PROTON-PROTON DYNAMICS IN PROTEINS BY NMR SPECTROSCOPY

by

Daniel Schmitt Weaver

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in the University of Michigan
2010

Doctoral Committee:

Professor Hashim Al-Hashimi, Chair
Professor Charles L. Brooks, III
Professor Erik R. P. Zuiderweg
Assistant Professor Jason E. Gestwicki

Truth is much too complicated to allow anything but approximations.

– John von Neumann

© Daniel Schmitt Weaver

2010

DEDICATION

To my family

ACKNOWLEDGEMENTS

Thanks to my dad Brian, my mom Heidi, and my brothers Emil and Tyler for being so majestically and astoundingly supportive through all the years and disasters and moves and hilarity. Thanks to my grandparents for all the support they gave us grandkids. No one could ever ask for a better family, or for more devotion. I love you guys.

Thanks to Dr. Erik Zuiderweg for being all the advisor any grad student could ever ask for – a teacher, a mentor, an inspiration, an example, and a great friend in good times and bad. Thanks to my committee for being my committee, and for much good advice, and particularly for their patience. Thanks especially to my committee chair Dr. Hashim al-Hashimi for getting me into this whole NMR business and for helping me out so much along the many halting steps from August 2003 to now.

Thanks to my fellow lab members with whom I've confronted the travails of graduate school. Thanks to Dr. Alex Kurochkin for loving maintenance of the NMR machines and for his unfathomable patience in the explanation of the particulars of spectroscopy and biochemistry. Thanks to Dr. Eric Bertelsen, as a tremendous role model of scientific discipline, clarity and purpose. Thanks to Dr. Akash Bhattacharya for being a great colleague and friend.

Thanks to the teachers who have inspired me in the past. Back in Mexico, Mrs. Hallenbeck from 4th grade for getting me started, Mr. Passer from 11th grade math for drilling the quadratic formula into my brain, and Fr. Stephen Wirkes for sermons that made me think. Drs. Charles Spencer, Peter Seligmann, and Paulen Smith from physics at Ithaca College, who got me started on this whole business, and Drs. Martin Brownstein from politics and Joe Tempesta from history. From Cornell: Drs. Alex Gaeta (quantum mechanics), Earl Kirkland (computational physics), Bruce Kusse (mathematical physics), Manfred Lindau (undergrad advisor, E&M), Lois Pollack (E&M), and Andy Ruina (MATLAB). Particular thanks to the Cornell engineering physics program, which provided exceptional preparation for graduate school and which I recommend wholeheartedly.

Thanks to my friends for their kindness and support. Special shouts out to the Cornell late night engineering physics crew, to the CUAUV team, to Dr. Samuel Stavis, to the halcyon days when UM Biophysics ruled Ashley's, to Von Cramm and Owen co-ops, to the Fleetwood Diner, and to Dr. Jen Hummel. Thanks to the Night Crew, to IRC, and to all the tanks and DPS. Finally, thanks to the designers of *Deus Ex* and thanks to William Gibson for writing *Neuromancer*.

TABLE OF CONTENTS

DEDICATION ii
ACKNOWLEDGEMENTS iii
LIST OF FIGURES vi
LIST OF TABLES ix
LIST OF APPENDICES x
ABSTRACT xi
CHAPTER 1: INTRODUCTION TO η_zK SPECTROSCOPY 1
1.1 Introduction 1
1.2 A brief technical history of fast NMR dynamics spin relaxation experiments 4
1.3 Calmodulin binding to smooth muscle myosin light chain kinase peptide binding domain peptide (smMLCKp) 8
1.4 Ubiquitin dynamics 10
1.5 Hsc70 nucleotide binding domain in the ADP.Pi-bound state 11
CHAPTER 2: THEORY OF η_zK RELAXATION 14
2.1 Theoretical description of the relaxation network 14
2.2 Symbolic representation of the relaxation process 16
2.3 Cross-relaxation and cross-correlated cross-relaxation 18
2.4 Cross-correlated cross-relaxation 25
2.5 Auto-relaxation and cross-correlated contributions to auto-relaxation 27
2.6 Approximating the LCLO system 30
CHAPTER 3: PRACTICAL CONDUCT OF THE η_zK EXPERIMENT 37
3.1 NMR spectroscopy 37
3.2 Data fitting and analysis 40
3.3 Validation of the approximate LCLO system 43
3.4 Calculation of theoretical K_{rig} rates 45
3.5 Estimation of the precision of X-ray coordinates 48
CHAPTER 4: PROTON-PROTON DYNAMICS OF CA^{2+}-SATURATED CALMODULIN 49
4.1 Calmodulin introduction 49
4.2 Calmodulin materials and methods 53
4.3 Calmodulin results 53
4.4 Calmodulin discussion 67
4.5 Effect of the limited precision of experimental protein structures 72
4.6 Structure-free use of *K*-rates 73
CHAPTER 5: PROTON-PROTON DYNAMICS OF UBIQUITIN 76
5.1 Ubiquitin introduction 76
5.2 Ubiquitin materials and methods 79
5.3 Ubiquitin results 80

5.4 Ubiquitin discussion 92
CHAPTER 6: PROTON-PROTON DYNAMICS OF HSC70 NUCLEOTIDE-BINDING DOMAIN 96
6.1 Hsc70 NBD introduction 96
6.2 NMR spectroscopy of the Hsc70 NBD 98
6.3 Hsc70 NBD materials and methods: protein synthesis 104
6.4 Hsc70 NBD materials and methods: spectroscopy 108
6.5 Hsc70 NBD results 109
CHAPTER 7: SUMMARY AND FUTURE DIRECTIONS 131
7.1 Summary 131
7.2 Theoretical advances 133
7.3 Summary of experimental results and advances 136
7.3.1 Calmodulin 136
7.3.2 Ubiquitin 137
7.3.3 Hsc70 NBD 139
7.4 Future directions 141
APPENDIX A: η_zK FITTING CODE (MATLAB) 145
APPENDIX B: η_zK FITTING CODE (FORTRAN 90) 155
REFERENCES 173

LIST OF FIGURES

Figure 2.1. Proton relaxation network corresponding to Figure 2.2 and Table 2.1 16
Figure 2.2: Definition of the energy levels in the proton relaxation network shown in Figure 2.1 17
Figure 3.1: The η_zK pulse sequence..... 37
Figure 3.2. Correspondence between detailed simulation of cross-relaxation processes and equidistant virtual neighbor proton approximation-based K fitting to these processes. 44
Figure 4.1. Ca^{2+}-saturated calmodulin free in solution 51
Figure 4.2. Ca^{2+}-saturated calmodulin bound to smooth muscle myosin light chain kinase CaM-binding domain peptide (smMLCKp)..... 52
Figure 4.3. I-type (NzHa to NzHa) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin 54
Figure 4.4. II-type (NzHa to NzHb) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin 55
Figure 4.5. III-type (NzHb to NzHa) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin..... 56
Figure 4.6. IV-type (NzHa to NzHb) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin..... 57
Figure 4.7. Representative X(t) and Y(t) fits for residue I100 of free Ca^{2+}- calmodulin. . 58
Figure 4.8. Representative X(t) and Y(t) fits for residue I100 of smMLCKp-bound Ca^{2+}-calmodulin. 59
Figure 4.9. Distribution of K fits for residue I100 of CaM/smMLCKp. 60
Figure 4.10. Distribution of η_z fits for residue I100 of CaM/smMLCKp. 61
Figure 4.11. Representative η_{xy} fit for residue I100 of CaM/smMLCKp. 62
Figure 4.12. D_i vs. $P_2(\cos \vartheta_i)$ fit used to determine rotational diffusion constants for free CaM (left) and smMLCKp-bound CaM (right). 63
Figure 4.13. *K*-rates fitted to the experimental relaxation curves for free CaM (top) and CaM/smMLCKp (bottom). 64
Figure 4.14. Q values for free CaM (top) and CaM/smMLCKp (bottom). 66
Figure 4.15. ΔQ between free and smMLCKp-bound calmodulin (Q(bound) – Q(free)). 68
Figure 4.16. Distribution of *Q* values for free calmodulin (open bars) and smMLCKp-bound calmodulin N-terminal domain (closed bars). 69
Figure 4.17. Structural statistics of Q. 70
Figure 4.18. Correlation plot of change in corrected *K* *vs*. change in *Q* upon smMLCKp binding to Ca^{2+}-saturated calmodulin. 73
Figure 4.19. Illustration of ten largest corrected *K* changes (by absolute value) upon smMLCKp binding to Ca^{2+}-saturated calmodulin..... 74

Figure 5.1. Structure of human ubiquitin. 78
Figure 5.2. Representative η_zK relaxation spectrum for human ubiquitin at 27 C with a spectrometer field of 500 MHz. 80
Figure 5.3. Representative X(t) and Y(t) fits for residue V5 of ubiquitin. 81
Figure 5.4. Distribution of K fits for residue V5 of ubiquitin. 82
Figure 5.5. Distribution of η_z fits for residue V5 of ubiquitin. 83
Figure 5.6. Representative η_{xy} fit for residue V5 of ubiquitin. 84
Figure 5.7. D_i vs. $P_2(\cos \vartheta_i)$ fit used to determine rotational diffusion constants for ubiquitin. 85
Figure 5.8. K rates for ubiquitin. 86
Figure 5.9. η_z rates for ubiquitin. 87
Figure 5.10. Histogram of correlation coefficients between experimental K rates and calculated K_{rig} rates within several ubiquitin NMR structures. 90
Figure 5.11. Q values from the 1D3Z ensemble (blue) along with NH S^2 amide backbone parameters for ubiquitin as measured by classical ^{15}N relaxation methodology (black). 92
Figure 5.12. Experimental K rates compared with K_{rig} rates calculated from all conformers of several NMR structures. 95
Figure 6.1. Representative spectrum of 0.3 mM protonated ^{15}N-labeled Hsc70 NBD spectra for an I-type η_zK experiment (NzHa to NzHa) at 10 ms relaxation time. 99
Figure 6.2. Representative Hsc70 NBD I-type (NzHa to NzHa) spectra at 400 ms relaxation time. 101
Figure 6.3. Representative Hsc70 NBD II-type (NzHa to NzHb) spectra at 400 ms relaxation time. 101
Figure 6.4. Representative Hsc70 NBD III-type (NzHb to NzHa) spectra at 400 ms relaxation time. 102
Figure 6.5. Representative Hsc70 NBD IV-type (NzHb to NzHb) spectra at 400 ms relaxation time. 102
Figure 6.6. Structure of bovine Hsc70 NBD. 104
Figure 6.7. Representative time course of relaxation for I, II, III and IV-type experiments on residue E27 of Hsc70 NBD. 110
Figure 6.8. η_zK fit to symmetric reconversion data for residue E27 of Hsc70 NBD. 111
Figure 6.9. Bootstrap distribution of K fits for residue E27 of Hsc70 NBD. 112
Figure 6.10. Bootstrap distribution of η_z fits for residue E27 of Hsc70 NBD. 113
Figure 6.11. η_{xy} fit for the Wang/Rance/Palmer η_{xy} experiment on residue E27 of Hsc70 NBD. 114
Figure 6.12. D_i vs. $P_2(\cos \vartheta_i)$ fit used to determine rotational diffusion constants for Hsc70 NBD. 115
Figure 6.13. K rates for the Hsc70 NBD. 117
Figure 6.14. K rates for the Hsc70 NBD (amide proton exchanging residues not included). 118
Figure 6.15. η_z rates for the Hsc70 NBD. 119
Figure 6.16. K_{rig} values derived from multiple structures of Hsc70 NBD. 120
Figure 6.17. Comparison of Cruickshank-calculated errors in K_{rig} with multiple-structure-based errors in K_{rig} for Hsc70 NBD. 121

Figure 6.18. Comparison of K and K_{rig} values for Hsc70 NBD. 123
Figure 6.19. Comparison of K and K_{rig} values for Hsc70 NBD (detail of lower K rates). 124
Figure 6.20. Correlation plot of K and K_{rig} values for Hsc70 NBD..... 125
Figure 6.21. Correlation plot of K and K_{rig} values for Hsc70 NBD (experimental error in K indicated). 126
Figure 6.22. Q as a function of percentage of K_{rig} rate contributed by amide protons reporting low K_{rig} rates. 128
Figure 6.23. Histogram of Q values calculated in Figure 6.15 for 2QWL. 130

LIST OF TABLES

Table 2.1. Longitudinal relaxation rates for a ^{15}N-^{1}H amide spin system with two neighbor protons (Figure 2.1).....29
Table 4.1. CaM axially symmetric diffusion tensor characteristics.....63
Table 4.2. Mean Q values for N- and C-terminal domains of free and smMLCKp-bound Ca^{2+}-saturated calmodulin.....67
Table 5.1. Ubiquitin axially symmetric diffusion tensor characteristics.....85
Table 5.2. CaM free/bound and ubiquitin X-ray structure K vs. K_{rig} correlations.....88
Table 5.3. Ubiquitin NMR structure K vs. K_{rig} correlations.....89
Table 5.4. Ubiquitin NMR structure mean Q and standard deviation of Q.....91
Table 6.1. Optical density (600 nm) measurements of E. coli growth in 1 L D_2O M9 culture.....106
Table 6.2. Hsc70 NBD axially symmetric diffusion tensor characteristics.....116
Table 6.3. Correlation coefficients R between K_{rig} and experimental K rates for Hsc70 NBD structures.....129

LIST OF APPENDICES

APPENDIX A: η_zK FITTING CODE (MATLAB)..... 145
APPENDIX B: η_zK FITTING CODE (FORTRAN 90) 155

ABSTRACT

PROTON-PROTON DYNAMICS IN PROTEINS BY NMR SPECTROSCOPY

by

Daniel Schmitt Weaver

Chair: Hashim Al-Hashimi

Proteins are dynamic objects that undergo motion over a wide range of timescales from femtoseconds to hours. Changes in protein dynamics affect important protein operating parameters such as overall protein stability, ligand affinity, allosteric regulation, and catalytic efficiency. In this work, I describe the measurement of proton-proton dynamics in proteins by means of NMR relaxation experiments.

The HSQC-TROSY-based η_zK NMR experiment constitutes the methodological basis of this dissertation. This experiment measures the K rate for each amide proton, which is the sum of all individual zero-quantum proton-proton relaxation rates between the amide proton and its proton neighbors. Each individual relaxation rate can be modulated by dynamics along its interproton vector. The η_zK experiment also measures η_z longitudinal amide N/NH DD/CSA cross-correlated relaxation rates. These rates are incorporated into a protocol for measuring protein rotational diffusion tensors.

Proton-proton dynamics in proteins are analyzed by comparing K rates with K_{rig} rates calculated for the case of a rigid protein. The ratio of K rates to K_{rig} rates, Q, provides a measurement of distance-weighted average proton-proton dynamics near amide proton sites.

Measurement of Ca^{2+}-saturated calmodulin Q parameters in free and peptide-bound forms indicates proton-proton dynamics throughout both forms. Significant rigidification of the proton network takes place on peptide binding.

η_zK data on ubiquitin is used to analyze agreement between experimental K rates and published NMR structures of ubiquitin, which all suggest proton-proton dynamics. Based on these observations, correlation between K rates and K_{rig} rates is introduced as a measurement of the quality of a proposed protein structure.

Finally, the η_zK experiment is applied to the nucleotide-binding domain (NBD) of the Hsc70 protein. This 44 kDa protein lies beyond the range of practical NMR relaxation spectroscopy prior to development of η_zK experimental techniques; thus, it serves as an explicit test of the method. Experimental K rates of Hsc70 NBD display good correlation with theoretical K rates calculated from three X-ray structures.

In conclusion, the η_zK experiment provides a means of measuring proton-proton dynamics in large proteins. These results will broaden our knowledge of protein dynamics and expand the experimental capabilities of the field.

CHAPTER 1: INTRODUCTION TO η_zK SPECTROSCOPY

1.1 Introduction

Gibbs free energy is the thermodynamic currency of chemistry in closed systems. Processes with overall negative Gibbs free energy proceed spontaneously; processes with positive Gibbs free energy do not. If a step with positive Gibbs free energy is required in a chemical process, it must be coupled to a greater, negative Gibbs free energy change in order to drive the process forward. In open systems, the grand potential plays a similar role, with additional considerations added for the chemical potential of molecules entering and leaving the system.

Proteins' ability to structurally couple positive and negative Gibbs free energy changes is a major underpinning of the chemistry of life. Binding and chemical modification events cause numerous localized changes in free energy throughout the protein [1; 2]. These changes are summed over the molecule to produce the total activation energy and Gibbs free energy change for the molecular action in question.

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

The Gibbs free energy equation is divided into enthalpic (ΔH) and entropic (TΔS) components. The enthalpic component reflects the internal energy of the system as determined by its stereochemical and electrical arrangements. The entropic component of the Gibbs free energy reflects the distribution of configurations available to the system.

Most systems of immediate research interest involve protein-ligand or protein-protein interactions. In these systems, the distribution of configurations incorporates all members of the system, including the solvent, ligand, and protein.

The connection between protein dynamics, protein conformational entropy and protein operation constitutes a major ongoing area of research in protein biophysics.[3; 4] Changes in protein conformational entropy have long been recognized as important contributors to change in overall entropy during protein processes.[5] Advances in macromolecular simulation techniques demonstrated that the residual entropy of a folded protein is an order of magnitude larger than the entropy of protein denaturation[6] while the Cooper-Dryden model of allosteric change[7] indicated a theoretical avenue for direct contribution of changes in protein entropy to allosteric communication. The richness of potential discovery in this field has stimulated great scientific interest.

Bridging the gap from experimental measurement of changes in protein dynamics to predictions of changes in protein entropy presents a thorny problem for protein biophysicists. The great complexity of protein structure presents one major obstacle, with entropically rich configurational diversity present at organizational levels stretching from the domain to the individual sidechain functional group. The other major difficulty lies in the numerous timescales of motion involved in protein dynamics, which reach from picoseconds to multiple seconds. Motion may take place over multiple timescales at the same length scale.

This breadth of length and time scales forces the employment of multiple experimental techniques capable of sensing different types of protein motion. It also places exceptional demands on molecular dynamics programs, which must accurately replicate protein dynamics at both extremely large and extremely small scales. Many types of dynamics take place at scales beyond the current practical reach of detailed molecular mechanics programs (CHARMM[8]/AMBER[9]/GROMOS[10]) and must be addressed by coarse-grain methods such as normal-mode analysis.

NMR spectroscopy is a central technique in the experimental study of protein dynamics and has provided a wealth of information about the extent of motions in the protein backbone and in selected sidechains. Experimental techniques have been developed for studies of protein dynamics at timescales including picoseconds to nanoseconds, microseconds to milliseconds, and even broad-based measures spanning nanoseconds to seconds.

The development of NMR theory[11] and experiment[12] capable of quantitating bond

vector dynamics in solution led to interest in relating the configurational entropy encapsulated in bond vector disorder to the order parameters used to quantitate NMR measurements of bond vector disorder. This interest bore fruit in the work of Akke et al.[13], Li et al.[14] and Yang and Kay[15], providing relations dependent on the choice of motional model for bond vector disorder.

To illustrate the arguments used, I summarize the results of the classical (as opposed to the quantum mechanical) version of Yang and Kay's formulation of the relation. The conformational entropy S_p contributed by the rotation of the *j*th bond vector in the presence of a vector orientation-dependent potential $U(q)$ is expressed as

$$S_p(j) = -k_B \int_V p(q) \ln\left(p(q)\right) dV \tag{2.2}$$

where $p(q)$ is the probability density function given by

$$p(q) = \frac{e^{-U(q)/k_B T}}{\int_V e^{-U(q)/k_B T} dV}, \tag{2.3}$$

k_B is Boltzmann's constant, T is the temperature in Kelvin, q represents the coordinates of the bond vector and volume integrals are conducted over the extent of the probability density function.

Meanwhile, the expression for the Lipari-Szabo NMR order parameter[11] also incorporates the probability density function in normalized form:

$$S_{LZ}^2 = (4\pi/5) \sum_r \left| \int_V p(q) Y_{2r}(q) dV \right|^2 \tag{2.4}$$

where the probability density function is assumed to be normalized and Y_{2r} are the second-order spherical harmonics. The normalized probability density function $p(q)$ derived from the bond potential $U(q)$ now serves as a bridge between the conformational entropy of the bond vector S_p and the Lipari-Szabo order parameter S^2_{LZ}. Making the (practically difficult) choice of a motional model manifest in $p(q)$ allows the

conformational entropy of a given bond vector to be derived from interpretation of S^2_{LZ} in terms of that motional model.

