;
NOE = 1 + (γH/γC)*d2*(6*J0p - J0m)*T1/1000//Chop;
AppendTo[T1temp, T1];
AppendTo[T2temp, T2];
AppendTo[NOEtemp, NOE];
Print[currparams];
Print[T1];
Print[T2];
Print[NOE];

Clear[corrvec];
//MCMC comparison algorithm, for details see comments made to similar algorithm in Appendix
V.A
If[trial == 1,
  lsdprev = -(((T1temp[[1]] - T1expt)^2)/(2*sdT*sdT) + ((T2temp[[1]] - T2expt)^2
)/(2*sdT*sdT)),
  lsdcurr = -(((T1temp[[1]] - T1expt)^2)/(2*sdT*sdT) + ((T2temp[[1]] - T2expt)^2
)/(2*sdT*sdT));
  lratio = lsdcurr - lsdprev;
  If[lratio ≥ 0,
    lsdprev = lsdcurr;
    prevparams = currparams;
    sdT *= simann;
    sdNOE *= simann;
    AppendTo[Allparams, {lsdcurr, T1temp, T2temp, NOEtemp, Exp[prevparams]}];
    Write[output, Allparams],
    rand2 = RandomReal[UniformDistribution[]];
    If[lratio ≥ Log[rand2],

```

```

    lsdprev = lsdcurr;
    prevparams = currparams;
    sdT *= simann;
    sdNOE *= simann;
    AppendTo[Allparams, {lsdcurr, T1temp, T2temp, NOEtemp, Exp[prevparams]}];
    Write[output, Allparams]
  ];
];

]

(*end of ntrials loop*);
Close[outputfile];
*****

```

Bibliography

- (1) Eskandari, S.; Prychyna, O.; Leung, J.; Avdic, D.; O'Neill, M. A. *J Am Chem Soc* **2007**, *129*, 11308-11309.
- (2) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A. N.; Woodside, M. T.; Block, S. M. *Science* **2008**, *319*, 630-633.
- (3) Lee, M. K.; Gal, M.; Frydman, L.; Varani, G. *Proc Natl Acad Sci U S A*, *107*, 9192-9197.
- (4) Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R. *Science* **2004**, *306*, 275-279.
- (5) Stoddard, C. D.; Montange, R. K.; Hennelly, S. P.; Rambo, R. P.; Sanbonmatsu, K. Y.; Batey, R. T. *Structure* **2010**, *18*, 787-797.
- (6) Aboul-ela, F.; Karn, J.; Varani, G. *J. Mol. Biol.* **1995**, *253*, 313-332.
- (7) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, *257*, 76-80.
- (8) Aboul-ela, F.; Karn, J.; Varani, G. *Nucleic Acids Res.* **1996**, *24*, 3974-3981.
- (9) Long, K. S.; Crothers, D. M. *Biochemistry* **1999**, *38*, 10059-10069.
- (10) Olsen, G. L.; Bardaro, M. F. J.; Echodu, D. C.; Drobny, G. P.; Varani, G. *J. Am. Chem. Soc.* **2010**, *132*, 303-308.
- (11) Bardaro, M. F. J.; Shajani, Z.; Patora-Komisarska, K.; Robinson, J. A.; Varani, G. *Nucleic Acids Res.* **2009**, *37*, 1529-1540.
- (12) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. *Science* **2006**, *311*, 653-656.
- (13) Hansen, A. L.; Al-Hashimi, H. M. *J. Am. Chem. Soc.* **2007**, *129*, 16072-16082.
- (14) Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M. *Nature* **2007**, *450*, 1263-1267.
- (15) Leulliot, N.; Varani, G. *Biochemistry* **2001**, *40*, 7947-7956.
- (16) Davidson, A.; Leeper, T. C.; Athanassiou, Z.; Patora-Komisarska, K.; Karn, J.; Robinson, J. A.; Varani, G. *Proc Natl Acad Sci U S A* **2009**, *106*, 11931-11936.
- (17) Kumar, S.; Kellish, P.; Robinson, W. E.; Wang, D. Y.; Appella, D. H.; Arya, D. P. *Biochemistry* **2012**, *51*, 2331-2347.
- (18) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. D. *Cell* **1994**, *76*, 357-369.
- (19) Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. *Cell* **1995**, *82*, 143-153.
- (20) Patra, S. K. *Exp Cell Res* **2008**, *314*, 1193-1201.