NMR spectroscopists are continually challenged to increase the sensitivity and molecular scope of their experiments in order to learn more about the full extent of protein motions and provide an improved foundation for molecular modeling. The extension of conformational entropy measurements to sidechains[16; 17; 18; 19] and non-amide backbone vectors[20] revealed major new reservoirs of residual protein entropy. With every new protein interatomic vector whose dynamics become quantifiable, new insights emerge regarding the conformational dynamics of proteins and their role in the essential biophysical thermodynamics of protein operation.

This thesis describes the development of new NMR methodology for studying fast ps-ns timescale motions in proteins of up to 70 kD. The experimental techniques involved are based on TROSY-HSQC experiments and measure rates of amide proton spin flip processes parallel to the large-molecule nuclear Overhauser effect. The bond vectors probed are nonbonded proton-proton vectors between the amide proton and its neighbor protons.

In order to place the work described into historical context, I present here a brief history of the use of NMR to study dynamics on the picosecond-nanosecond timescale.

1.2 A brief technical history of fast NMR dynamics spin relaxation experiments

The recent history of fast NMR dynamics begins with the 1989 measurement of ^{15}N amide backbone dynamics in staphylococcal nuclease.[12] This experiment sets the tone for NMR dynamics investigations over the next two decades. Nearly 400 publications were based on similar methodology as of 2006, when the last major review of the field was published.[21] These publications comprise the vast bulk of the field of fast NMR dynamics.

^{15}N amide backbone dynamics experiments measure the relaxation of nitrogen magnetization stemming from random motions of the ^{15}N-^{1}H$_N$ bond vector. In their simplest form, these experiments constitute the measurement of ^{15}N R_1, ^{15}N R_2 and ^{15}N-

$^{1}H_{N}$ heteronuclear NOE relaxation rates and subsequent extraction of model-free dynamics parameters[11; 21] from the relaxation data using ModelFree[22; 23] or similar programs.

Relaxation rates are conventionally measured by tracking the time evolution of the intensity of ^{15}N-^{1}H crosspeaks in HSQC-based[24] relaxation experiments.[25; 26] With the advent of the TROSY-HSQC methodological revolution[27] significant development of TROSY-based amide backbone dynamics experiments[28; 29; 30] has taken place although TROSY-based measurements comprise a smaller fraction of published results (45 citations of Zhu et al. as of this writing).

Methyl dynamics studies in sidechains constitute the second major current within fast NMR dynamics. These experiments provided the first robust probes of sidechain motion in proteins. Methyl groups are distributed unevenly through the protein and consequently do not provide an ecumenical picture of motion. The study of their dynamics has, however, been invaluable for revealing levels of motion significantly greater than those found along the amide backbone.[31]

Methyl dynamics can be probed with ^{13}C- or ^{2}H-based[32; 33; 34] experiments. In either case, specialized labeling strategies are required in order to create 'clean' spin systems which can be prepared, evolved and observed free from the interference of spurious relaxation mechanisms. In modern work, ^{13}C-based experiments have largely been supplanted by ^{2}H-based experiments due to more straightforward labeling and interpretation of the latter[31] as well as the development of TROSY-HMQC methyl spectroscopy[35] for use in larger proteins of modern NMR interest. As of this writing there are 114 citations of Nicholson et al.[32], 95 citations of Muhandiram et al.[33], and 13 citations of Ollerenshaw et al.[36]

Modern ^{2}H-based methyl dynamics methods measure the relaxation of spin states prepared in $^{13}CH_2D$ (two ^{1}H and one ^{2}H nuclei)[34] or $^{13}CHD_2$ (one ^{1}H and two ^{2}H nuclei)[36] methyl groups. Several different types of carbon-deuterium spin states are prepared. These states are allowed to relax, and the remaining magnetization is observed as ^{13}C-^{1}H methyl crosspeaks in TROSY-HMQC-based relaxation experiments.

Model-free dynamics parameters explaining the relaxation behavior of these several carbon-deuterium spin states are extracted using spectral density mapping.[37; 38]

The dynamically active bond vector characterized by these parameters is the symmetry axis of the methyl group and is variously referred to as S^2_{axis} or O^2. Methyl dynamics experiments typically report more extensive dynamics than backbone dynamics experiments[16; 21; 31] suggesting a 'rigid backbone, mobile sidechains' model of protein dynamics.

Numerous fast NMR dynamics experiments exist apart from amide backbone and methyl dynamics categories. The array of spin states and relaxation mechanisms probed by these experiments is broad. Notable examples include $^{13}C\alpha$-$^{1}H\alpha$ backbone dynamics, $^{13}C'$ backbone dynamics. and $^{13}C'$-$^{13}C\alpha$ dynamics[21; 39; 40; 41; 42] as well as studies of histidines, methylenes, and cross-correlated relaxation in methyl groups.[31]

The proton-proton homonuclear NOE[43] is a relaxation-driven process. NOE buildup rates, and consequently the observed strength of NOE intensities, are affected by fast internal motions along the nonbonded vector connecting the NOE donor and acceptor protons.[44] Since the proton NOE constitutes the foundation of NMR structure determination methods, and proper binning of NOE intensities into strong, moderate, and weak intensities is essential to structural accuracy, considerable research effort has gone into determining the degree to which fast internal motions might diminish the apparent strength of NOE intensities. Numerous molecular dynamics efforts[45; 46; 47; 48; 49; 50; 51] to address the question indicate that, in general, NOE structure determination methods are qualitative and robust enough to survive the inaccuracies introduced by internal motions as long as globular proteins are under consideration. Unfolded proteins and peptides are more challenging subjects[52; 53] whose apparent NOE contacts suffer from seriously misleading motional averaging.

Experimental attempts to measure dynamics based on the homonuclear NOE are comparatively few in number. The sole direct approach in proteins[54] uses 1D NMR to measure the buildup curves of a handful of strong proton resonances in BPTI sidechains and compare these buildup curves with those expected freom theoretical calculations based on sidechain molecular structure. Significant technical difficulties with experiment design and interpretation have prevented further traditional experimental investigation of the nonbonded internal motions probed by homonuclear NOEs.

A more recent approach to proton-proton dynamics in proteins relies on compensation of NOE and ROE effects during off-resonance ROESY experiments.[55] These experiments observe crosspeaks between Hα and Hβ atoms in sidechains of small proteins and indicate the presence of significant motion along the Hα-Hβ vector.

With the exception of the methyl dynamics experiments, the protein dynamics experiments listed above are limited to employment in smaller proteins. For many of these experiments, this limitation stems from their construction on the foundation of the HSQC experiment. HSQC chemical shift labeling, mixing, and observation periods are carried out without regard for selection of slowly-relaxing magnetic populations, leading to severe line broadening. In general, relaxation experiments developed for the large-protein environment must be constructed so as to place their observables in magnetization states that can be easily processed by large-protein spectroscopic techniques such as HSQC-TROSY or methyl TROSY.

Relaxation experiments designed for employment in large proteins must also spend a minimum of time in preparation of initial magnetization states. Spending too much time in preparation of magnetization leads to poor initial magnetization intensity and consequent low signal to noise. Excessive time spent with magnetization in the transverse plane during the preparation period is especially costly due to extremely rapid transverse relaxation in large proteins. Experiments requiring quantitative re-establishment of equilibrium magnetization, such as those measuring the 15N-1H heteronuclear NOE as in classical backbone dynamics studies, are particularly difficult to carry out in large proteins. This is due to slow longitudinal relaxation rates (long T_1) in large proteins, which retard the process of return to equilibrium.

Despite methyl dynamics experiments' suitability for use in large proteins, we are interested in developing new fast protein dynamics experiments in order to explore motions along new vectors and broaden our knowledge of the extent of dynamics in proteins. The development of methyl and non-classical backbone dynamics experiments opened the door to a substantially expanded picture of dynamics contributions to residual protein entropy, and discovered considerable entropic content in sidechains and within backbone motions poorly probed by classical amide backbone dynamics.

The η_zK experiment detailed in this dissertation represents the fruit of this interest, and combines elements from several of the experimental approaches described above. As in ^{15}N amide backbone dynamics experiments, magnetization is prepared within the amide nitrogen-proton spin system. However, due to the way magnetization is prepared, the relaxation of the system is dominated by proton-proton homonuclear relaxation. Symmetric reconversion[56] allows partial decoupling of the effects of proton-proton homonuclear relaxation from the effects of nitrogen-proton heteronuclear relaxation, and allows both to be characterized at once.

One of the relaxation rates measured in our experiment, the longitudinal cross-correlated relaxation rate η_z, is part of the heteronuclear relaxation behavior of the ^{15}N-^1H_N system. This relaxation rate is used in the determination of protein global rotational correlation time, a subject we will treat on below. The other relaxation rate measured in our experiment is somewhat novel for protein relaxation studies, and constitutes the sum of all proton-proton homonuclear NOE buildup rates between the amide proton and its neighbor protons. We label this rate K.

Structural information derived from other sources – high resolution crystal structures are particularly convenient – can be used to establish a benchmark for what we might expect K rates to be, given a rigid structure. Motion along interproton vectors will tend to reduce K rates below this rigid structural benchmark. A complete analysis of the η_zK experiment measures a distance-weighted average of order parameters for the vectors between amide protons and their neighbor protons. This provides a unique 'regional average' picture of dynamics within NOE range of the amide protons.

1.3 Calmodulin binding to smooth muscle myosin light chain kinase peptide binding domain peptide (smMLCKp)

Calmodulin binding to smooth muscle myosin light chain kinase peptide binding domain peptide (smMLCKp) serves as the initial demonstration of the η_zK experiment and its use in studying protein dynamics. As is obvious from the name, smMLCKp is a peptide homologue of the calmodulin binding site of smooth muscle myosin light chain

kinase (smMLCK).[57; 58; 59] The peptide in question extends over smMLCK residues 470-490.[16] Calmodulin binding to this domain is an essential part of smMLCK activation by disruption of interactions between the smMLCK catalytic site and autoinhibitory domain[60]; however, calmodulin binding is involved in the regulation of an enormous list of cellular processes.[59] Calmodulin must therefore tune its binding energy to achieve high specificity across a range of binding targets, while retaining moderate affinity so that CaM-activated proteins can be efficiently deactivated as necessary.

In order to achieve these goals, calmodulin employs an energetic mechanism consisting of a Gibbs free energy tradeoff between entropy and enthalpy.[19; 61; 62] The ΔG between Ca^{2+}-saturated calmodulin free in solution and Ca^{2+}-saturated calmodulin bound to smMLCKp is a relatively moderate -11 kcal/mol. This ΔG is a combination of a negative ΔH of around -16 kcal/mol and a $-T\Delta S$ contribution of around +5 kcal/mol. The small positive free energy contribution is surprising for a process in which a great deal of free calmodulin's exposed hydrophobic surface area is buried against the interface with the peptide, a phenomenon normally thought to present a very entropically favorable contribution to Gibbs free energy changes. The change in buried surface area would be expected to produce a $-T\Delta S$ contribution of around -45 kcal/mol. Unbalanced, this large negative entropic Gibbs free energy change would vastly overtune the affinity of calmodulin for smMLCKp.

NMR experiments[16; 19] indicate that this negative Gibbs free energy change is counterbalanced by a large positive conformational entropic contribution to ΔG that constitutes approximately +50 kcal/mol and stems from rigidification of the free calmodulin ensemble into a narrowed structural ensemble capable of binding calmodulin. This rigidification is parcelled among macroscale entropy loss (binding of the EF-hand domains around the peptide-binding domain) and quenching of dynamics within the molecular bond network.

The burial of exposed hydrophobic surface area against matching peptide sidechains and the presence of enthalpic interactions with the peptide leads to high specificity, since only an exact match with the smMLCKp target sequence will allow the rigidification cost to be paid back by favorable enthalpic and hydrophobic entropic

interactions. This high specificity is accompanied by only moderate affinity, since the large counterbalanced entropic and enthalpic costs lead to a total $\Delta G = -11$ kcal/mol.

The quenching of calmodulin molecular bond dynamics upon smMLCKp binding has been studied extensively by experimental NMR spectroscopy[16; 19] and MD simulation.[63] These studies have lent unambiguous support to the concept that quenching of fast dynamics contributes a large portion of the entropy loss seen upon peptide binding. Ultra-high-resolution X-ray crystal structures of Ca^{2+}-saturated calmodulin in free (1EXR)[64] and bound (2O5G)[65] states are available, as are highly concentrated NMR samples of both free and bound forms of the molecule. This combination of circumstances made calmodulin binding an excellent testbed for η_zK relaxation methodology development.

In the fourth chapter of this thesis, I describe the use of the η_zK experiment to study proton-proton dynamics in free and bound calmodulin, and compare the results of these studies with earlier NMR dynamics studies. These experiments confirm quenching of fast dynamics on Ca^{2+}-calmodulin binding to smMLCKp, broaden the demonstrate the basic experimental and statistical techniques used in measuring NMR proton-proton dynamics, and illustrate the capabilities and limitations of the η_zK experimental environment.

1.4 Ubiquitin dynamics

Ubiquitin is a 76-residue, 8.5 kDa regulatory protein present in all eukaryotic cells. It is the covalent label used in the essential cellular post-translational modification process of protein ubiquitination, in which one or more copies of ubiquitin are covalently attached to target proteins by attaching ubiquitin's flexible C-terminal tail to a lysine on the target protein. Since ubiquitin itself is rich in exposed lysines, several ubiquitin molecules can be ligated to a given protein, forming polyubiquitin chains.

The process of ubiquitination marks the target protein for one of several cellular fates. Most famous is the proteasomal degradation pathway that leads to proteolysis, but

ubiquitin-based targeting underlies dozens of other cellular targeting localization processes.[66]

Ubiquitin has a long history of involvement in NMR dynamics experiments for several reasons. It is small (76 residues) and has an exceptionally well-resolved NMR spectrum. It can be concentrated to a high degree (1-2 mM) and tumbles rapidly and almost isotropically, leading to excellent signal-to-noise. It is extremely stable in solution, allowing prepared samples to last for years. Finally, as is often the case in science, ubiquitin is well-studied because it is well-studied. Most aspects of the molecule's biophysics have been extensively characterized, and it is therefore easy to draw inferences about the consequences of ubiquitin observations.

A large number of NMR-based structures of ubiquitin exist, each generated in the course of research into new methods of generating NMR structures. Five notable recent structures include 1D3Z[67] (a 10-member ensemble refined with a combination of NOE and RDC data), 1XQQ[68] (a 128-member ensemble refined with a combination of NOE and MD data), 2NR2[69] (a 144-member ensemble using a combination of NOE and MD data with a different refinement procedure), 2K39[70] (a 116-member ensemble based on extensive RDC analysis), and 2KLG[71] (a 20-member ensemble based on paramagnetic solvent structure determination techniques).

This cornucopia of structural hypotheses constitutes an interesting opportunity to put the η_zK experiment to work. In the fifth chapter of this thesis, I will compare the five NMR structures described above with K rates derived from experiment, under the hypothesis that higher structural fidelity to the true solution state of the structure will lead to a closer relationship between theoretical K_{rig} rates calculated from structure and experimental K rates derived from the η_zK experiment. The results of this comparison suggest that K rates hold considerable promise as an additional restraint for use in NMR structural biology.

1.5 Hsc70 nucleotide binding domain in the ADP.Pi-bound state

The highly conserved Hsp70 family of chaperone proteins are present throughout the dominion of life and play vital roles in numerous cellular processes[72]. Several members of the family have been studied in our laboratory group, including bovine Hsc70, E. coli DnaK, and Thermus thermophilus DnaK. All members of the Hsp70 family consist of three domains. These are the 44 kD N-terminal nucleotide binding domain (NBD), which binds adenosine nucleotides that serve as allosteric regulators of substrate binding domain binding state; the 15 kD substrate binding domain (SBD), which is conncted to the NBD by a 6-10 residue hydrophobic linker and binds exposed hydrophobic regions of misfolded proteins; and a 10 kD C-terminal lid domain that tunes the kinetics of substrate binding and release. The structural and biochemical characteristics of each of these domains are well studied. Although the biochemical operation of the complete machine is also well understood, the structural basis of this operation remains an active topic of study.

In the ADP.Pi nucleotide state, bovine Hsc70 NBD provides an ideal testbed for studies of NMR dynamics methodology in large proteins. It is stable in solution, has a well-dispersed NMR spectrum, can be prepared at high sample concentrations, and is large enough to provide a serious challenge to the spectroscopist without being so large as to make NMR spectroscopy impossible. Several static X-ray structures of the Hsc70 NBD exist (3HSC[73], 1HPM[74], and 2QWL[75]) providing several independent and relatively high-resolution (1.75-1.93 Å) benchmarks for purposes of comparing dynamics experiment with static structure.

The ultimate test of the η_zK experiment is whether it can provide information on dynamics in an actual large protein. Measuring any form of dynamics in bovine Hsc70 NBD is an ambitious goal: due to a paucity of existing methodology, there are only a handful of NMR dynamics studies on proteins of this size or larger. In the fifth chapter of this work, I demonstrate the employment of the η_zK experiment on triple-labeled, perdeuterated bovine Hsc70 NBD (1-386) and compare the results of the η_zK experiment to expectations from the available X-ray structures. η_zK methodology is shown to be viable for studying proton-proton dynamics in moderately large proteins such as bovine Hsc70 NBD. Experimental K rates demonstrate excellent agreement with structural K_{rig} trends. The validity of the Cruickshank-based error estimation method is called into

question, with significantly less variability among K_{rig} rates calculated from independent structures than is indicated by Cruickshank-type error propagation.

CHAPTER 2: THEORY OF η_zK RELAXATION

2.1 Theoretical description of the relaxation network

Consider the longitudinal magnetization of an amide NH spin system in a protein at the large-molecule limit. We prepare √2NzHa and √2NzHb magnetization from linear combinations of nitrogen one-spin order Nz and two-spin order 2NzHz.

$$\sqrt{2}N_zH_\alpha = \frac{\sqrt{2}}{2}\left(N_z + 2N_zH_z\right) \tag{2.1}$$

$$\sqrt{2}N_zH_\beta = \frac{\sqrt{2}}{2}\left(N_z - 2N_zH_z\right) \tag{2.2}$$

These are the proton-up and proton-down components of nitrogen Z magnetization, respectively. We will refer to them collectively as LCLO major states, LCLO standing for Linear Combinations of Longitudinal Order. We are interested in modeling the auto- and cross-relaxation behavior of the LCLO major states. For an isolated amide two-spin system, auto and cross-relaxation behavior are described by the following Liouville representation:

$$\frac{d}{dt}\begin{bmatrix} \sqrt{2}N_zH_\alpha \\ \sqrt{2}N_zH_\beta \end{bmatrix} = -\begin{bmatrix} \lambda+\eta_z & -\zeta \\ -\zeta & \lambda-\eta_z \end{bmatrix}\begin{bmatrix} \sqrt{2}N_zH_\alpha \\ \sqrt{2}N_zH_\beta \end{bmatrix} \tag{2.3}$$

with

$$\lambda = \frac{d_{NH}^2}{8}\begin{pmatrix} 6J(\omega_N) + J(\omega_H - \omega_N) \\ +3J(\omega_H) + 6J(\omega_H + \omega_N) \end{pmatrix} + \frac{c_N^2}{3}J(\omega_N) + \frac{c_N^2}{6}J(\omega_H) \quad (2.4)$$

and

$$\eta_z = c_N d_{NH} J_X^{NH/N}(\omega_N) \quad (2.5)$$

and

$$\xi = -\frac{d_{NH}^2}{8}\left(J(\omega_H - \omega_N) - 3J(\omega_H) + 6J(\omega_H + \omega_N)\right) - \frac{c_H^2}{6}J(\omega_H) \quad (2.6)$$

where $c_X=B_0\gamma_X\Delta_X$, $d_{XY}=(\mu_0/4\pi)\gamma_X\gamma_Y r_{XY}^{-3}$, B_0 is the spectrometer static magnetic field, γ_X is the gyromagnetic moment of nucleus X, Δ_X is the magnitude of the axial CSA of nucleus X, μ_0 is the permeability of free space, r_{XY} is the distance between nuclei X and Y, ω_X is the Larmor frequency for nucleus X, $J(\omega_X)$ is the autocorrelation spectral density function for rotational motion of a particular vector at Larmor frequency X, and $J^X_{XY/X}(\omega)$ is the cross-correlation spectral density function for relative motion between the internuclear vector between nuclei X and Y and the CSA principal axis of nucleus X at Larmor frequency ω.

For the isolated two-spin system, all relaxation pathways are dominated by spectral densities at frequencies ω_N or larger, the relevant rates are on the order of 0.01-0.1 s-1. However, the LCLO states experimentally (see Chapters 4-6) cross-relax at rates one to two orders of magnitude faster due to interactions with extraneous protons.

2.2 Symbolic representation of the relaxation process

The explanation for this fast cross-relaxation behavior lies in the network of protons surrounding the NH spin system. The presence of these protons, each of which may be in an a or b state, splits each of the two LCLO major states into several substates. √2NzHa splits into NzHaH'aH"a..., NzHaH'aH"b..., NzHaH'bH"a... and so on, while √2NzHb splits into NzHbH'aH"a..., NzHbH'aH"b..., NzHbH'bH"a... and so on. (For clarity of notation, we will neglect the normalizing coefficients of LCLO substates.) The total population in each LCLO state is the sum of their substate populations (see Figure 2.2).

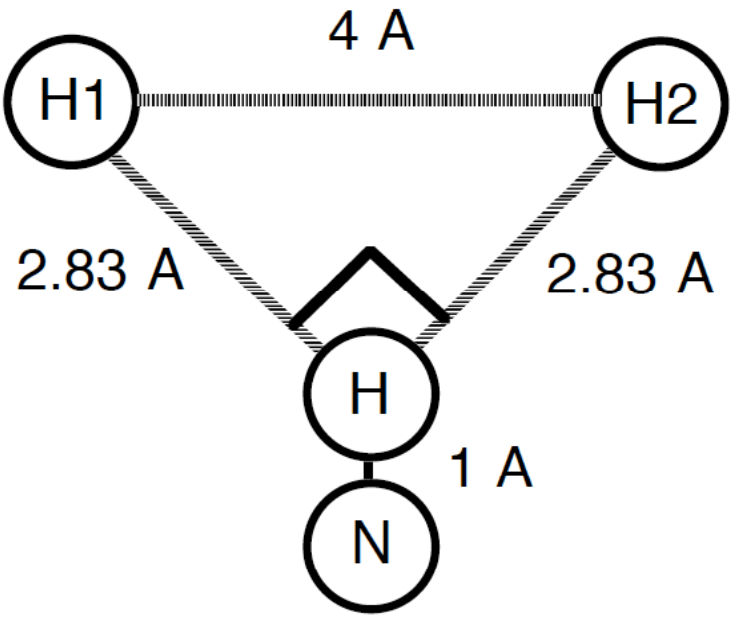

Figure 2.1. Proton relaxation network corresponding to Figure 2.2 and Table 2.1

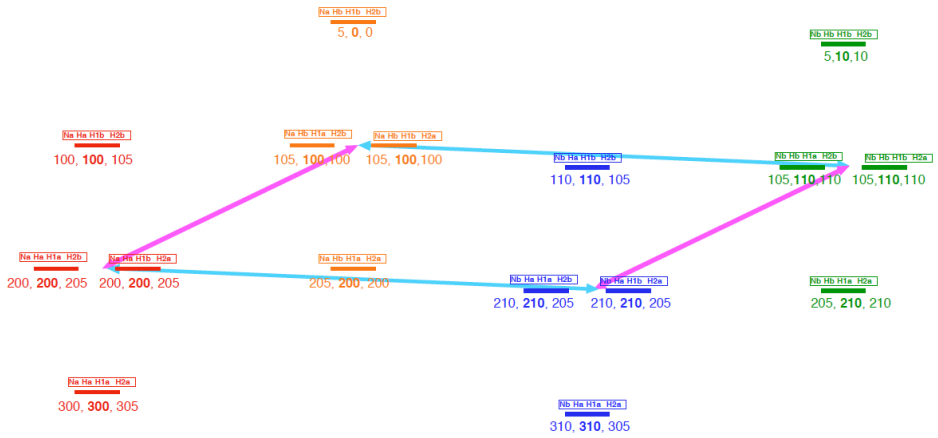

Figure 2.2: Definition of the energy levels in the proton relaxation network shown in Figure 2.1
The energy levels in Figure 2.2 are used for the relaxation matrices and Table 2.1 in the theory section of this work. NaHa levels are shown in red; NaHb levels in orange; NbHa in blue; NbHb in green. Cyan arrows symbolize 4 ^{15}N transitions each; magenta arrows symbolize 4 amide hydrogen transitions each. The numbers indicate the relative populations in the NzHa, equilibrium, and NzHb state respectively.

Figure 2.1 depicts an amide NH spin system with two neighbor protons, and Figure 2.2 shows the associated energy level diagram. As is apparent from Figure 2.2, the inclusion of the proton network surrounding the NH spin system presents another pathway for cross-relaxation between the LCLO major states. This pathway consists of zero-quantum 'spin flip' transitions between one proton in a given spin state and a neighbor proton in the opposite spin state. For example, NzHaH'bH"a... may undergo a zero-quantum transition to NzHbH'aH"a..., or to NzHaH'aH"b.... For LCLO network protons i and j, zero-quantum cross-relaxation will take place between states Nz...Hia...Hjb... and Nz...Hib...Hja.... Density matrix calculations give the relaxation rate governing this spin-flip process as:

$$(2.7)\; K_{ij} = \frac{1}{4} d^2_{HiHj} J_{HiHj}(0)$$

where the spectral density function $J_{HiHj}(0)$ is

$$J_{HiHj}(0) \approx S^2 R \frac{2}{5} \tau_c \quad (2.8)$$

as defined in Post.[48] This definition of $J_{HiHj}(0)$ incorporates the effects of radial (R) and angular (S^2) motion of the relaxing protons i and j with respect to one another, assuming that these motions are not correlated. This is a necessary extension in situations where the relaxing nuclei are not constrained by a connecting chemical bond. Further information on systems of this kind can be found in Olejniczak et al.[45], Brüschweiler et al.[49], and Abseher et al.[50]

The zero-quantum transition rate defined in Equation (2.8) is approximately 1-3 s^{-1} for close interproton distances in proteins with characteristic global diffusion times τ_c = 4-7 ns, and faster for larger proteins with slower global diffusion times. These relaxation rates match the order of magnitude of experimentally observed LCLO cross-relaxation behavior. We will first develop the fitting model in the large molecule/high magnetic field limit, defined such that zero-quantum relaxation terms dwarfs all other relaxation mechanisms. Subsequently, we will bring back all small omitted terms and assess their cumulative effect on the relaxation processes.

In this work, we use the η_zK pulse sequence[76] incorporating symmetrical reconversion techniques[56]. In the following, we assume that this separation has been made; hence we will exclusively focus on the other mechanisms that contribute to K. The theoretical treatment provided here is found in large part in our publication on amide proton spin flip rates and changes in Ca^{2+}-saturated calmodulin order parameter on smooth muscle myosin light chain kinase peptide binding[77], which also forms the basis of sections of Chapter 3 and Chapter 4.

2.3 Cross-relaxation and cross-correlated cross-relaxation

We now construct the system of first order differential equations governing cross-relaxation in the LCLO substate system depicted in Figure 2.2, considering only zero-quantum spin-flip terms in the relaxation matrix. In the following we denote the amide

proton as H^A, the first neighbor proton as H^1, and the second neighbor proton as H^2. The matrix is related to a NOE relaxation matrix, in which the diagonal R_1 terms and "leakage" terms are removed. Rate matrices of this type distribute initial perturbations of the Boltzmann distribution throughout the system, without driving the system back to equilibrium. Below, we will refine the LCLO matrix by adding additional terms that do drive the system back to thermodynamic equilibrium.

$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\alpha^2 \\ N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\beta^1 H_\beta^2 \end{bmatrix} = -\bar{R}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\alpha^2 \\ N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\beta^1 H_\beta^2 \end{bmatrix} \tag{2.9}$$

where R is the relaxation matrix

(2.10)

$$\bar{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & K_{12}+K_{A2} & -K_{12} & 0 & -K_{A2} & 0 & 0 & 0 \\ 0 & -K_{12} & K_{12}+K_{A1} & 0 & -K_{A1} & 0 & 0 & 0 \\ 0 & 0 & 0 & K_{A1}+K_{A2} & 0 & -K_{A1} & -K_{A2} & 0 \\ 0 & -K_{A2} & -K_{A1} & 0 & K_{A1}+K_{A2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -K_{A1} & 0 & K_{12}+K_{A1} & -K_{12} & 0 \\ 0 & 0 & 0 & -K_{A2} & 0 & -K_{12} & K_{12}+K_{A2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Terms equal to the negative sum of the off-diagonal cross-relaxation terms have been added to the diagonal (auto-relaxation) rate entry of the relaxation matrix for each LCLO substate in order to preserve detailed balance.

In order to clearly delineate the difference between true auto-relaxation terms and diagonal terms which arise as a result of detailed balance of cross-relaxation, the detailed

balance terms will be added during consideration of the cross-relaxation elements of the matrix. Although they are located on the diagonal of the matrix, they are not true auto-relaxation processes.

As can be seen from the abbreviated form of the matrix, LCLO substates may be conveniently represented by binary numbers, with 0s and 1s replacing a's and b's. Bitwise logic operations may be used to explore the operational relaxation pathways between LCLO substates. For instance, if the bitwise XOR of the binary representations of two LCLO substates contains two and only two 1's, a 0 in the first state has changed into a 1 in the second state (a to b) while a 1 in the first state has changed into a 0 in the second state (b into a). This indicates the presence of a zero-quantum spin flip relaxation pathway between these two states. Shortcuts such as these are convenient in the computational treatment of LCLO relaxation.

Examination of this relaxation matrix (Eq. (2.10)) leads to two observations. First, when only the zero-quantum terms are considered, the relaxation matrix may be blocked into submatrices representing isoenergetic levels in Figure 2.2 (labeled block 0, block 1, etc.). Since only zero-quantum terms are being considered, each block exchanges magnetization only within itself. Each block is characterized by a constant number of a and b spins, corresponding to the sum of the spin quantum number of all protons considered, m_J. Zero-quantum relaxation processes conserve the total number of a and b spins while rearranging their location within the proton network.

The energetic blocks in Eq. (2.10) are as follows:

Block 0: $$\frac{d}{dt} N_z H_\alpha^A H_\alpha^1 H_\alpha^2 = 0$$

Block 1: $$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \end{bmatrix} = -\begin{bmatrix} K_{12}+K_{A1} & -K_{12} & -K_{A1} \\ -K_{12} & K_{12}+K_{A2} & -K_{A2} \\ -K_{A1} & -K_{A2} & K_{A1}+K_{A2} \end{bmatrix}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \end{bmatrix}$$

Block 2:

$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \end{bmatrix} = -\begin{bmatrix} K_{A1}+K_{A2} & -K_{A1} & -K_{A2} \\ -K_{A1} & K_{12}+K_{A1} & -K_{12} \\ -K_{A2} & -K_{12} & K_{12}+K_{A2} \end{bmatrix}\begin{bmatrix} N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \end{bmatrix}$$

Block 3:

$$\frac{d}{dt} N_z H_\beta^A H_\beta^1 H_\beta^2 = 0$$

Second, a system containing N neighbor protons contains 2^{N+1} substates in total, arranged among N+2 energy level blocks. Let us refer to these blocks as Block 0, Block 1, Block 2 ... Block N+1, as above. In our labeling system, block 0 contains the single substate with all protons in the a state, block 1 contains substates having one b proton, block 2 contains substates having two b protons, and so on, all the way up to block N+1 which contains the single substate having N+1 b protons. In general, block m_J contains $\binom{N+1}{m_J}$ substates, $\binom{N}{m_J}$ of which will be a substates and $\binom{N}{m_J-1}$ of which will be b substates.

Here we use the common definition $\binom{N}{m_J} \equiv \frac{N!}{(N-m_J)!m_J!}$.

We next consider several other terms that contribute to the cross-relaxation of the amide NH LCLO substate network. These terms are united in the fact that they all contain spectral density functions on the order of $J(\omega_H)$. These spectral density functions are at least two orders of magnitude smaller in magnitude than the zero-frequency spectral density functions J(0) that govern the zero-quantum spin flip transitions. For this reason, the non-J(0) terms that we will examine here should be considered second-order contributions to the overall relaxation behavior of the system.

Single-quantum homonuclear proton-proton dipole-dipole interactions (hereafter abbreviated SQHH interactions) cause LCLO substates to exchange a single a spin for a single b spin, or vice versa. For instance, NzHaH'bH"a... may transition to NzHbH'bH"a... or to NzHaH'aH"a... due to SQHH interactions. In general, SQHH interactions operating on the ith proton in the LCLO system connect states Nz...Hia... and

Nz...Hib.... We may sum the SQHH interactions for the state-switching proton i over every other proton j in the system, leading to a total rate:

$$SQHH_i = \sum_j SQHH_{ij} = \sum_j \frac{3}{8} d_{HiHj}^2 J_{HiHj}(\omega_H) \tag{2.11}$$

In order to preserve legibility, we label SQHH interactions with the single letter A when illustrating their placement in the Liouville relaxation matrix. SQHH interactions enter the Liouville-space relaxation matrix as:

(2.12)

$$\bar{R} = \begin{bmatrix}
\sum A & -A_2 & -A_1 & 0 & -A_A & 0 & 0 & 0 \\
-A_2 & K_{12} + K_{A2} + \sum A & -K_{12} & -A_1 & -K_{A2} & -A_N & 0 & 0 \\
-A_1 & -K_{12} & K_{12} + K_{A1} + \sum A & -A_2 & -K_{A1} & 0 & -A_A & 0 \\
0 & -A_1 & -A_2 & K_{A1} + K_{A2} + \sum A & 0 & -K_{A1} & -K_{A2} & -A_N \\
-A_N & -K_{A2} & -K_{A1} & 0 & K_{A1} + K_{A2} + \sum A & -A_2 & -A_1 & 0 \\
0 & -A_A & 0 & -K_{A1} & -A_2 & K_{12} + K_{A1} + \sum A & -K_{12} & -A_1 \\
0 & 0 & -A_A & -K_{A2} & -A_1 & -K_{12} & K_{12} + K_{A2} + \sum A & -A_2 \\
0 & 0 & 0 & -A_A & 0 & -A_1 & -A_2 & \sum A
\end{bmatrix}$$

wherein $\sum A = \sum_i A_i$.

Heteronuclear NOE contributions to the LCLO relaxation matrix have already been mentioned. Heteronuclear NOE processes connect LCLO substates along the same pathways as SQHH interactions, allowing cross-relaxation between single a spins and single b spins, or vice versa. For instance, NzHaH'bH"a... may transition to NzHbH'bH"a... or to NzHaH'aH"a... due to heteronuclear NOE interactions.

However, there is a complication caused by the mixed-basis representation presented here, in which heteronuclear NOE processes are operating between heteroatom spins in the z state and proton spins in the a/b state. This causes the familiar double-quantum/zero-quantum relaxation processes produced by the heteronuclear NOE to be split across cross- and auto-relaxation entries in the Liouville-space relaxation matrix.

For each proton in the relaxation network, we may sum all NOE interactions with heteronuclei over the entirety of the protein. This sum yields the total rate at which NOE-induced spin transitions take place for each individual proton in the network.

The structure of heteronuclear NOE-type interactions in the Liouville relaxation matrix consists of a split of equal terms (here abbreviated B) between the auto and cross relaxation components of a given single proton a-b state pair. For instance: the state NzHaH'aH"b will have an NOE term B with which it relaxes to itself (autorelaxation along the diagonal into NzHaH'aH"b) and another NOE term B with which it enters each other state available by a single-quantum proton transition. These terms B both have the same sign.

The heteronuclear NOE term B_i affecting the state-switching behavior of the ith proton is the sum over all j of the ith state-switching proton's heteronuclear NOE interaction with the jth heteronucleus (atom type X):

$$B_i = \sum_j B_{ij} = \sum_j \frac{1}{8} d_{HiXj}^2 \left(J_{HiXj}\left(\omega_H - \omega_X\right) - 3J_{HiXj}\left(\omega_H\right) + 6J_{HiXj}\left(\omega_H + \omega_X\right)\right) \quad (2.13)$$

When these summed heteronuclear NOE interactions are added to the Liouville relaxation matrix for the LCLO substate system, the relaxation matrix becomes:

$$\bar{R} = \begin{bmatrix}
\sum A + \sum B & -A_2 + B_2 & -A_1 + B_1 & 0 & -A_A + B_A & 0 & 0 & 0 \\
-A_2 + B_2 & K_{12} + K_{A2} + \sum A + \sum B & -K_{12} & -A_1 + B_1 & -K_{A2} & -A_A + B_A & 0 & 0 \\
-A_1 + B_1 & -K_{12} & K_{12} + K_{A1} + \sum A + \sum B & -A_2 + B_2 & -K_{A1} & 0 & -A_A + B_A & 0 \\
0 & -A_1 + B_1 & -A_2 + B_2 & K_{A1} + K_{A2} + \sum A + \sum B & 0 & -K_{A1} & -K_{A2} & -A_A + B_A \\
-A_A + B_A & -K_{A2} & -K_{A1} & 0 & K_{A1} + K_{A2} + \sum A + \sum B & -A_2 + B_2 & -A_1 + B_1 & 0 \\
0 & -A_A + B_A & 0 & -K_{A1} & -A_2 + B_2 & K_{12} + K_{A1} + \sum A + \sum B & -K_{12} & -A_1 + B_1 \\
0 & 0 & -A_A + B_A & -K_{A2} & -A_1 + B_1 & -K_{12} & K_{12} + K_{A2} + \sum A + \sum B & -A_2 + B_2 \\
0 & 0 & 0 & -A_A + B_A & 0 & -A_1 + B_1 & -A_2 + B_2 & \sum A + \sum B
\end{bmatrix} \quad (2.14)$$

where $\sum B = \sum_i B_i$.

Single-quantum proton CSA relaxation operates along the same relaxation pathways as the previous two relaxation mechanisms, allowing cross-relaxation between

single a spins and single b spins. It depends on the strength of the proton CSA, however, a quantity much less rigorously quantified than the dipolar interaction distances on which previous relaxation mechanisms have been based. A CSA of 10 ppm for the amide proton is consistent with orders of magnitude reported in the literature and will serve to illustrate the size of the interaction.

Furthermore, since proton CSA relaxation is based on a monic and not a dyadic interaction, the size of the proton CSA relaxation rate does not scale with the number of protons in the system. This causes proton CSA relaxation to be a particularly minor contributor to LCLO relaxation; in general, relaxation rates that do not scale with the number of other atoms in the protein should be considered third-order in the LCLO system. For the ith proton in the LCLO network, the proton CSA relaxation (here abbreviated C_i)

$$C_i = \frac{1}{6} c_{Hi}^2 J(\omega_H) \tag{2.15}$$

is added to the Liouville relaxation matrix as follows:

(2.16)

$$\bar{R} = \begin{bmatrix}
\cdots & -A_2+B_2-C_2 & -A_1+B_1-C_1 & 0 & -A_N+B_N-C_N & 0 & 0 & 0 \\
-A_2+B_2-C_2 & \cdots & -K_{12} & -A_1+B_1-C_1 & -K_{N2} & -A_N+B_N-C_N & 0 & 0 \\
-A_1+B_1-C_1 & -K_{12} & \cdots & -A_2+B_2-C_2 & -K_{N1} & 0 & -A_N+B_N-C_N & 0 \\
0 & -A_1+B_1-C_1 & -A_2+B_2-C_2 & \cdots & 0 & -K_{N1} & -K_{N2} & -A_N+B_N-C_N \\
-A_N+B_N-C_N & -K_{N2} & -K_{N1} & 0 & \cdots & -A_2+B_2-C_2 & -A_1+B_1-C_1 & 0 \\
0 & -A_N+B_N-C_N & 0 & -K_{N1} & -A_2+B_2-C_2 & \cdots & -K_{12} & -A_1+B_1-C_1 \\
0 & 0 & -A_N+B_N-C_N & -K_{N2} & -A_1+B_1-C_1 & -K_{12} & \cdots & -A_2+B_2-C_2 \\
0 & 0 & 0 & -A_N+B_N-C_N & 0 & -A_1+B_1-C_1 & -A_2+B_2-C_2 & \cdots
\end{bmatrix}$$

The next cross-relaxation pathway we will take note of, is, unlike the previous pathways, in that its characteristic spectral density function is $J(2\omega_H)$. Double-quantum proton-proton dipolar interactions cause two protons in an LCLO substate to change at once from either an a state to a b state, or vice versa. For example, NzHaH'bH"a... may transition to NzHbH'bH"b... due to a double-quantum proton-proton dipolar interaction.