- (21) Meints, G. A.; Miller, P. A.; Pederson, K.; Shajani, Z.; Drobny, G. *J Am Chem Soc* **2008**, *130*, 7305-7314.
- (22) Pederson, K.; Meints, G. A.; Shajani, Z.; Miller, P. A.; Drobny, G. P. *J. Am. Chem. Soc.* **2008**, *130*, 9072-9079.
- (23) Wallach, D. *J. Chem. Phys.* **1967**, *47*, 5258-&.
- (24) Freed, J. H. *J. Chem. Phys.* **1977**, *66*, 4183-4199.
- (25) Wittebort, R. J.; Szabo, A. *J. Chem. Phys.* **1978**, *69*, 1722-1736.
- (26) Edholm, O.; Blomberg, C. *Chemical Physics* **1979**, *42*, 449-464.
- (27) Tarroni, R.; Zannoni, C. *J. Chem. Phys.* **1991**, *95*, 4550-4564.
- (28) Meirovitch, E.; Shapiro, Y. E.; Polimeno, A.; Freed, J. H. *Journal of Physical Chemistry A* **2006**, *110*, 8366-8396.

- (29) Echodu, D.; Goobes, G.; Shajani, Z.; Pederson, K.; Meints, G.; Varani, G.; Drobny, G. *J. Phys. Chem. B* **2008**, *112*, 13934-13944.
- (30) Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1982**, *79*, 386-389.
- (31) Wittebort, R. J.; Olejniczak, E. T.; Griffin, R. G. *J. Chem. Phys.* **1987**, *86*, 5411-5420.
- (32) Nadler, W.; Schulten, K. *J. Chem. Phys.* **1985**, *82*, 151-160.
- (33) Korzhnev, D. M.; Orekhov, V. Y.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127*, 713-721.
- (34) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Biomol. NMR* **2008**, *41*, 113-120.
- (35) Alam, T. M.; Drobny, G. P. *Chemical Reviews* **1991**, *91*, 1545-1590.
- (36) Breen, N. F.; Weidner, T.; Li, K.; Castner, D. G.; Drobny, G. P. *J. Am. Chem. Soc.* **2009**, *131*, 14148-14149.
- (37) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.
- (38) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559-4570.
- (39) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989-4991.
- (40) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269-5278.
- (41) King, G. C.; Harper, J. W.; Xi, Z. *J. Nuclear Magnetic Resonance and Nucleic Acids* **1995**, *261*, 436-450.
- (42) <http://ambermd.org>.
- (43) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668-1688.
- (44) Ponder, J. W.; Case, D. A. *Protein Simulations* **2003**, *66*, 27-85.
- (45) Cheatham, T. E.; Young, M. A. *Biopolymers* **2001**, *56*, 232-256.
- (46) www.charmm.org.
- (47) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187-217.
- (48) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128-132.
- (49) Horton, J. R.; Ratner, G.; Banavali, N. K.; Huang, N.; Choi, Y.; Maier, M. A.; Marquez, V. E.; MacKerell, A. D., Jr.; Cheng, X. *Nucleic Acids Res.* **2004**, *32*, 3877-3886.
- (50) Riccardi, L.; Nguyen, P. H.; Stock, G. *J. Phys. Chem. B* **2009**, *113*, 16660-16668.
- (51) www.rosettacommons.org.
- (52) Bradley, P.; Baker, D. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 922-929.
- (53) Das, R.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14664-14669.
- (54) Das, R.; Karanicolas, J.; Baker, D. *Nat Methods* **2010**, *7*, 291-294.
- (55) Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G. H.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* **2010**, *327*, 1014-1018.
- (56) Rose, M. E. *Elementary Theory of Angular Momentum*; Wiley: New York, 1957.

- (57) Tinkham, M. *Group Theory and Quantum Mechanics*; McGraw-Hill: New York, 1964.
- (58) Shajani, Z.; Varani, G. *Biopolymers* **2007**, *86*, 348-359.
- (59) Duer, M. J. *Introduction to Solid-State NMR Spectroscopy*; Blackwell Science Ltd.: Oxford, 2004.
- (60) Slichter, C. P. *Principles of Magnetic Resonance*; Springer-Verlag: Berlin, 1990.
- (61) Vold, R. R.; Vold, R. L. *Adv. Mag. Opt. Res.* **1991**, *16*, 85-171.
- (62) Ying, J. F.; Grishaev, A.; Bryce, D. L.; Bax, A. *J. Am. Chem. Soc.* **2006**, *128*, 11443-11454.