For LCLO network protons i and j, double-quantum cross-relaxation, which we will label D_{ij}, will take place between states Nz...Hia...Hja... and Nz...Hib...Hjb.... at the rate:

$$D_{ij} = \frac{3}{2} d_{HiHj}^2 J_{HiHj}(2\omega_H) \tag{2.17}$$

These double-quantum interactions modify the Liouville relaxation matrix in the following manner:

$$\bar{R} = \begin{bmatrix}
\cdots & -A_2+B_2-C_2 & -A_1+B_1-C_1 & -D_{12} & -A_N+B_N-C_N & -D_{N2} & -D_{N1} & 0 \\
-A_2+B_2-C_2 & \cdots & -K_{12} & -A_1+B_1-C_1 & -K_{N2} & -A_N+B_N-C_N & 0 & -D_{N1} \\
-A_1+B_1-C_1 & -K_{12} & \cdots & -A_2+B_2-C_2 & -K_{N1} & 0 & -A_N+B_N-C_N & -D_{N2} \\
-D_{12} & -A_1+B_1-C_1 & -A_2+B_2-C_2 & \cdots & 0 & -K_{N1} & -K_{N2} & -A_N+B_N-C_N \\
-A_N+B_N-C_N & -K_{N2} & -K_{N1} & 0 & \cdots & -A_2+B_2-C_2 & -A_1+B_1-C_1 & -D_{12} \\
-D_{N2} & -A_N+B_N-C_N & 0 & -K_{N1} & -A_2+B_2-C_2 & \cdots & -K_{12} & -A_1+B_1-C_1 \\
-D_{N1} & 0 & -A_N+B_N-C_N & -K_{N2} & -A_1+B_1-C_1 & -K_{12} & \cdots & -A_2+B_2-C_2 \\
0 & -D_{N1} & -D_{N2} & -A_N+B_N-C_N & -D_{12} & -A_1+B_1-C_1 & -A_2+B_2-C_2 & \cdots
\end{bmatrix} \tag{2.18}$$

2.4 Cross-correlated cross-relaxation

Cross-correlated proton-proton cross-relaxation mechanisms also exist in the LCLO system. These mechanisms take two forms. The first consists of cross-correlation between proton-proton dipolar relaxation and proton CSA relaxation, while the second term consists of cross-correlation between one proton-proton dipolar interaction and another proton-proton dipolar interaction.

Other dipolar/CSA terms consist of cross-correlation between heteroatom-proton dipolar relaxation and proton CSA relaxation, as well as cross-correlation between one heteroatom-proton dipolar interaction and another nitrogen-proton dipolar interaction. Spectroscopic means (180° pulses during relaxation) are used to average, and consequently eliminate, the effects of these heteroatom cross-correlated cross-relaxation mechanisms on relaxation during the η_zK experiment.

Although the mechanisms vary, both active types of cross-correlated cross-relaxation operate along the same pathway as single-quantum cross relaxation. They cause LCLO substates to switch a single a spin for a single b spin, or vice versa. The

sign of each cross-correlated cross-relaxation term changes depending on the spin states of the other proton states coupled by that term.

The rules for these sign changes are different for the dipolar/CSA and dipolar/dipolar terms. First, identify the proton whose state is being switched. In the case of a switch from NzHaH'bH"a... to NzHaH'aH"a..., this is the H' proton.

The proton-proton dipolar/proton CSA terms will have this proton involved as one member of the dipolar pair, and also as the CSA contributor. Next, consider the other member of the dipolar pair. If the other member of the dipolar pair is in the a state during the transition, the sign will be negative; if the other member of the dipolar pair is in the b state during the transition, the sign will be positive. The contribution from the dipolar pair with state-switching proton i and dipolar partner proton j will therefore be:

$$XC_{HH/H}^{DD/CSA} = \frac{1}{2} c_{Hi} d_{HiHj} J^{X}_{HiHj/Hi}(\omega_H) \quad \text{(dipolar partner in a state)} \tag{2.19}$$

$$XC_{HH/H}^{DD/CSA} = -\frac{1}{2} c_{Hi} d_{HiHj} J^{X}_{HiHj/Hi}(\omega_H) \quad \text{(dipolar partner in b state)} \tag{2.20}$$

The proton-proton dipolar/proton-proton dipolar cross-correlated cross-relaxation terms follow a different rule. The proton whose state is being switched, will be involved in both dipolar interactions. Again, we label this proton i. Two other protons, j and k, are the dipolar partners of proton i, one in each of the dipolar interactions. If the states of proton j and k are the same, the sign of the dipolar/dipolar cross-correlated cross-relaxation term will be negative; if the states of proton j and k are different, the sign of the term will be positive. The contribution from the dipolar/dipolar terms with state-switching proton i and dipolar partner protons j and k will therefore be:

$$XC_{HiHj/HiHk}^{DD/DD} = \frac{3}{4} d_{HiHj} d_{HiHk} J^{X}_{HiHj/HiHk}(\omega_H) \quad \text{(j and k same state (a/a, b/b))} \tag{2.21}$$

$$XC_{HiHj/HiHk}^{DD/DD} = -\frac{3}{4} d_{HiHj} d_{HiHk} J^{X}_{HiHj/HiHk}(\omega_H) \quad \text{(j and k different state (a/b, b/a))} \tag{2.22}$$

2.5 Auto-relaxation and cross-correlated contributions to auto-relaxation

The auto-relaxation behavior of the LCLO substates consists of a common longitudinal relaxation rate shared between all LCLO substates, with LCLO substate-specific modifications based on cross-correlated and auto-relaxation processes. We will first describe the common longitudinal relaxation rate and then move through the various cross-correlation effects on auto-relaxation within the LCLO system.

The auto-relaxation rate common to all LCLO substates is:

$$R = \frac{1}{3}c_N^2 J_{NCSA}(\omega_N) + \sum_i DDR_i =$$

$$\frac{1}{3}c_N^2 J_{NCSA}(\omega_N) + \sum_i \frac{1}{4}d_{NHi}^2 \left(J_{NHi}(\omega_N - \omega_H) + 3J_{NHi}(\omega_N) + 6J_{NHi}(\omega_N - \omega_H) \right) \quad (2.23)$$

where N is the amide nitrogen involved in the NH spin system being probed.

Several cross-correlated auto-relaxation terms exist in the LCLO system. These terms change sign depending on the a/b substate of the protons within each substate, causing each LCLO substate to display varying auto- as well as cross-relaxation behavior. Their origin lies in cross-correlation between nitrogen-proton dipolar and nitrogen CSA relaxation, or between the nitrogen-proton dipolar relaxation mechanisms of two different protons.

The ith proton in an LCLO substate will contribute a nitrogen-proton dipolar/nitrogen CSA cross-correlated auto-relaxation rate of the form:

$$\eta_{zi} = c_N d_{NHi} J^X_{NHi/N}(\omega_N) \quad \text{(proton state a)} \quad (2.24)$$

$$\eta_{zi} = -c_N d_{NHi} J^X_{NHi/N}(\omega_N) \quad \text{(proton state b)} \quad (2.25)$$

where the sign of the contribution depends on the a or b state of the ith proton.

The spectroscopic combination of the LCLO major states ensures that only the amide proton H_N's DD/CSA cross-correlated auto-relaxation contribution will be observed in the final analysis of the experiment.

Meanwhile, the ith and jth proton in an LCLO substate will contribute a nitrogen-proton dipolar/nitrogen-proton dipolar cross-correlated auto-relaxation rate of the form:

$$DDDD_{ij} = \frac{3}{2} d_{NHi} d_{NHj} J^{X}_{NHi/NHj}(\omega_N) \quad \text{(i,j states same (a,a or b,b))} \tag{2.26}$$

$$DDDD_{ij} = -\frac{3}{2} d_{NHi} d_{NHj} J^{X}_{NHi/NHj}(\omega_N) \quad \text{(i,j states different (a,b or b,a))} \tag{2.27}$$

Numerical values for all different longitudinal relaxation rates for the spin system shown in Figures 2.1 and 2.2 were computed for a ^{15}N-^{1}H amide spin system with $B_0 = 11.47$ T ($\omega_H = 500$ MHz) , $t_c = 7$ ns, $S^2 = 1$. These values are presented in Table 2.1.

Term	J (w) term	Rate (s^{-1})
K_{A1} (ZQ A-1)	J (0)	0.779
K_{12} (ZQ 1-2)	J (0)	0.097
^{15}N CSA	J (N)	0.3568
^{15}N- (^{1}H) Dip	J (N) J (N+H) J (N-H)	2.07
η_z	J (N)	-1.71
^{1}H-^{15}N NOE	J (N+H) J (N-H)	0.007
SQ A (1a, 2a)*	J (H)	0.0034
SQ A (1a, 2b)*	J (H)	0.0082
SQ 1 (Aa, 2a)*	J (H)	0.0031
SQ 1 (Aa, 2b) *	JH, JH2	0.0023
DQ A-1	J (2H)	0.0024
DQ 1-2	J (2H)	0.0003
^{1}H-amide CSA	J (H)	0.002
ζ	J (H) J (N+H) J (N-H)	0.003

Table 2.1. Longitudinal relaxation rates for a ^{15}N-^{1}H amide spin system with two neighbor protons (Figure 2.1).
B_0 = 11.47 T (ω_H = 500 MHz ^{1}H), t_c=7 ns, S^2=1. A=$^{1}H_N$, 1=$^{1}H_1$, 2=$^{1}H_2$ as in the relaxation matrices. Because the network is symmetric, some values for $^{1}H^2$ are not listed. ^{15}N CSA is 150 ppm, $^{1}H_N$ CSA is 10 ppm. Symbols in parentheses indicate the spin states of the other two protons, which have an effect on dipole-dipole / dipole-dipole cross-correlated relaxation. SQ A (1b, 2b) = SQ A (1a, 2a); SQ A (1a, 2b)= SQ A (1b, 2a); etc.

It is clear that, individually, the rates governed by spectral densities at frequencies higher than ω_N are all very small compared to the J(0)-driven spin flips and J(ω_N)-driven relaxation of the diagonal terms. For larger proteins, K rates will dominate the spectral density terms. At higher magnetic field, all J(ω) terms other than J(0) will be even smaller, making the K domination even larger. In the results section, we will show that the cumulative effect of the rates driven by J(ω) terms other than J(0) can be neglected in practice, even when working with calmodulin on a 500 MHz spectrometer. Those interested in further investigations into the theory of relaxation in AX_n spin systems should consult the classic review of Werbelow and Grant.[78]

2.6 Approximating the LCLO system.

The detailed LCLO system contains ten different relaxation mechanisms operating over literally thousands of proton-proton and heteroatom-proton interactions, each with its own set of spectral density functions. These interactions cannot be practically separated and individually characterized with the current state of the art of NMR spectroscopy.

We can, however, approximate the detailed LCLO system by a simpler system that retains the important features of the LCLO system while being accessible to experiment. We will develop this approximation in two stages.

The first stage of the approximation is to neglect the operation of non-zero-quantum terms in the cross-relaxation elements of the matrix. Relevant auto-relaxation terms (the common LCLO auto-relaxation rate, and the dipole-dipole/CSA cross-correlated relaxation rate) are retained along the diagonal. This is equivalent to returning to Equation (2.10) in the detailed description of the relaxation matrix given previously.

The qualitative basis for this approximation is that, in the large-molecule limit for which the experiment is designed, zero-quantum processes dominate the relaxation behavior of the LCLO system. Under conditions of relative molecular rigidity, the total non-zero-quantum cross-relaxation contribution from a given proton-proton interaction is 1-2% of the total zero-quantum relaxation contribution. This is due to the relative sizes of the spectral density function evaluated at zero frequency and at the proton Larmor frequency ω_H; under semi-rigid molecular conditions, $J(0)$ is roughly 100x larger than $J(\omega_H)$ at 500 MHz. Non-zero-quantum processes are consequently lost in the experimental error of measuring the zero-quantum processes.

In the following chapter, we will return to this approximation and make a quantitative appraisal of its validity based on actual protein structures. For now, we will assume its validity and discuss its consequences.

As mentioned earlier, retaining only the zero-quantum terms in the LCLO relaxation matrix allows the matrix to be split into submatrices describing zero-quantum relaxation within a single energetic block. For the relaxation matrix these submatrices are, as illustrated previously:

Block 0: $$\frac{d}{dt}N_z H_\alpha^A H_\alpha^1 H_\alpha^2 = -(R+\eta_z)N_z H_\alpha^A H_\alpha^1 H_\alpha^2$$

Block 1:

$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \end{bmatrix} = -\begin{bmatrix} R+\eta_z+K_{12}+K_{A1} & -K_{12} & -K_{A1} \\ -K_{12} & R+\eta_z+K_{12}+K_{A2} & -K_{A2} \\ -K_{A1} & -K_{A2} & R-\eta_z+K_{A1}+K_{A2} \end{bmatrix}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \end{bmatrix}$$

Block 2:

$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \end{bmatrix} = -\begin{bmatrix} R+\eta_z+K_{A1}+K_{A2} & -K_{A1} & -K_{A2} \\ -K_{A1} & R-\eta_z+K_{12}+K_{A1} & -K_{12} \\ -K_{A2} & -K_{12} & R-\eta_z+K_{12}+K_{A2} \end{bmatrix}\begin{bmatrix} N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \end{bmatrix}$$

Block 3: $$\frac{d}{dt}N_z H_\beta^A H_\beta^1 H_\beta^2 = -(R-\eta_z)N_z H_\beta^A H_\beta^1 H_\beta^2$$

The second step of the approximation is the replacement of the zero-quantum dipolar relaxation rates K_{Aj} (j=1...N) between the amide proton and the N neighbor protons involved in the LCLO system by the mean rate across the N neighbor protons, $K = \sum_j K_{Aj}/N$. This is equivalent to replacing the real proton relaxation network with an approximate proton relaxation network in which all protons are equidistant from the amide proton. This equidistance is set such that the sum of zero-quantum relaxation rates between the amide proton and the virtual equidistant protons in the approximation is equal to the sum of zero-quantum relaxation rates between the amide protons and all true neighbor protons in the molecule.

We justify this approximation by making a brief detour into the conceptual similarities between the approximate LCLO system and the study of chemical dynamics.

In chemical dynamics, chemical reaction processes are examined for the effects of the energetic substate distributions of reactants and products. For instance, one vibrational energy state of a molecule may be much more likely to undergo a given reaction process than other vibrational energy states of the same molecule.

Consider a chemical system with an initial state i and several possible final states f. The initial state converts to each final state at a rate k_{if}. The chemical dynamics expression $K = \sum_j K_{Aj} / N$ for the total rate of decrease of the initial state is then simply:

$$k_i = \sum_f k_{if} \tag{2.28}$$

Now, consider an initial LCLO substate i which has several possible final states f. For instance, NzHaH'bH"a... has among its possible final states NzHbH'aH"a... and NzHaH'aH"b.... The initial state converts to each final state at a rate K_{if}. The expression for the total rate of decrease of the initial LCLO substate is then:

$$K_i = \sum_f K_{if} \tag{2.29}$$

Although we cannot observe the individual K_{if} rates (the rate at which individual neighbor protons generate spin flips), we can observe their sum, the K_i rate (the rate at which all neighbor protons together generate spin flips). This suggests that we can make an approximation to the real system (which has varying, unknowable K_{if} rates) by constructing a virtual system with the same total K_i rate, distributed equally across N equidistant virtual neighbor protons. Each of these virtual neighbor protons has the same K_{if} rate. The idea is that splitting the relaxation contributions equally among equidistant virtual neighbor protons allows the virtual system to approximate the equilibration behavior of a real system and preserve the same cross-relaxation rate K_i while sidestepping the observationally intractable problem of dissecting K_i into its component K_{if}s. We test this approximation in the results section. For now, again, we will take it as given and move along with exploration of its consequences.

Under the equidistant virtual neighbor proton approximation, the interproton relaxation rates that involve zero-quantum spin flips between LCLO substates sharing the same LCLO major state, are removed from the system. (An example would be the pathway connecting NzHaH'bH"a and NzHaH'aH"b, labeled K_{12} in the relaxation matrix equations.) Because we can only observe changes in LCLO major state (i.e. changes in the amide proton spin state), we may neglect substate exchanges that do not cause a change of major state.

These simplifications drastically lower the number of parameters required to describe the relaxation behavior of the system. The system can now be described in terms of the sum of zero-quantum relaxation rates *K* and the number of virtual approximation protons N. The zero-quantum relaxation submatrix blocks of the simplified system become:

Block 0:
$$\frac{d}{dt} N_z H_\alpha^A H_\alpha^1 H_\alpha^2 = -\left(R + \eta_z\right) N_z H_\alpha^A H_\alpha^1 H_\alpha^2$$

Block 1:
$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \end{bmatrix} = -\begin{bmatrix} R+\eta_z+K & \cdots & -K \\ \cdots & R+\eta_z+K & -K \\ -K & -K & R-\eta_z+2K \end{bmatrix}\begin{bmatrix} N_z H_\alpha^A H_\alpha^1 H_\beta^2 \\ N_z H_\alpha^A H_\beta^1 H_\alpha^2 \\ N_z H_\beta^A H_\alpha^1 H_\alpha^2 \end{bmatrix}$$

Block 2:
$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \end{bmatrix} = -\begin{bmatrix} R+\eta_z+2K & -K & -K \\ -K & R-\eta_z+K & \cdots \\ -K & \cdots & R-\eta_z+K \end{bmatrix}\begin{bmatrix} N_z H_\alpha^A H_\beta^1 H_\beta^2 \\ N_z H_\beta^A H_\alpha^1 H_\beta^2 \\ N_z H_\beta^A H_\beta^1 H_\alpha^2 \end{bmatrix}$$

Block 3:
$$\frac{d}{dt} N_z H_\beta^A H_\beta^1 H_\beta^2 = -\left(R - \eta_z\right) N_z H_\beta^A H_\beta^1 H_\beta^2$$

This representation gives only two kinds of rows in each submatrix, one corresponding to substates belonging to the NzHa major state and the other corresponding to substates belonging to the NzHb major state. For the m_Jth energetic block, each NzHa-type substate passes magnetization to m_J NzHb-type substates, with all magnetization leaving at an individual pathway rate K. Each NzHb-type substate passes magnetization to N+1-m_J NzHa-type substates, with all magnetization entering and

leaving at a rate K. Effectively, the behavior of the individual substates has been homogenized within the major states..

One may now remove the individual substates from the energetic block picture and replace them with two single states which represent substates belonging to NzHa and NzHb major states. The NzHa-type substate passes magnetization to the NzHb-type substate at a rate (N+1- m_J)K, while the NzHb-type substate passes magnetization to the NzHa-type substate at a rate m_JK. The dynamics of the NzHa- and NzHb-type substate in this representation are identical to the dynamics of an individual NzHa- and NzHb-affiliated substate in the block picture above. This results in the replacement of every block, *no matter its size,* by a 2x2 representative matrix:

This representation gives only two kinds of rows in each submatrix, one corresponding to substates belonging to the NzHa major state and the other corresponding to substates belonging to the NzHb major state. For the m_Jth energetic block, each NzHa-type substate passes magnetization to m_J NzHb-type substates, with all magnetization leaving at an individual pathway rate K. Each NzHb-type substate passes magnetization to N+1-z NzHa-type substates, with all magnetization entering and leaving at a rate K. Effectively, the behavior of the individual substates has been homogenized within the major states.

One may now remove the individual substates from the energetic block picture and replace them with two single states which represent substates belonging to √2NzHa and √2NzHb major states. The NzHa-type substate passes magnetization to the NzHb-type substate at a rate (N+1-m_J)K, while the NzHb-type substate passes magnetization to the NzHa-type substate at a rate m_JK. The dynamics of the NzHa- and NzHb-type substate in this representation are identical to the dynamics of an individual NzHa- and NzHb-affiliated substate in the block picture above. This results in the replacement of every block, *no matter its size,* by a 2x2 representative matrix:

(2.30)

$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha \\ N_z H_\beta \end{bmatrix}_{N,m_J} = -\begin{bmatrix} R+\eta_z+m_J K & -m_J K \\ -\left(N+1-m_J\right)K & R-\eta_z+\left(N+1-m_J\right)K \end{bmatrix}\begin{bmatrix} N_z H_\alpha \\ N_z H_\beta \end{bmatrix}_{N,m_J}$$

The intra-NzHa and -NzHb relaxation pathways -- those which transfer magnetization from NzHa-affiliated substates to NzHa-affiliated substates, or from NzHb-affiliated substates to NzHb-affiliated substates -- have disappeared from this representation. These relaxation pathways do not produce 'visible' changes in magnetization state in the 2x2 representative matrix, and can be therefore be ignored. The only productive pathways in the representative energetic block matrix are those leading from NzHa-affiliated substates to NzHb-affiliated substates, and *vice versa*.

For a given set of initial conditions, the solution of the system of equations represented by this 2x2 matrix yields the time evolution of a representative NzHa- and NzHb-type substate within the m_Jth energetic block. This can be used to calculate the time evolution of the NzHa and NzHb major states within the m_Jth energetic block after properly weighting the representative substates to account for the number of substates within each energetic block.

As mentioned previously, block m_J contains $\binom{N+1}{m_J}$ substates, $\binom{N}{m_J}$ of which will be a substates and $\binom{N}{m_J-1}$ of which will be b substates. There are zero a substates in the case of $m_J=N+1$, and zero b substates in the case of $m_J=0$ (in these cases, the binomial formulation is nonsensical). Multiplying the NzHa and NzHb solutions of the system of differential equations given in Equation (2.30) by, respectively, $\binom{N}{m_J}$ and $\binom{N}{m_J-1}$ allows proper accounting for the LCLO substate multiplicity within each energetic block.

The time evolution of the complete approximate LCLO system can then be calculated by summing Equation (2.30) over all N+2 energetic blocks. This leaves us with the final system of differential equations for the LCLO approximation:

$$\frac{d}{dt}\begin{bmatrix} N_z H_\alpha \\ N_z H_\beta \end{bmatrix}_N = \sum_{m_J=0}^{N+1} \frac{d}{dt}\begin{bmatrix} N_z H_\alpha \\ N_z H_\beta \end{bmatrix}_{N,m_J}$$

$$= \sum_{m_J=0}^{N+1} -\begin{bmatrix} \binom{N}{m_J} & 0 \\ 0 & \binom{N}{m_J-1} \end{bmatrix} \begin{bmatrix} R+\eta_z+m_J K & -m_J K \\ -(N+1-m_J)K & R-\eta_z+(N+1-m_J)K \end{bmatrix} \begin{bmatrix} N_z H_\alpha \\ N_z H_\beta \end{bmatrix}_{N,m_J}$$

(2.31)

CHAPTER 3: PRACTICAL CONDUCT OF THE η_zK EXPERIMENT

3.1 NMR spectroscopy

The η_zK experiment consists of acquiring sets of four two-dimensional ^1H_N-^{15}N TROSY-type spectra[27; 79; 80; 81; 82] at several different time points. The η_zK experimental pulse sequence is illustrated in Figure 3.1 and is essentially an S^3E-filtered (spin state selective element[83]) clean TROSY[84] sequence with relaxation delays included.

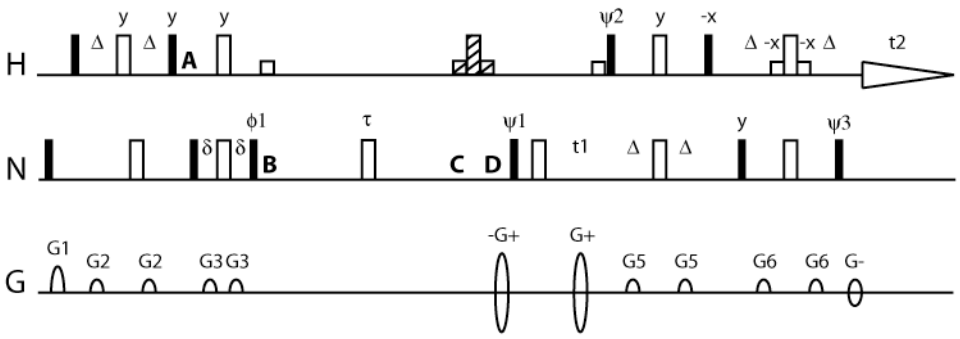

Figure 3.1: The η_zK pulse sequence.
Closed pulses indicate 90° pulses. Open pulses indicate 180° pulses. Short open pulses indicate selective square 90° pulses on water of ~800 ms duration. Delays are $\Delta = 1/4J_{NH}$, $\delta \approx 1/8J_{NH}$ (S^3E filter[29]) and τ = the relaxation time for a given experiment. δ is adjusted to minimize cross relaxation intensity for $\tau \approx 0$. Shaded pulses are issued during NzHb-observing experiments to place NzHb magnetization into the NzHa state for observation during TROSY chemical shift evolution, mixing, and readout. For Bruker instruments, $\varphi 1$ = [$\pi/8$ $\pi/8$ $5\pi/8$ $5\pi/8$] for I and II-type experiments and = [$3\pi/8$ $3\pi/8$ $7\pi/8$ $7\pi/8$] for III and IV-type experiments. $\psi 1$, $\psi 2$, $\psi 3$, and receiver phase are phase-alternated so as to accomplish axial peak suppression and Rance-Kay coherence selection. G+ and G- indicate coding and decoding gradients as implemented in TROSY-HSQC spectroscopy. For Varian instruments, phase cycling will need to be altered per Roerhl et al.[85]

The HSQC-based[86] approach of Bouguet-Bonnet et al.[87] provides a point of departure for this development of a completely 'TROSY-native' approach to preparing and measuring $\sqrt{2}N_zH_\alpha$ and $\sqrt{2}N_zH_\beta$ magnetization. Each set of η_zK spectra consists of four symmetric reconversion[39; 56; 88] sub-experiments: (I) $\sqrt{2}N_zH_\alpha$ prepare → $\sqrt{2}N_z H_\alpha$ detect; (II) $\sqrt{2}N_zH_\alpha$ prepare → $\sqrt{2}N_zH_\beta$ detect; (III) $\sqrt{2}N_zH_\beta$ prepare → $\sqrt{2}N_zH_\beta$ detect; (IV) $\sqrt{2}N_zH_\beta$ prepare → $\sqrt{2}N_zH_\alpha$ detect. The symmetric reconversion procedure allows the elimination of relaxation processes common to all magnetization pathways and allows simultaneous measurement of cross-relaxation and cross-correlated relaxation processes.

In brief, the η_zK experiment prepares the desired starting magnetization with an INEPT[89; 90] element followed by an S^3E element, and measures the result of relaxation processes operating on this magnetization with an appropriately manipulated HSQC-TROSY readout. In the following several paragraphs, I explain the progress of magnetization throughout the experiment and the choices of pulse elements used in experimental development.

The initial 90° pulse on nitrogen and subsequent gradient destroys equilibrium nitrogen magnetization recovering from the end of the previous iteration. This allows preparation of the desired complex magnetization states from pure proton magnetization. Longitudinal two-spin order is created by the INEPT element between A and B. The S^3E filter between B and C selects either $\sqrt{2}N_zH_\alpha$ or $\sqrt{2}N_zH_\beta$ to pass through to the relaxation period τ between C and D. The final pulse in this S^3E filter selects the type of magnetization desired (y, y, -y, -y for NzHa and x, x, -x, -x NzHb) and is phase cycled in concert with the receiver in order to mitigate the effects of asymmetric return to equilibrium from +z and –z magnetization during relaxation experiments. The length of the delay used in the S^3E element is optimized for minimal positive intensity in crossover experiments. Water magnetization is kept along the –y axis until C, at which point a selective 90° pulse on water returns water to +z during relaxation.

Cross-relaxation diffuses the initial pure state of $\sqrt{2}N_zH_\alpha$ or $\sqrt{2}N_zH_\beta$ into a mixture of $\sqrt{2}N_zH_\alpha$ and $\sqrt{2}N_zH_\beta$ magnetization. At the same time, auto-relaxation processes decrease the intensity of both $\sqrt{2}N_zH_\alpha$ and $\sqrt{2}N_zH_\beta$ magnetization. The composite 180°

pulse on nitrogen midway through the relaxation period averages the effects of cross-correlated relaxation processes operating differentially on Na-type and Nb-type magnetization, including NH dipolar/H CSA cross-correlated cross-relaxation and NH dipolar/NH dipolar cross-correlated cross-relaxation.

The relaxation period concludes at D with the issuance of an optional proton WATERGATE 180° pulse[91] depending on the component of magnetization desired for detection. Since TROSY is designed to accept the NxHa component of magnetization at the start of t_1, the WATERGATE 180° pulse is not issued if the desired symmetric reconversion subexperiment involves measuring $\sqrt{2}N_zH_\alpha$ magnetization. In this case, magnetization passes directly to the nitrogen 90° pulse beginning t_1, which is phase alternated between y and –y in concert with the receiver in order to achieve axial peak suppression. This pulse converts $\sqrt{2}N_zH_\alpha$ magnetization to NxHa magnetization, which is properly accepted by TROSY. The WATERGATE 180° pulse is issued if the desired symmetric reconversion subexperiment involves measuring $\sqrt{2}N_zH_\beta$ magnetization. Issuing this pulse causes the contents of $\sqrt{2}N_zH_\beta$ magnetization to be converted to $\sqrt{2}N_zH_a$ magnetization, and vice versa. The ex-NzHb magnetization, now in NzHa form, is immediately converted to NxHa magnetization by the nitrogen 90 pulse, and t_1 begins. A WATERGATE 180° pulse is used instead of a normal 180° pulse in order to ensure that water remains in +z, aiding water suppression efforts.

From E onwards the experiment proceeds as a standard ^{15}N-$^{1}H_N$ TROSY experiment. The 180° pulse on nitrogen at F is used to compensate time taken during t_1 for suppression of carbon scalar couplings and cross-correlated relaxation, and for the selective 90° pulse on water at G. This selective pulse is used to ensure that water enters the TROSY mixing and readout period aligned along –y in accordance with the design of TROSY.

The TROSY mixing and readout period between G and H consists of two S^3CT (spin-state-selective coherence transfer) elements bracketed by coding and decoding gradients. After chemical shift labeling of nitrogen coherence during t_1, these S^3CT elements and gradients selectively transfer either of the slowly relaxing nitrogen coherence quadruplet components N_+H_a or N_+H_a to the slowly relaxing proton quadruplet

component N_aH_+ for detection during t_2, depending on whether the pulse sequence is in antiecho or echo mode. Antiecho and echo transients are combined in order to yield a pure absorption spectrum with quadrature detection.[23; 92; 93; 94; 95; 96]

3.2 Data fitting and analysis

Experimental η_zK data is fit in several steps. Peak intensities for I, II, III and IV-type experiments are extracted using jitter mode in the Rate Analysis tool in NMRViewJ. These intensities are saved and imported into MATLAB, where post-spectral data processing begins.

The first step in fitting the data is extracting per-residue K rates, initial state purity information, and number of virtual equidistant protons N from a fit of the symmetric reconversion time evolution function X(t), where

$$(2.32)\ X(t) = \sqrt{\frac{II(t)\ III(t)}{I(t)\ IV(t)}}$$

and I(t), II(t), III(t) and IV(t) represent the four symmetric rweconversion sub-experiments listed above. The X(t) function for each residue is fit to a target function consisting of Equation (2.32) by Levenberg-Marquardt nonlinear least squares[97] while varying the number of equidistant protons N in a grid search. The range of N used in these studies varies from 1 to 60 (see Figure 2.2). The number of virtual neighbor protons needed for the fit is related to the number of actual protons significantly contributing to the relaxation behavior of the amide proton, but the dependence is not strict and is likely to vary with the degree of asymmetry in these protons' contribution to the total relaxation of the amide proton. For ^{15}N-labeled, protonated calmodulin, optimal N values vary between 5 and 11.

The dependence of the model on the number of equidistant virtual protons N places curve fitting to the model in the challenging realm of mixed-integer nonlinear programming (MILP) problems. These NP-hard problems involve optimization of a mix

of parameters taking integer values and continuous values. Worse yet, the several methods developed to efficiently attack MILP problems rely on the optimization technique of *relaxation* (no relation to NMR relaxation) in which parameters constrained to integer values are allowed to take on continuous values and the solution is nudged in the direction of integer values for these parameters. These techniques are inapplicable to the curve fitting problem encountered herein, since the choice of N used in the model governs inherently integral quantities such as the number of energy level blocks and the n, choose k functions used in accounting for LCLO substate degeneracy. In these circumstances, the grid search over N is the best and most robust technique available.

The fit also includes parameters governing the purity of initial magnetization state preparation (either NzHa or NzHb, typically >95%). The fit K rates, initial state purity parameters, and number of equidistant virtual protons N are then used to fit the symmetric reconversion time evolution function Y(t) for the value of the longitudinal NH/N DD/CSA cross-correlated relaxation rate η_z and the efficiency of observing NzHa or NzHb magnetization, where

$$(2.33)\, Y(t) = \frac{IV(t)}{I(t)}$$

The Y(t) function for each residue is fit to a target function consisting of Equation (2.33) by Levenberg-Marquardt nonlinear least squares, with fit K rates, initial state purity parameters, and number of equidistant virtual protons N held constant. This results in a total of five parameters for each X(t) and Y(t) symmetric reconversion fit.

NMR relaxation experiments are time-consuming. Physical repetition of relaxation experiments in order to construct a conventional statistical ensemble is impractical. Under these conditions, bootstrap statistical methods[98; 99] are used to construct confidence intervals for NMR relaxation parameters.

The first step in these methods consists of randomly selecting several *bootstrap sets* of data from the original data with replacement, such that the bootstrap sets of data contain as many total data points as the original 'true' data set, although data points will normally be repeated within the bootstrap data sets. In order to improve the quality of

the bootstrap ensemble and ensure complete coverage of the permutations of interpretation of existing data, as many bootstrap data sets should be generated as is computationally tractable. The work delineated in this thesis used 200 bootstrap data sets for each fitting problem.

The next step in bootstrap statistical methodology involves calculating the quantity of interest (e.g. curve fitting to some function by minimization of the χ^2 goodness of fit parameter) for each member of the bootstrap data ensemble and constructing an ensemble of the derived parameters. The result, in the case of the $\eta_z K$ experiment, consists of five distributions, one for each of the five fit parameters, based on bootstrap permutations of the available data. The statistics of these bootstrap distributions of fit parameters are then taken as estimators of the statistics of the underlying true distribution. The simplest approach, and the one used in this work, is to use the mean and standard deviation of the bootstrap distributions of fit parameters as estimated values and error bounds for these parameters. Alternative approaches including estimation of error bounds from percentiles of the distribution are also permissible, and likely more accurate in the case of severely asymmetric distributions.

Since the Y(t) fit parameters are dependent on the X(t) fit parameters, but the X(t) fit parameters are not dependent on the Y(t) fit parameters, the quality of the fit resulting from a given set of functional parameters is evaluated from the χ^2 goodness of fit parameter of the X(t) stage of fitting alone. The best Y(t) fit resulting from each best X(t) fit is appended to that X(t) fit to complete the full set of five best-fit parameters. Experience indicates that an X(t) fit resulting in good agreement with the X(t) data will always be able to lead to a Y(t) fit resulting in good agreement with the Y(t) data.

$\eta_z K$ fitting and bootstrap distribution construction code has been implemented in MATLAB and in FORTRAN90. The MATLAB code was used for algorithm development and provides a more convenient form of the code for use in casual glances at data. The FORTRAN90 code is written with speed in mind and incorporates functions from the MINPACK[100] minimization and LAPACK[101] linear algebra code libraries, which are in turn dependent on the fast scalar, vector and matrix operations implemented in the BLAS Basic Linear Algebra Subprograms[102; 103] routines. These routines are machine architecture-specific and should be available wherever fine supercomputing is

sold. Those interested in using the FORTRAN90 code will need to compile it in their local environment. The runtime for the FORTRAN90 code running on a single processor of the University of Michigan Nyx computing cluster in late September 2009 was 17 hours 9 minutes in the case of Hsc70 NBD data containing 319 residues with 17 data points each. The fit conditions for this trial consisted of grid searching over all values of N from 1-60 and constructing bootstrap ensembles containing 200 members.

3.3 Validation of the approximate LCLO system

In order to validate the equidistant proton network approximation, I generate synthetic relaxation data for the 2O5G X-ray structure of smMLCKp-bound Ca^{2+}-saturated calmodulin.

The structure is first protonated and checked for steric clashes and geometry using MolProbity.[104] Synthetic relaxation data is then generated for each amide proton under the influence of all relaxation interactions with its 8 nearest neighbor protons, as detailed in Chapter 2. All zero-quantum, single-quantum and double-quantum terms are included in the simulations. The governing conditions of these relaxation interactions consist of the experimentally determined conditions of axially symmetric diffusion for CaM/smMLCKp. All spectral density functions are set to an order parameter of 0.8 with a local correlation time of 1 ps.

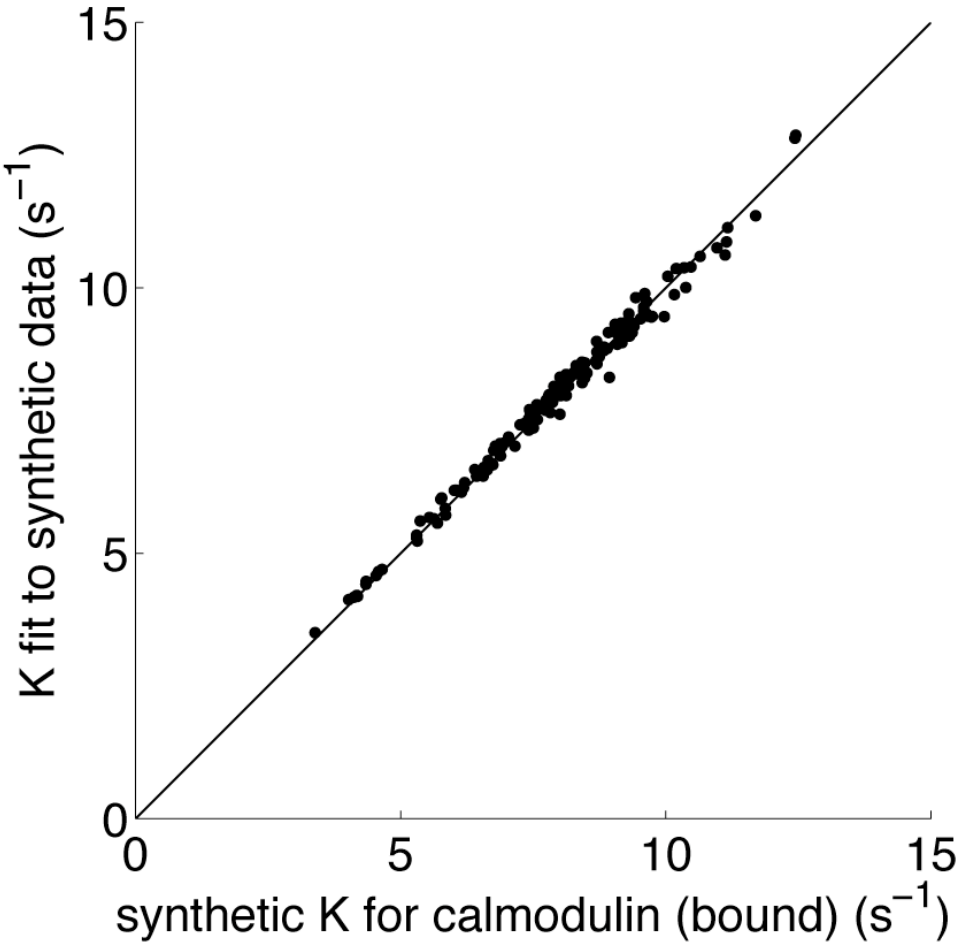

Figure 3.2. Correspondence between detailed simulation of cross-relaxation processes and equidistant virtual neighbor proton approximation-based K fitting to these processes.
Synthetic relaxation data (NzHa (t=0) → NzHa (t), NzHa (t=0) → NzHb (t), NzHb (t=0) → NzHa (t) and NzHb (t=0) → NzHb (t)) is generated from the PDB file for all residues of CaM/smMLCKp, taking the 8 closest neighbors for each amide into account and using all possible relaxation mechanisms with all spectral density terms; using axially symmetric rotational diffusion as in Table 4.1; $S^2 = 0.8$; $t_e = 1$ ps; 500 MHz. These synthetic relaxation data were fitted with the equidistant proton approach, using terms with spectral densities J(0) only. The horizontal axis contains the *K*-rates derived from the crystal structure and used to generate the synthetic relaxation data. The vertical axis contains the approximate K-rates resulting from the equidistant virtual neighbor proton approximation fitting process described above. The solid line is y=x. The R^2 is 0.99.