- (63) Carr, H. Y.; Purcell, E. M. *Phys. Rev.* **1954**, *94*, 630-638.
- (64) Meiboom, S.; Gill, D. *Review of Scientific Instruments* **1958**, *29*, 688-691.
- (65) Shajani, Z.; Varani, G. *J. Mol. Biol.* **2005**, *349*, 699-715.
- (66) Peng, J. W.; Thanabal, V.; Wagner, G. *J. Magn. Reson.* **1991**, *94*, 82-100.
- (67) Dethoff, E. A.; Hansen, A. L.; Musselman, C.; Watt, E. D.; Andricioaei, I.; Al-Hashimi, H. M. *Biophys. J.* **2008**, *95*, 3906-3915.
- (68) Miller, P. A.; Shajani, Z.; Meints, G. A.; Caplow, D.; Goobes, G.; Varani, G.; Drobny, G. P. *J Am Chem Soc* **2006**, *128*, 15970-15971.
- (69) Koplín, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. *Structure* **2005**, *13*, 1255-1267.
- (70) Al-Hashimi, H. M. *Biopolymers* **2007**, *86*, 345-347.
- (71) Hennig, M.; Williamson, J. R. *Nucleic Acids Res.* **2000**, *28*, 1585-1593.
- (72) Brodsky, A. S.; Williamson, J. R. *J. Mol. Biol.* **1997**, *267*, 624-639.
- (73) Puglisi, J. D.; Chen, L.; Frankel, A. D.; Williamson, J. R. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3680-3684.
- (74) Long, K. S.; Crothers, D. M. *Biochemistry* **1995**, *34*, 8885-8895.
- (75) Emani, P. S.; Olsen, G. L.; Echodu, D. C.; Varani, G.; Drobny, G. P. *J. Phys. Chem. B* **2010**, *114*, 15991-16002.
- (76) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471-1475.
- (77) Zhang, Q.; Al-Hashimi, H. M. *RNA* **2009**, *15*, 1941-1948.
- (78) Tugarinov, V.; Liang, Z.; Shapiro, Y. E.; Freed, J. H.; Meirovitch, E. *J. Am. Chem. Soc.* **2001**, *123*, 3055-3063.
- (79) Olsen, G. L.; Echodu, D. C.; Shajani, Z.; Bardaro, M. F. J.; Varani, G.; Drobny, G. P. *J. Am. Chem. Soc.* **2008**, *130*, 2896-2897.
- (80) Nuutero, S.; Fujimoto, B. S.; Flynn, P. F.; Reid, B. R.; Ribeiro, N. S.; Schurr, J. M. *Biopolymers* **1994**, *34*, 463-480.
- (81) Zhang, Q.; Al-Hashimi, H. M. *Nat. Meth.* **2008**, *5*, 243-245.
- (82) Wong, V.; Case, D. A.; Szabo, A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11016-11021.
- (83) Favro, L. D. *Phys. Rev.* **1960**, *119*, 53-62.
- (84) Woessner, D. E. *J. Chem. Phys.* **1962**, *37*, 647-654.
- (85) Olsen, G. L.; Bardaro, M. F. J.; Echodu, D. C.; Drobny, G. P. *J. Biomol. NMR* **2009**, *45*, 133-142.
- (86) Abragam, A. *Principles of Nuclear Magnetism*; Clarendon Press: Oxford, 1961.

- (87) Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, W. C. *J. Chem. Phys.* **1973**, 59, 915-924.
- (88) Duchardt, E.; Schwalbe, H. *J. Biomol. NMR* **2005**, 32, 295-308.
- (89) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. *J. Chem. Soc., Perkin Trans. 2* **1987**, S1-S19.
- (90) Carrasco, B.; de la Torre, J. G. *Biophys. J.* **1999**, 76, 3044-3057.
- (91) Yamakawa, H. *J. Chem. Phys.* **1970**, 53, 436-443.
- (92) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson.* **2000**, 147, 138-146.
- (93) Millero, F. J.; Dexter, R.; Hoff, E. *J. Chem. Eng. Data* **1971**, 16, 85-87.
- (94) Fernandes, M. X.; Ortega, A.; Martinez, M. C. L.; de la Torre, J. G. *Nucleic Acids Res.* **2002**, 30, 1782-1788.