I then fit this synthetic data using the equidistant virtual neighbor proton network approximation including only proton-proton spin-flip (*K*) rates. Figure 3.2 compares the *K*-rates resulting from this fit with the detailed *K*-rates resulting from the generation of

the synthetic dataset. The correspondence is excellent ($R^2 = 0.99$) over the range of *K*-rates between 3 and 13 s^{-1}. This is strong proof that the neglect of higher frequency spectral density terms and use of the equidistant virtual neighbor proton approximation is justified.

3.4 Calculation of theoretical K_{rig} rates

Equation (2.7) gives the zero-quantum spin flip transition rate between two protons as

$$K_{ij} = \frac{1}{4} d_{HiHj}^2 J_{HiHj}(0)$$

$K_{rig,ij}$ is the zero-quantum spin flip transition rate experienced between two protons *i* and *j* which are fixed with respect to each other on the ps-ns timescale and whose relaxation dominated to good approximation by their zero-quantum relaxation behavior. This equation is used to calculate K_{rig} for each amide proton by summing the influence of all neighbor protons *j* on each amide proton *i*. Each *i*-*j* interaction consists of two parts: the interproton distance d_{HiHj} and the zero-quantum spectral density $J_{HiHj}(0)$.

As previously mentioned, distance interactions d_{HiHj} between protons *i* and *j* are calculated by means of the equation

$$d_{HiHj} = (\mu_o / 4\pi) \gamma_H^2 \hbar r_{HiHj}^{-3}$$

where γ_H is the proton gyromagnetic moment, μ_0 is the permeability of free space, and r_{HiHj} is the distance between protons *i* and *j*. The interproton distance is extracted from macromolecular structure files in PDB format. If NMR structures are used, protons will already be present; if X-ray structures are used, computational protonation with MolProbity will be required before d_{HiHj} values can be calculated.

Interproton distances must be corrected for interactions between amide protons and sidechain methyl group protons due to rapid rotational averaging of the methyl proton positions about the methyl symmetry axis. Correction for rapid motional averaging was performed according to the method of Tropp[44] and Yip and Case[105] as detailed in Neuhaus and Williamson[43], in which the distance from the amide proton to each methyl proton is replaced by

$$(2.34)\; r_{Tropp} = \left\{ \frac{1}{18} \sum_{\mu=1}^{3} \sum_{\nu=1}^{3} \left[\frac{3\left(\vec{r}_{ij,\mu} \bullet \vec{r}_{ij,\nu}\right)^2 - r_{ij,\mu}^2 r_{ij,\nu}^2}{r_{ij,\mu}^5 r_{ij,\nu}^5} \right] \right\}^{-1/6}$$

The next step in the construction of K_{rig} is calculation of the spectral density function $J_{HiHj,rigid}(0)$. Since the spectral density function is expressly for the case of rigid molecular structure, all order parameters are set to 1, leading to a corresponding simplification of spectral densities. Depending on the model for molecular rotational diffusion in use, the spectral density function may take on more or less elaborate forms. The simplest form of the spectral density function, used in the simplest case of isotropic rotational diffusion, is the time-honored model-free spectral density function developed by Lipari and Szabo in 1982[11]:

$$(2.35)\; J(\omega) = \frac{2}{5} \frac{\tau_c}{1 + (\omega\tau_c)^2}$$

where τ_c is the isotropic overall molecular rotational diffusion time and ω is the characteristic Larmor frequency of the relaxation interaction involved. Since $\omega = 0$ for K rates, this simplest form of the spectral density function reduces to

$$(2.36)\; J(0) = \frac{2}{5}\tau_c$$

More involved spectral density functions include axially symmetric and fully asymmetric diffusion models[106; 107; 108]. The axially symmetric spectral density function is used throughout this work and is given by

$$(2.37)\ J(\omega) = \frac{2}{5}\sum_{j=1}^{3} A_{ji}\frac{\tau_j}{1+\left(\omega\tau_j\right)^2}$$

where $\tau_1^{-1} = 6D_\perp$, $\tau_2^{-1} = 5D_\perp + D_{\|}$, $\tau_3^{-1} = 2D_\perp + 4D_{\|}$, $A_{1i} = (3\cos^2\Theta_i - 1)^2/4$, $A_{2i} = 3\sin^2\Theta_i\cos^2\Theta_I)^2$, $A_{3i} = (3/4)\sin^4\Theta_i$, Θ_I is the angle between the NH bond vector of the *i*th backbone amide system and the principal axis of the diffusion tensor, and $D_\perp$ and $D_{\|}$ are the constants of diffusion about the minor and major axes of the ellipsoid respectively. Since $\omega = 0$ for K rates, the spectral density function for axially symmetric diffusion reduces to

$$(2.38)\ J(\omega) = \frac{2}{5}\sum_{j=1}^{3} A_{ji}\tau_j$$

η_z rates are combined with η_{xy} transverse ^{15}N-^{1}H/^{15}N CSA/DD cross-correlated relaxation rates in order to determine the macromolecular rotational diffusion tensor and consequently the diffusion constants entering the spectral density function J(0). The idea of this combination is derived from the work of Kroenke et al.[109]. η_{xy} rates are determined using TROSY-based techniques described by Wang et al.[110] Experimental site-specific rotational diffusion constants are calculated from

$$(2.39)\ D_i^{\exp} = \frac{\omega_N}{3}\left(6\frac{\eta_{xy}}{\eta_z} - 7\right)^{-1/2}$$

PDB orientation and diffusion tensor fits are carried out for these calculated D_i values using the programs PDBINERTIA and QUADRIC_DIFFUSION (AG Palmer III, Columbia University) and the relevant diffusion constants are extracted from the

resulting fits. In all cases studied herein, F-testing indicates that axially symmetric models of rotational diffusion provide a statistically significant improvement in fit over isotropic models of rotational diffusion. Fully asymmetric models of rotational diffusion do not provide a statistically significant improvement in fit over axially symmetric models of rotational diffusion.

3.5 Estimation of the precision of X-ray coordinates

Following Cruickshank[111], the precision of the position of an atom of 'typical' B-factor in an X-ray structure is (very) approximately

$$(2.40)\ \sigma\left(r, B_{avg}\right) = 3^{1/2}\left(N_i / p\right)^{1/2} C^{-1/3} R d_{\min}$$

where N_i is the total number of atoms in the structure, p is $n_{obs} - n_{params}$, n_{obs} is the number of reflections, n_{params} is the number of parameters (~$4n_{atoms}$ for isotropic refinement, ~$9n_{atoms}$ for fully anisotropic refinement), C is the completeness for range of the data used in refinement, R is the crystallographic R value, and d_{min} is the resolution of the structure. Assuming that the 'typical' B-factor can be used for computationally added protons, that any additional error introduced by MolProbity protonation is subsumed in Cruickshank's generous error estimation, and that distance errors between protons are uncorrelated, Eqn. (2.40) leads to a point-to-point distance error estimate for X-ray structures of

$$(2.41)\ \sigma\left(r, B_{avg}\right) = 6^{1/2}\left(N_i / p\right)^{1/2} C^{-1/3} R d_{\min}$$

based on propagating the error of the difference between two position estimates. Based on this distance error estimate, proton-proton distances may be assumed to be a Gaussian distribution with a mean established by the protonated X-ray structure and a standard deviation given by the point-to-point distance error given above.

CHAPTER 4: PROTON-PROTON DYNAMICS OF CA^{2+}-SATURATED CALMODULIN

4.1 Calmodulin introduction

Calmodulin (CaM) is a 148-residue soluble protein (16.7 kDa) constitutively expressed in nearly all eukaryotic cells and essential to cellular survival.[59] All vertebrates express the same calmodulin sequence, which provides an idea of the degree of conservation of this fundamental protein. CaM constitutes approximately 0.1% of the protein in all cells and is expressed at higher levels in rapidly growing cells.[112]

Calmodulin's role in the cell is that of a calcium sensor. Numerous proteins exhibiting calcium-sensitive functionality (nearly 300 reported as of 2004) bind CaM at CaM-binding domain motifs[58] upon cellular calcium uptake. There are five main types of CaM-binding domain, each marked by the presence of a handful of conserved hydrophobic residues. Upon an increase in intracellular calcium concentration from basal levels around 10^{-7} M to 10^{-6} M, CaM binds four Ca^{2+} ions and the CaM-binding domain, demonstrating strong positive cooperativity and high affinity for the CaM-binding domain ($K_d = 10^{-7}$-10^{-11} M)[112; 113]. Ca^{2+}-saturated CaM binding leads, in turn, to activation of the CaM-binding protein's calcium-sensitive functionality. This frequently occurs by a process in which autoinhibitory downregulation of protein activity (either steric or chemical) is removed by CaM binding and returned upon CaM release.

Free in solution, calmodulin consists of four EF-hand Ca^{2+}-binding motifs organized into globular N- and C-terminal domains connected by a flexible central linker. Each domain contains two EF-hand domains and folds in a similar fashion, with 46% sequence identity between domains. The Ca^{2+}-saturated form of calmodulin, which is our target for study, contains exclusively alpha-helical and loop structure. The C-terminal domain of CaM binds Ca^{2+} with three-to-five-fold higher affinity than the N-terminal

domain and exhibits significant positive cooperativity.[59; 112; 113] Several protein interactions with CaM depend on this partial Ca^{2+}-binding behavior as part of their regulatory process. These interactions include those with the CaM-binding domain under study in this chapter, that of smooth muscle myosin light chain kinase (smMLCK). These partial interactions are beyond the scope of this chapter; all CaM samples used in the course of this investigation exist under conditions of Ca^{2+}-saturation.

Calmodulin is extensively studied for several reasons. It plays a fundamental role in cellular physiology, being throughly involved in calcium sensing and cellular signalling. It is the most heavily studied of the calcium sensor proteins and provides a model for this important class of systems. It is a complex system capable of exhibiting a wide range of binding behaviors to different types of CaM-binding domains and over different ranges of calcium concentration.

Speaking strictly in terms of structural biology, Ca^{2+}-saturated CaM diffracts extremely well, is highly soluble, and gives excellent NMR spectra under a range of sample conditions. The quality of X-ray structures available for Ca^{2+}-saturated calmodulin is within the realm of fantasy for most other protein systems. The 1EXR structure of free Ca^{2+}-saturated calmodulin[64] diffracts to 1.0 Å resolution (although its rigid central helix is not seen in solution[114]), while the 2O5G structure of Ca^{2+}-saturated calmodulin bound to a peptide mimic of the smMLCK binding domain (smMLCKp) diffracts to 1.08 Å.[65] These structures are illustrated in Figures 4.1 and 4.2.

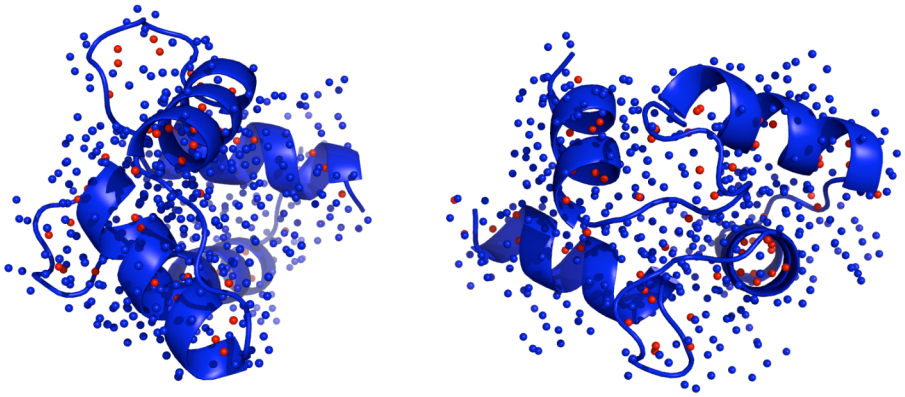

Figure 4.1. Ca^{2+}-saturated calmodulin free in solution
The structure portrayed is 1EXR.[64] Spheres indicate proton locations, with red spheres denoting amide protons and blue spheres denoting all other protons. Sphere scale is set to 1/4th the proton van der Waals radius for clarity.

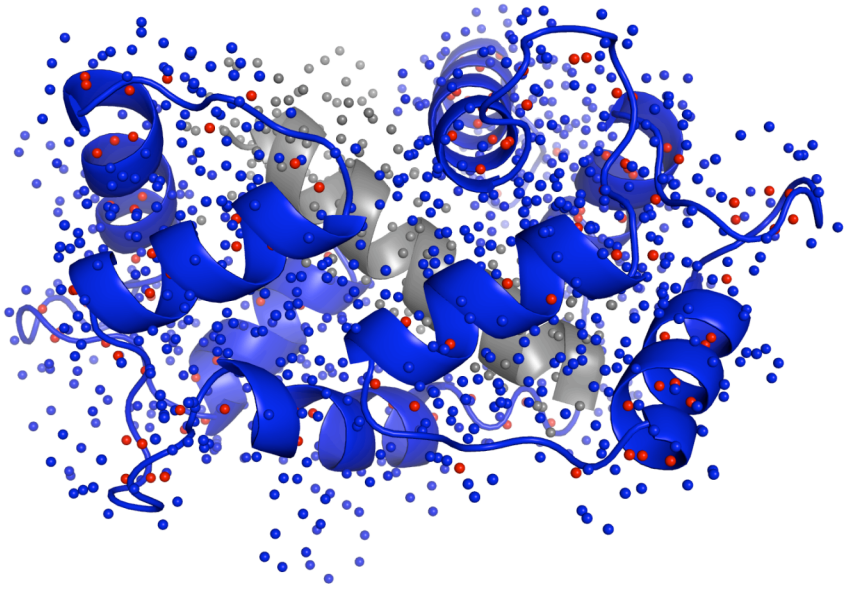

Figure 4.2. Ca^{2+}-saturated calmodulin bound to smooth muscle myosin light chain kinase CaM-binding domain peptide (smMLCKp)
The structure portrayed is 2O5G.[65] Spheres indicate proton locations, with red spheres denoting amide protons and blue spheres denoting all other protons. Sphere scale is set to 1/4th the proton van der Waals radius for clarity.

Most tantalizing of all, Ca^{2+}-saturated CaM appears to exhibit highly interesting dynamic changes throughout methyl groups in its sidechains during the process of binding smMLCKp[16; 19], pointing to a major role for protein dynamics in CaM binding to CaM-binding domains. This presents an opportunity to use proton-proton dynamics to establish a picture of dynamics changes that extend beyond the limited view provided by methyl probes. With this possibility in mind, we chose CaM Ca^{2+}-saturated CaM binding to smMLCKp as the first test of η_zK relaxation methodology in the study of protein dynamics.

4.2 Calmodulin materials and methods

We carried out η_zK relaxation experiments (Chapter 3) on a 1.0 mM sample of calcium-saturated chicken calmodulin (CaM) in 10% D_2O and on a 1.0 mM sample of calcium-saturated chicken calmodulin in complex with the calmodulin-binding domain of chicken gizzard smooth muscle myosin light chain kinase (CaM/smMLCKp), also in 10% D_2O. All experiments are conducted at 35 °C and 500 MHz on a Bruker Avance 500 MHz spectrometer equipped with a triple-resonance triple-axis gradient room-temperature probe.

η_zK relaxation experiments consist of 2D ^1H_N-^{15}N TROSY-HSQC spectra acquired at several different relaxation delays. All η_zK experiments are conducted with 2048 complex points in the proton dimension and 160 complex points in the nitrogen dimension, with the carrier set at 4.773 and 117.473 ppm respectively. The recycle delay is set to 1.0 s.

The experiment is conducted with 17 values of the relaxation delay τ: 20, 30, 40, 60, 60, 80, 90, 100, 120, 120, 140, 150, 160, 180, 180, 200, and 210 ms. 16 transients are recorded for each data point Experimental time was about 1 hour 50 min per spectrum leading to a total experiment time of 270 h over both samples. η_{xy} experiments[110] are performed at values of N = 1, 2, 3, 4 and 5 for the relaxation delay τ, according to the relation $2\tau = N/J^1_{NH}$.

4.3 Calmodulin results

Figures 4.3, 4.4, 4.5, and 4.6 illustrate representative CaM and CaM/smMLCKp spectra for I, II, III and IV-type experiments at 60 ms relaxation time. Signal intensities are quantified using peak heights as measured in NMRView's Rate Analysis tool.

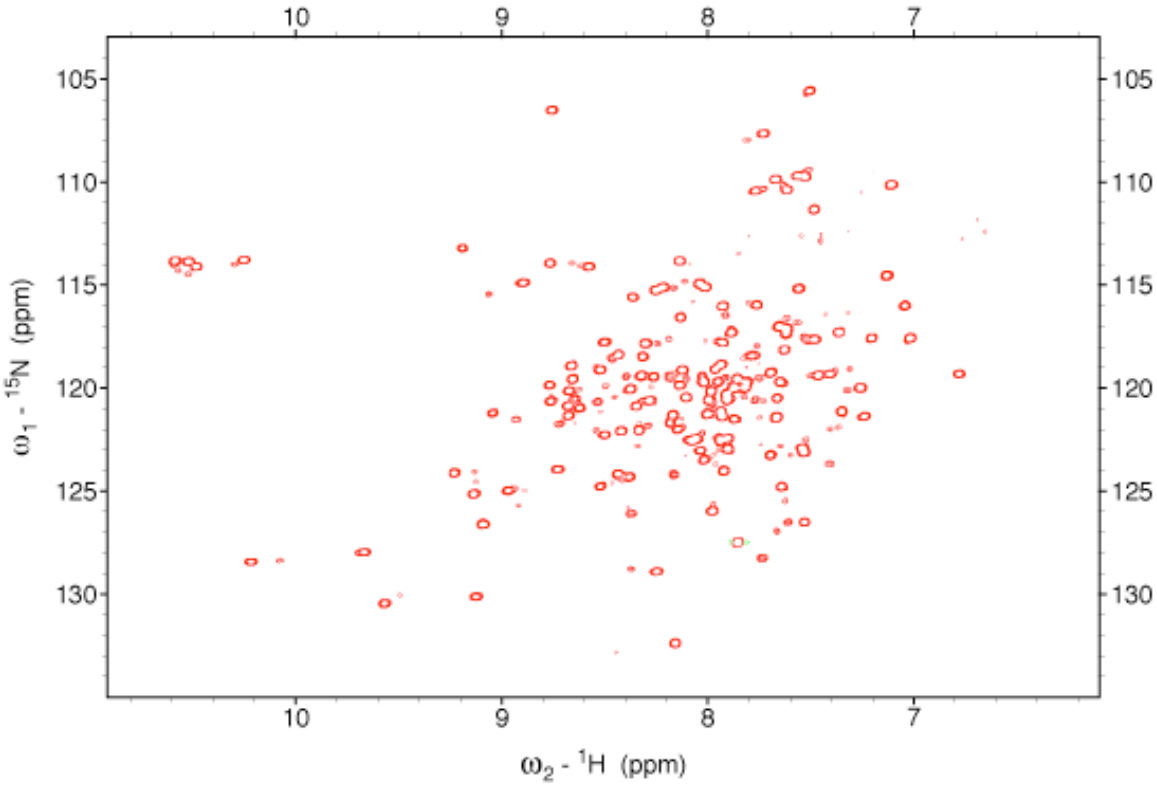

Figure 4.3. I-type (NzHa to NzHa) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin
The spectrum displayed is acquired at 35 C with a spectrometer field of 500 MHz and a relaxation time τ = 60 ms.

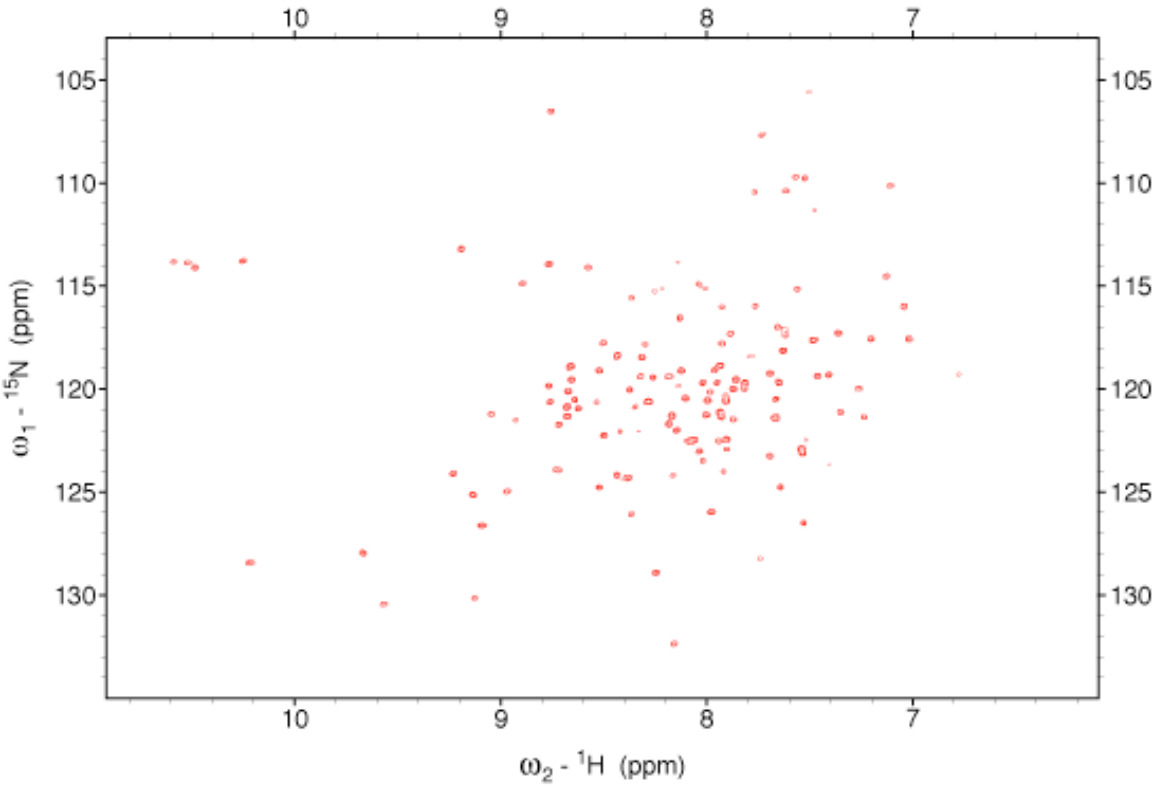

Figure 4.4. II-type (NzHa to NzHb) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin
The spectrum displayed is acquired at 35 C with a spectrometer field of 500 MHz and a relaxation time τ = 60 ms.

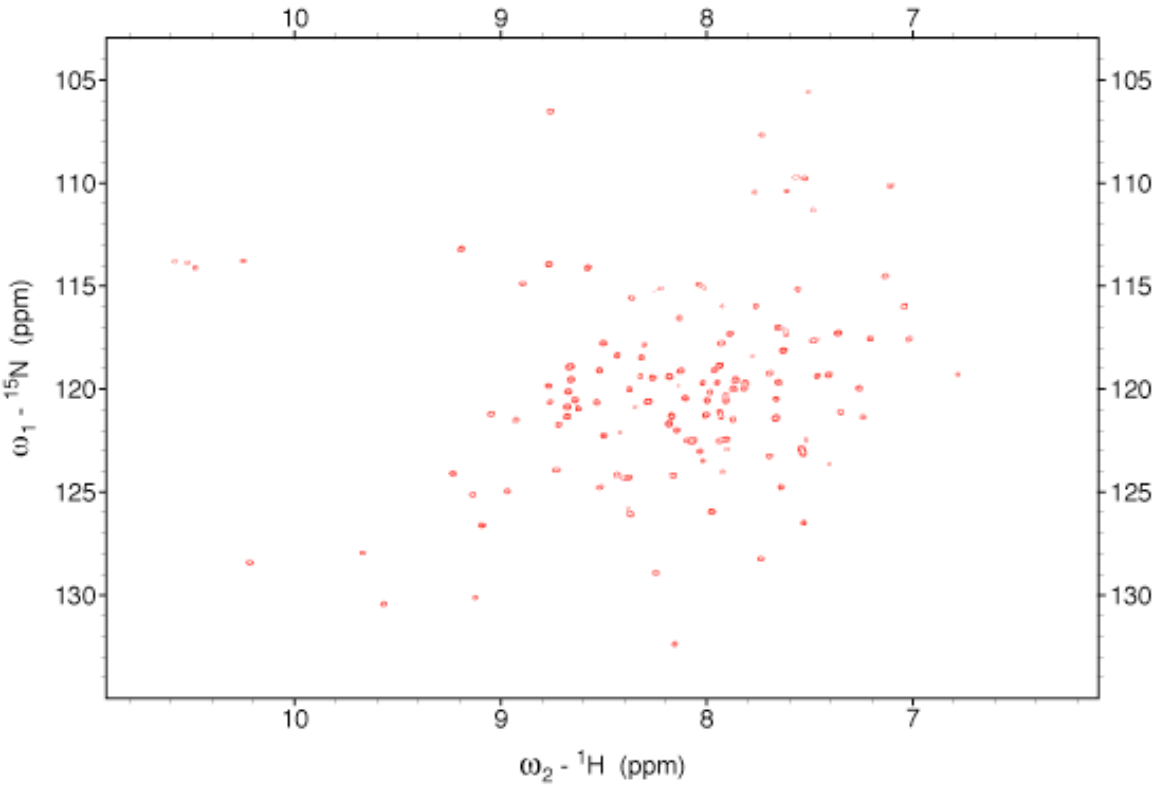

Figure 4.5. III-type (NzHb to NzHa) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin
The spectrum displayed is acquired at 35 C with a spectrometer field of 500 MHz and a relaxation time τ = 60 ms.

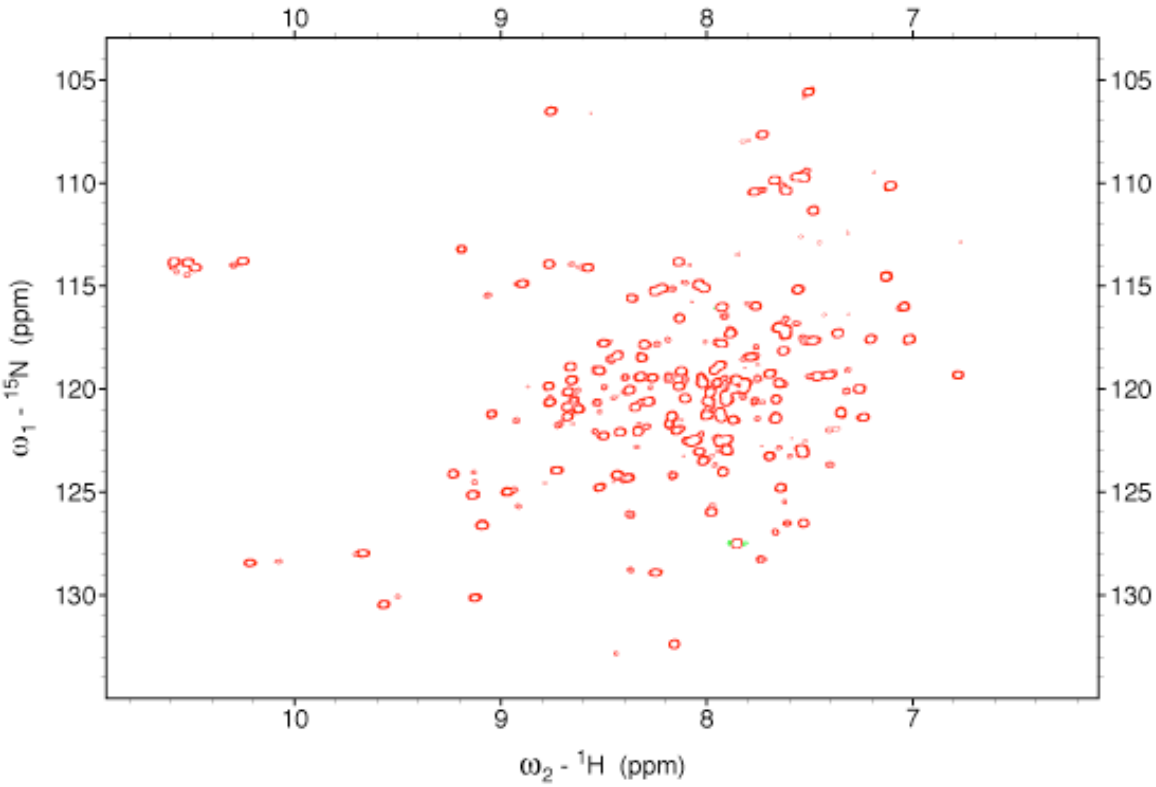

Figure 4.6. IV-type (NzHa to NzHb) η_zK relaxation spectrum for smMLCKp-bound Ca^{2+}-saturated chicken calmodulin
The spectrum displayed is acquired at 35 C with a spectrometer field of 500 MHz and a relaxation time τ = 60 ms.

112 of 146 non-proline CaM residues and 133 of 146 non-proline CaM/smMLCKp residues are assigned. In CaM, 70 of 112 residues are alpha-helical, with the remaining 42 residues having loop structure. In CaM/smMLCKp, 84 of 133 residues are alpha-helical, with the remaining 49 residues having loop structure. N42, F68, M76, and D118 are assigned in CaM and unassigned in CaM/smMLCKp; M36, S38, G40, E47, N53, G59, N60, T62, R74, D80, F92, G96, E104, R106, H107, V108, N111, G113, E119, E123, M124, G132, Y138, E139, and M144 are assigned in CaM/smMLCKp and unassigned in CaM.

Figures 4.7 and 4.8 illustrate representative X(t) and Y(t) fits for residue I100 of free and smMLCKp-bound calmodulin, respectively. Fit results at ±1 bootstrap ensemble population standard deviation (as detailed in Chapter 3 and illustrated in Figures 4.4A and 4.4B) are marked by green lines.

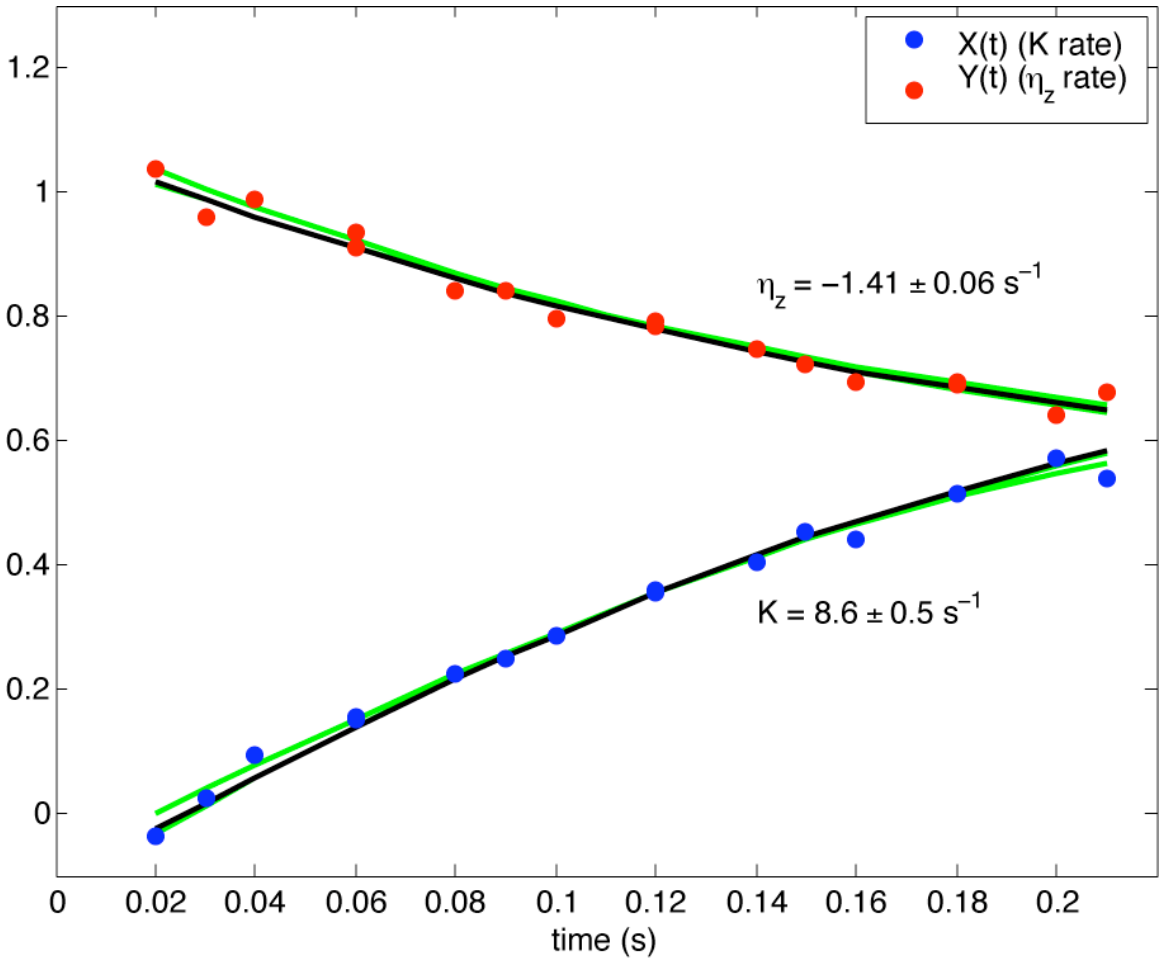

Figure 4.7. Representative X(t) and Y(t) fits for residue I100 of free Ca^{2+}-calmodulin.
Fit results at ±1 bootstrap ensemble population standard deviation (as detailed in Chapter 3 and illustrated in Figures 4.9 and 4.10) are marked by green lines.

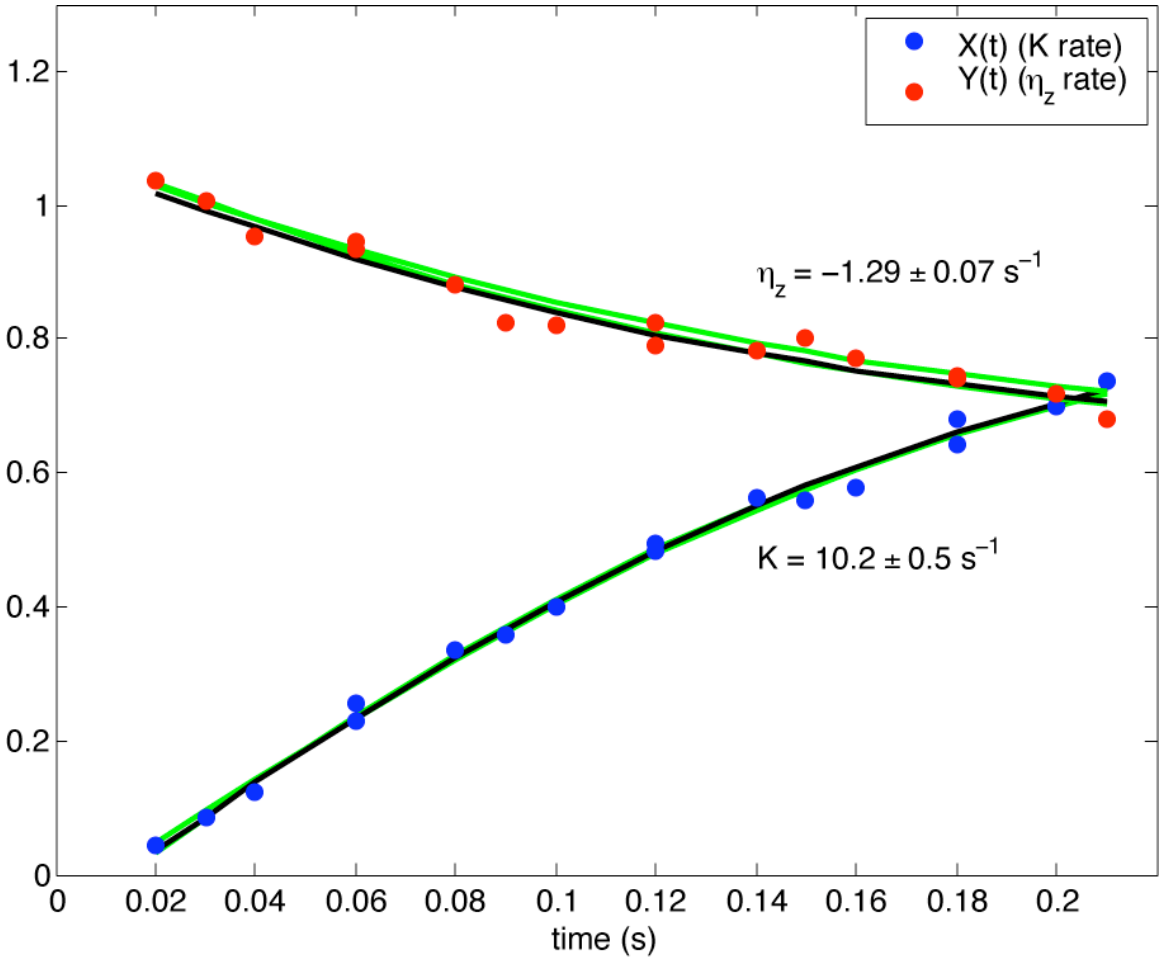

Figure 4.8. Representative X(t) and Y(t) fits for residue I100 of smMLCKp-bound Ca^{2+}- calmodulin.
Fit results at ±1 bootstrap ensemble population standard deviation (as detailed in Chapter 3 and illustrated in Figures 4.9 and 4.10) are marked by green lines.

Figures 4.9 and 4.10 illustrate the distribution of K and η_z fits, respectively, for residue I100 of smMLCKp-bound CaM. This distribution is based on a bootstrap ensemble of 200 permutations of I100's symmetric reconversion data. Similar distributions are prepared for all residues and the mean and standard deviation of each distribution are used as estimators of the true values of K and η_z.