- (95) Aragon, S. R.; Perez, M.; Ng, K.; Eden, D. Optical Kerr Effect of DNA Oligomers and RNA. In *Molecular and Colloidal Electro-Optics*; Stoylov, S. P., Stoimenova, M. V., Eds.; Taylor & Francis: Boca Raton, 2006.
- (96) Aragon, S. R.; Flamik, D. *Macromolecules* **2009**, 42, 6290-6299.
- (97) Shajani, Z.; Deka, P.; Varani, G. *Trends Biochem. Sci.* **2006**, 31, 421-424.
- (98) Razga, F.; Zacharias, M.; Reblova, K.; Koca, J.; Sponer, J. *Structure* **2006**, 14, 825-835.
- (99) Falb, M.; Amata, I.; Gabel, F.; Simon, B.; Carlomagno, T. *Nucleic Acids Res.* **2010**, 38, 6274-6285.
- (100) Shajani, Z.; Varani, G. *Biochemistry* **2008**, 47, 7617-7625.
- (101) Bothe, J. R.; Lowenhaupt, K.; Al-Hashimi, H. M. *J. Am. Chem. Soc.* **2011**, 133, 2016-2018.
- (102) Perrin, F. *Le Journal de Physique* **1936**, 7, 1-11.
- (103) Huntress, W. T. *J. Chem. Phys.* **1968**, 48, 3524-3533.
- (104) Freed, J. H. *J. Chem. Phys.* **1964**, 41, 2077-2083.
- (105) Jackson, J. D. *Classical Electrodynamics*, 3rd ed.; John Wiley & Sons, Inc., 1998.
- (106) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, 31, 5108-5121.
- (107) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, 25, 1605-1612.
- (108) Blackledge, M. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005**, 46, 23-61.
- (109) Stelzer, A. C.; Frank, A. T.; Bailor, M. H.; Andricioaei, I.; Al-Hashimi, H. M. *Methods* **2009**, 49, 167-173.
- (110) Ryabov, Y.; Clore, G. M.; Schwieters, C. D. *J Chem Phys* **2012**, 136, 034108.
- (111) Schurr, J. M.; Fujimoto, B. S.; Nuutero, S. *J. Magn. Reson., Ser A* **1994**, 106, 1-22.
- (112) Schurr, J. M.; Fujimoto, B. S.; Diaz, R.; Robinson, B. H. *J. Magn. Reson.* **1999**, 140, 404-431.
- (113) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. *Nature* **2011**, 470, 498-U484.
- (114) Cedar, H. *Cell* **1988**, 53, 3-4.
- (115) Meints, G. A.; Drobny, G. P. *Biochemistry* **2001**, 40, 12436-12443.
- (116) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *J. Am. Chem. Soc.* **1947**, 69, 2483-2488.

- (117) Meints, G. A.; Karlsson, T.; Drobny, G. P. *J. Am. Chem. Soc.* **2001**, *123*, 10030-10038.
- (118) Greenfield, M. S.; Ronemus, A. D.; Vold, R. L.; Vold, R. R.; Ellis, P. D.; Raidy, T. E. *J. Magn. Reson.* **1987**, *72*, 89-107.
- (119) Nadler, W.; Schulten, K. *J. Chem. Phys.* **1986**, *84*, 4015-4025.
- (120) Mathematica; Version 8.0 ed.; Wolfram Research Inc.: Champaign, IL, 2010.
- (121) Korzhnev, D. M.; Kay, L. E. *Accounts of Chemical Research* **2008**, *41*, 442-451.
- (122) Havlin, R. H.; Tycko, R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3284-3289.
- (123) Hu, K. N.; Yau, W. M.; Tycko, R. *J. Am. Chem. Soc.* **2010**, *132*, 24-+.
- (124) Emani, P. S.; Olsen, G. L.; Varani, G.; Drobny, G. P. *J Phys Chem A* **2011**, *115*, 12055-12069.
- (125) Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. *Science* **1996**, *273*, 1547-1551.
- (126) Zemmell, R. W.; Kelley, A. C.; Karn, J.; Butler, P. J. *J Mol Biol* **1996**, *258*, 763-777.
- (127) Olsen, G. L.; Edwards, T. E.; Deka, P.; Varani, G.; Sigurdsson, S. T.; Drobny, G. P. *Nucleic Acids Res.* **2005**, *33*, 3447-3454.
- (128) Huang, W.; Varani, G.; Drobny, G. P. *J Am Chem Soc* **2010**, *132*, 17643-17645.
- (129) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (130) Murray, L. J.; Arendall, W. B., 3rd; Richardson, D. C.; Richardson, J. S. *PNAS* **2003**, *100*, 13904-13909.
- (131) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732-738.
- (132) Ottiger, M.; Bax, A. *JACS* **1998**, *120*, 12334-12341.
- (133) Hansen, M. R.; Mueller, L.; Pardi, A. *Nature Structural Biology* **1998**, *5*, 1065-1074.
- (134) Tycko, R.; Blanco, F. J.; Ishii, Y. *JACS* **2000**, *122*, 9340-9341.
- (135) Ruckert, M.; Otting, G. *JACS* **2000**, *122*, 7793-7797.
- (136) Zweckstetter, M.; Bax, A. *JACS* **2000**, *122*, 3791-3792.

Curriculum Vitae

Prashant S. Emani

Educational Experience

University of Washington, Seattle: Doctorate of Philosophy (PhD.) in Physics (September 2007 – June 2012)

University of Washington, Seattle: Masters of Science (M.S.) in Physics (September 2006-September 2007)

University of Minnesota, Twin Cities: Bachelor of Science (B.S.) in Physics and Astrophysics with Honors (Summa cum Laude) and High Distinction (January 2002 – May 2006)

Research experience

University of Washington – Seattle, Department of Physics: PhD Thesis research (March 2008 – present)

Development of analytical and numerical methods to describe local and global dynamics in RNA and DNA using Nuclear Magnetic Resonance (NMR) experimental data

Advisor: Dr. Gary P. Drobny, Professor of Chemistry and Adjunct Professor in Department of Physics

University of Washington – Seattle, Department of Physics: Research rotation (April 2007 – December 2007)

Experimental test of Newton's inverse squared-distance law of gravitation on sub-millimeter scale

Advisor: Dr. Jens H. Gundlach, Professor of Physics

University of Minnesota – Twin Cities, Department of Physics and Astronomy: Undergraduate Senior thesis research (September 2005 – May 2006)

Statistical study of mass distributions of gravitational lenses using PixeLens (mass-distribution estimating software)

Advisor: Dr. Liliya L.R. Williams, Professor of Astronomy

Cornell University, Department of Astronomy Summer Research Experience for Undergraduates (REU) program:

Summer research project (June 2005 – August 2005)

Study of statistical methods to improve searches through observational data for extra-solar planets

Advisor: Dr. David Chernoff, Professor of Astronomy

Positions held

Research Assistant, Graduate Research with Dr. Gary P. Drobny, University of Washington – Seattle, Department of Chemistry (June 2008 – present, except when teaching in Physics department as seen below)

Teaching Assistant, University of Washington – Seattle, Department of Physics (September 2011 – December 2011, January 2010 – March 2010, January 2008 – June 2008)

Research Assistant, Graduate Research with Dr. Jens H. Gundlach, University of Washington – Seattle, Department of Physics (April 2007 – December 2007)

Courses taught

- Introductory Physics for Scientists and Engineers (Freshman level), Discussion sections and Laboratory Assistance
- Quantum Physics (Junior level) Theory, Discussion section
- Optics Laboratory (Junior level), Laboratory Assistance

Awards and Honors

2006 – 2007 Graduate Opportunities and Minority Achievement Program (GOMAP) Fellowship, University of Washington – Seattle

2006 Hagstrum Award in Physics, University of Minnesota – Twin Cities, Department of Physics and Astronomy

Publications

Emani, P. S.; Olsen, G. L.; Varani, G.; Drobny, G. P.; “Theory of nonrigid rotational motion applied to NMR relaxation in RNA”, *J. Phys. Chem. A*, **2011**, *115*, 12055.

Emani, P. S.; Olsen, G. L.; Echodu, D. C.; Varani, G.; Drobny, G. P.; “Slow exchange model of nonrigid rotational motion in RNA for combined solid-state and solution NMR studies”, *J. Phys. Chem. B*, **2010**, *114*, 15991.

Pederson, K.; Echodu, D. C.; **Emani, P.**; Olsen, G.; Bardaro, Jr., M. F.; Shajani, Z.; Meints, G.; Miller, P.; Varani, G.; Drobny, G. P.; “Unifying Solution and Solid State NMR studies of Nucleic Acid Dynamics”, *Encyclopedia of Magnetic Resonance: Solid State NMR studies of Biopolymers*, **2010**, 535.