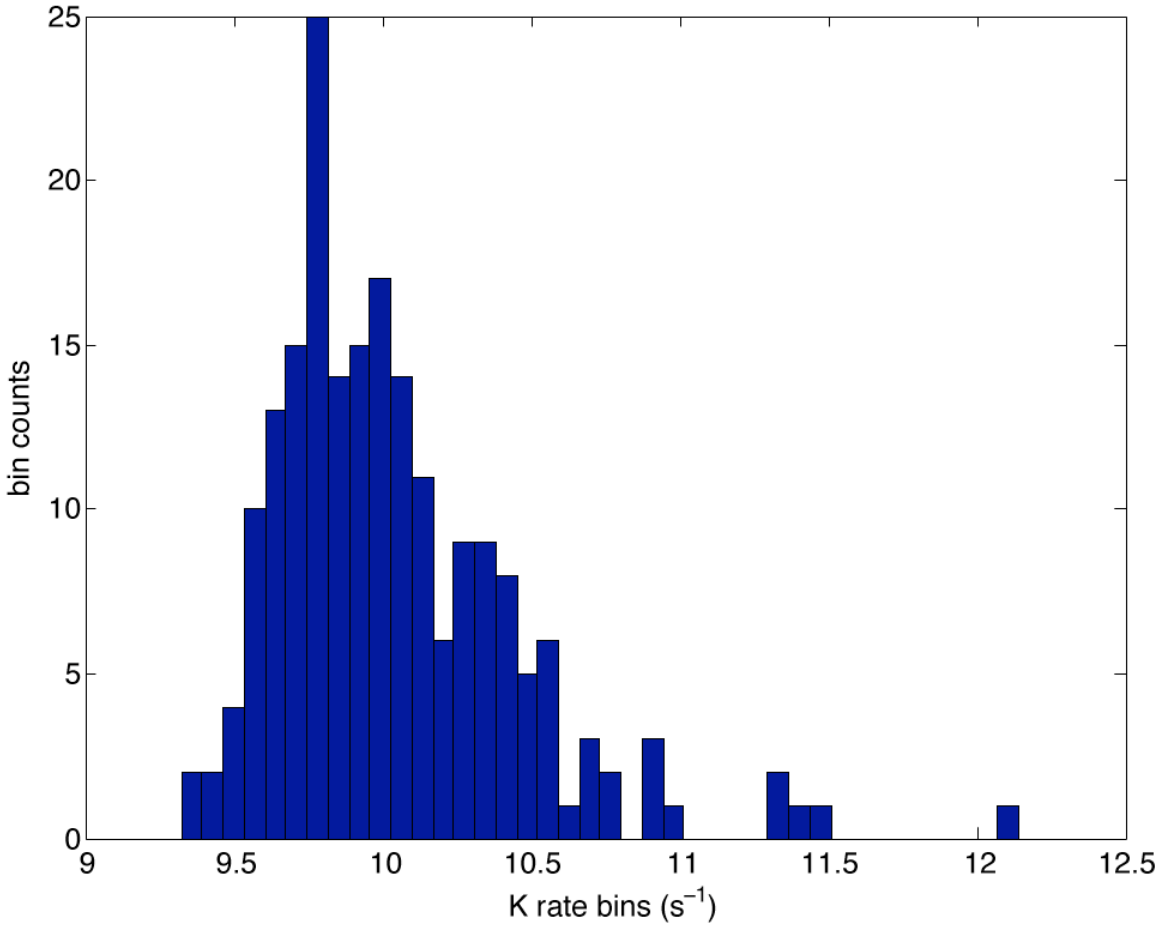

Figure 4.9. Distribution of K fits for residue I100 of CaM/smMLCKp.
This distribution is based on a bootstrap ensemble of 200 permutations of I100's symmetric reconversion data.

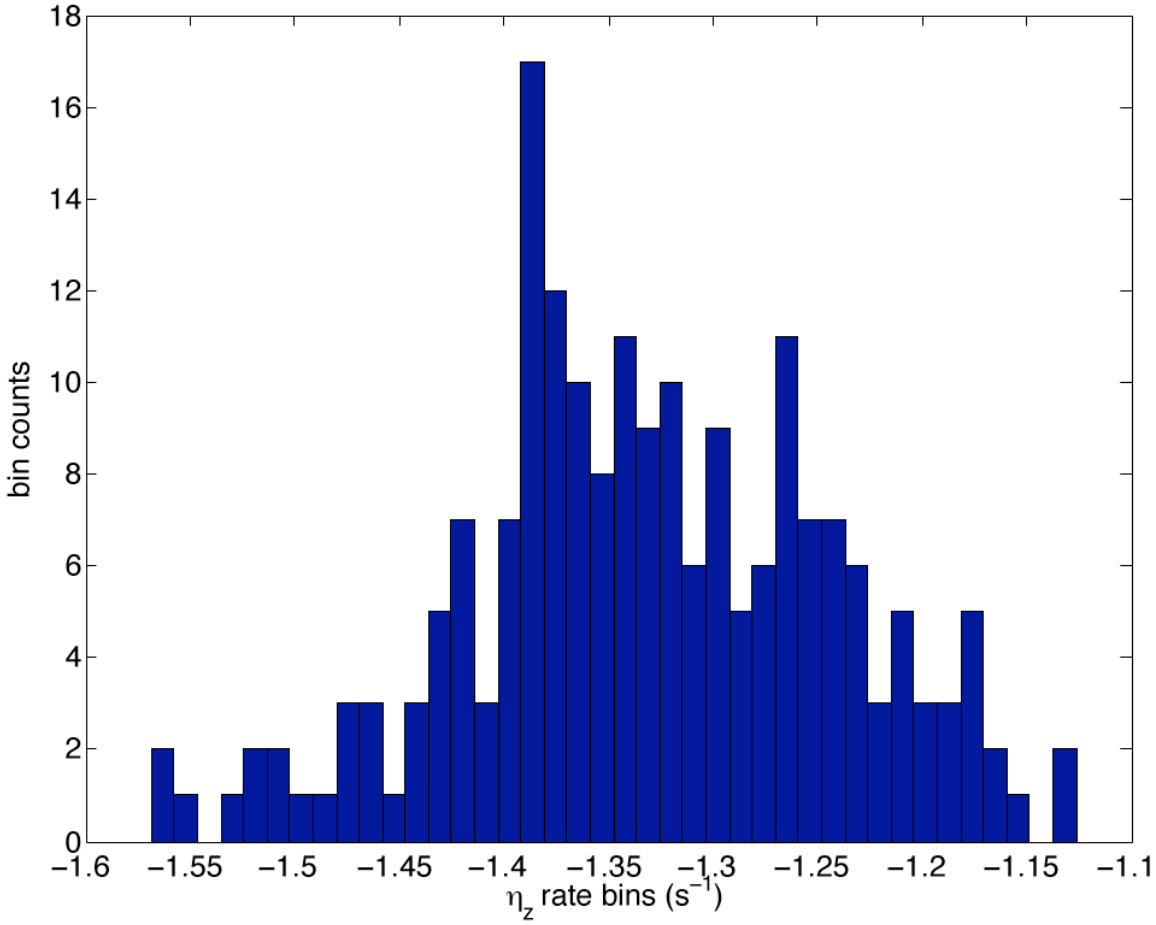

Figure 4.10. Distribution of η_z fits for residue I100 of CaM/smMLCKp. This distribution is based on a bootstrap ensemble of 200 permutations of I100's symmetric reconversion data.

Figure 4.11 illustrates an η_{xy} fit for the Wang/Rance/Palmer η_{xy} experiment[110] on residue I100 of CaM/smMLCKp. Fit results at ±1 standard deviation are marked by green lines.

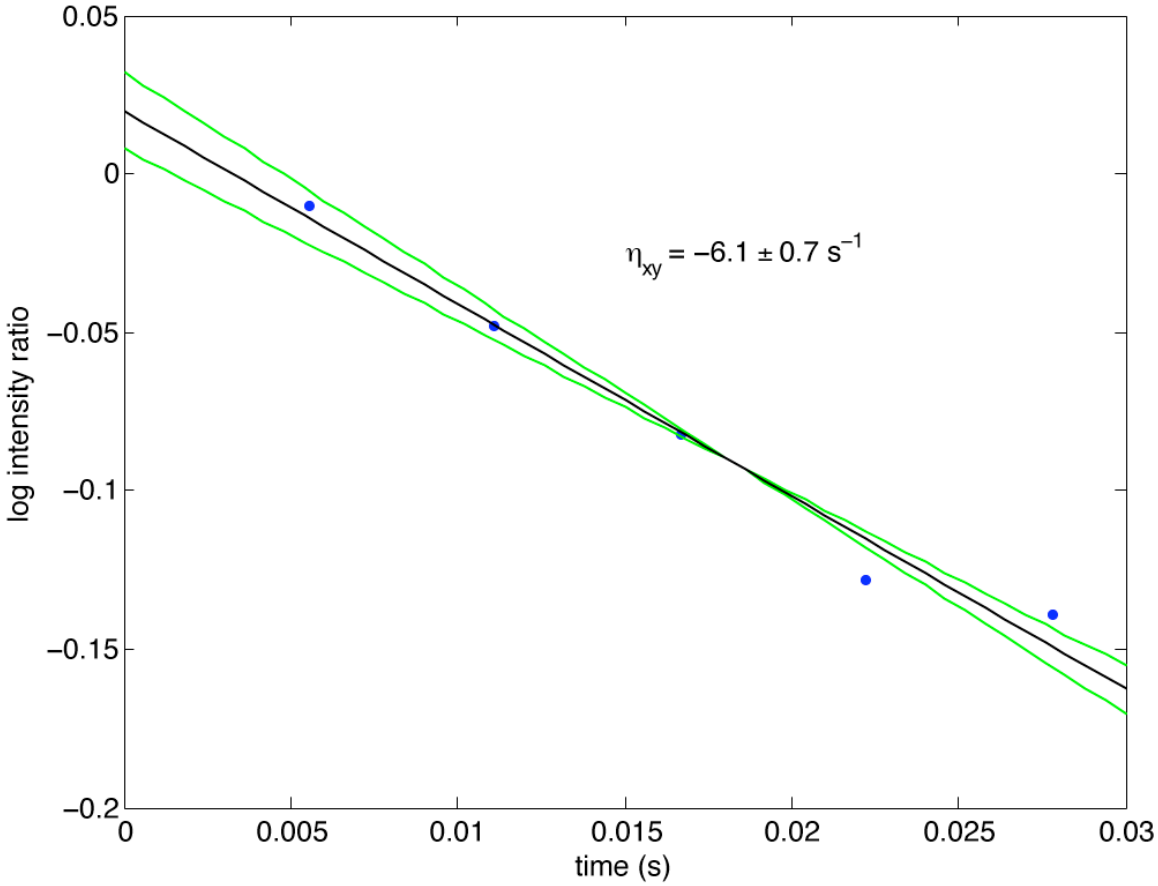

Figure 4.11. Representative η_{xy} fit for residue I100 of CaM/smMLCKp. Fit results at ±1 standard deviation are marked by green lines.

η_{xy} and η_z rates are used to calculate D_i values for individual residues, and individual residues are fit to produce an overall diffusion tensor picture for the entire molecule. Figure 4.12 illustrates D_i calculated from the η_{xy}/η_z ratios of free CaM and CaM/smMLCKp, respectively, as a function of $P_2(\cos \vartheta_i)$, ϑ_i being the off-angle of the i^{th} ^{15}N-$^{1}H_N$ amide bond vector axis from the symmetry axis of the axially symmetric diffusion tensor resulting from PDBINERTIA. QUADRIC_DIFFUSION is then used to calculate an overall diffusion tensor estimate from the D_i data as a function of ϑ_i (AG Palmer III,

Columbia University). The resulting axially symmetric characteristic global diffusion times are listed in Table 4.1.

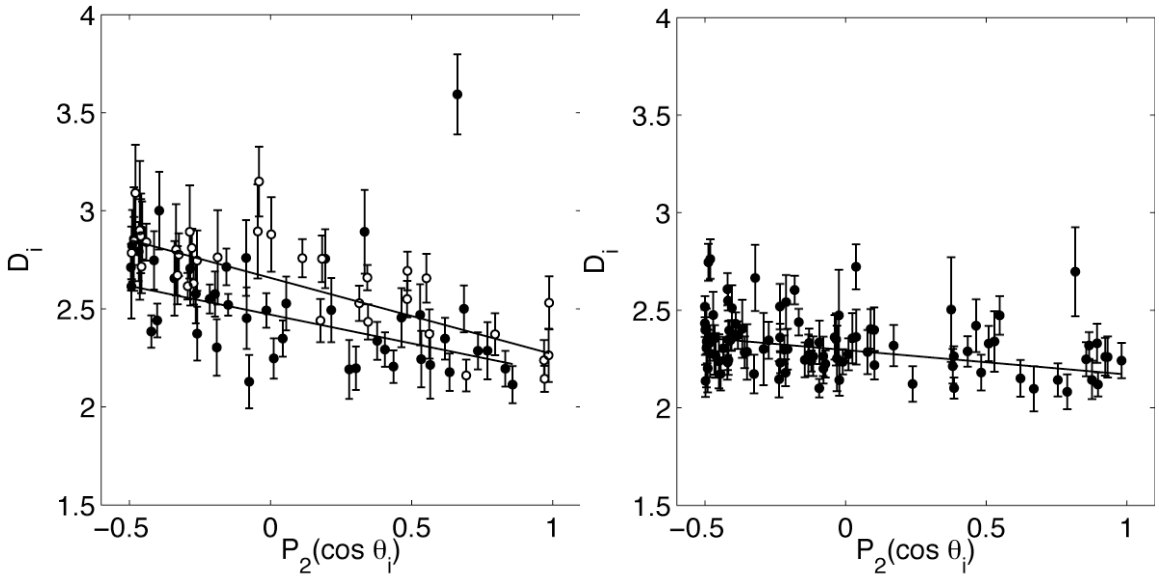

Figure 4.12. D_i vs. $P_2(\cos \vartheta_i)$ fit used to determine rotational diffusion constants for free CaM (left) and smMLCKp-bound CaM (right).
For free CaM, closed symbols indicate N-terminus data and open symbols indicate C-terminus data. Lines indicate best fit axially symmetric rotational diffusion tensor from the program QUADRIC_DIFFUSION.

	$\tau_c \parallel$ (ns)	$\tau_c \perp$ (ns)
CaM free, N-terminal	5.5 ± 0.2	7.7 ± 0.2
CaM free, C-terminal	4.9 ± 0.3	7.3 ± 0.3
CaM/smMLCKp	6.4 ± 0.2	7.8 ± 0.1

Table 4.1. CaM axially symmetric diffusion tensor characteristics.

Figure 4.13 illustrates K rates for free and bound calmodulin, respectively. Error bars in each case are derived from 200 Monte Carlo trials as in Figure 4.9. *K* values vary strongly over the protein due to diversity in local proton environments. Error bars are seen to be smaller than the variation in *K*-rates. Several residues in free and bound form

demonstrate very fast *K*-rates. These outliers can be explained by amide proton exchange: exchange between amide protons and water protons will scramble the prepared NzHa and NzHb states, giving rise to large *K* values.

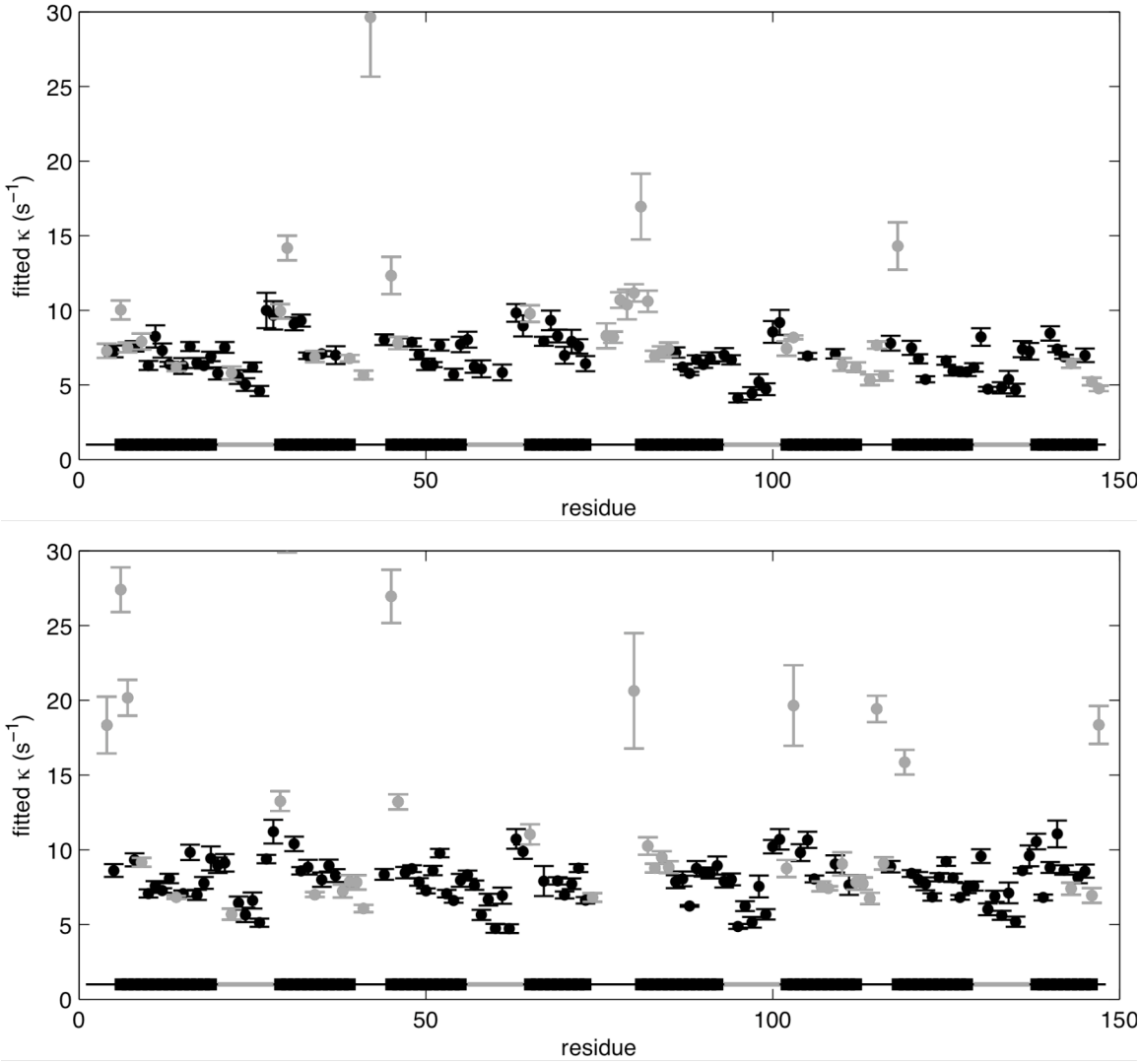

Figure 4.13. *K*-rates fitted to the experimental relaxation curves for free CaM (top) and CaM/smMLCKp (bottom).
The error bars were obtained from 200 Monte Carlo trials. Residues for which amide protons exchange faster than 1 s^{-1} are indicated in gray. These residues are not included in the data analysis of the following figures. In this and subsequent figures, thick horizontal bars denote helical regions of CaM, gray horizontal bars denote Ca^{2+}-binding loops, and thin horizontal lines denote other loops of CaM.

Amide proton exchange has been extensively studied for Ca^{2+}-CaM free in solution and bound to M13 skeletal muscle MLCK[114], a peptide similar to the smooth

muscle MLCK peptide used in this study.[58] Residues for which the free or M13-bound exchange rate was found to be larger than 1 s^{-1} are rendered in grey in Figure 4.14 and 4.15. The outliers in the raw *K* data for the CaM free/smMLCK system all correspond to amide protons with very fast exchange rates in the free/M13 system. For instance, the outlier for N42 at 30 s^{-1}, was reported to have an amide proton exchange rate k_{HX} = 120 s^{-1}. In principle, one should be able to correct experimental *K*-rates for amide proton mass-exchange rates and possible K contributions from water molecules, provided that all experiments are conducted under identical conditions and that amide proton mass-exchange rates and amide proton-water NOEs can be quantitatively measured. Currently, this data is not available. In addition, integration of the theory of K relaxation with theories of amide proton mass exchange poses a challenging (albeit interesting) theoretical task that lies beyond the scope of this thesis work. For this reason, residues with reported amide exchange rates faster than 1 s^{-1} in free or bound form are not included in further analysis of proton-proton dynamics in calmodulin.

Figure 4.14 illustrates Q, the ratio of the experimental *K*-rates to K_{rig} rates computed from structure, for free CaM and CaM/smMLCKp respectively. Due to X-ray structural imprecision and its effects on theoretical K_{rig} rates (see Chapter 2), relative errors in *Q* are considerably larger than relative errors in experimental *K*-rates. Deviations of *Q* from 1 are due to sub-nanosecond motion of the proton network in the vicinity of the amide protons. Local angular motions of the interproton vectors will reduce Q. In this case, *Q* is similar in concept to a classic S^2 order parameter.

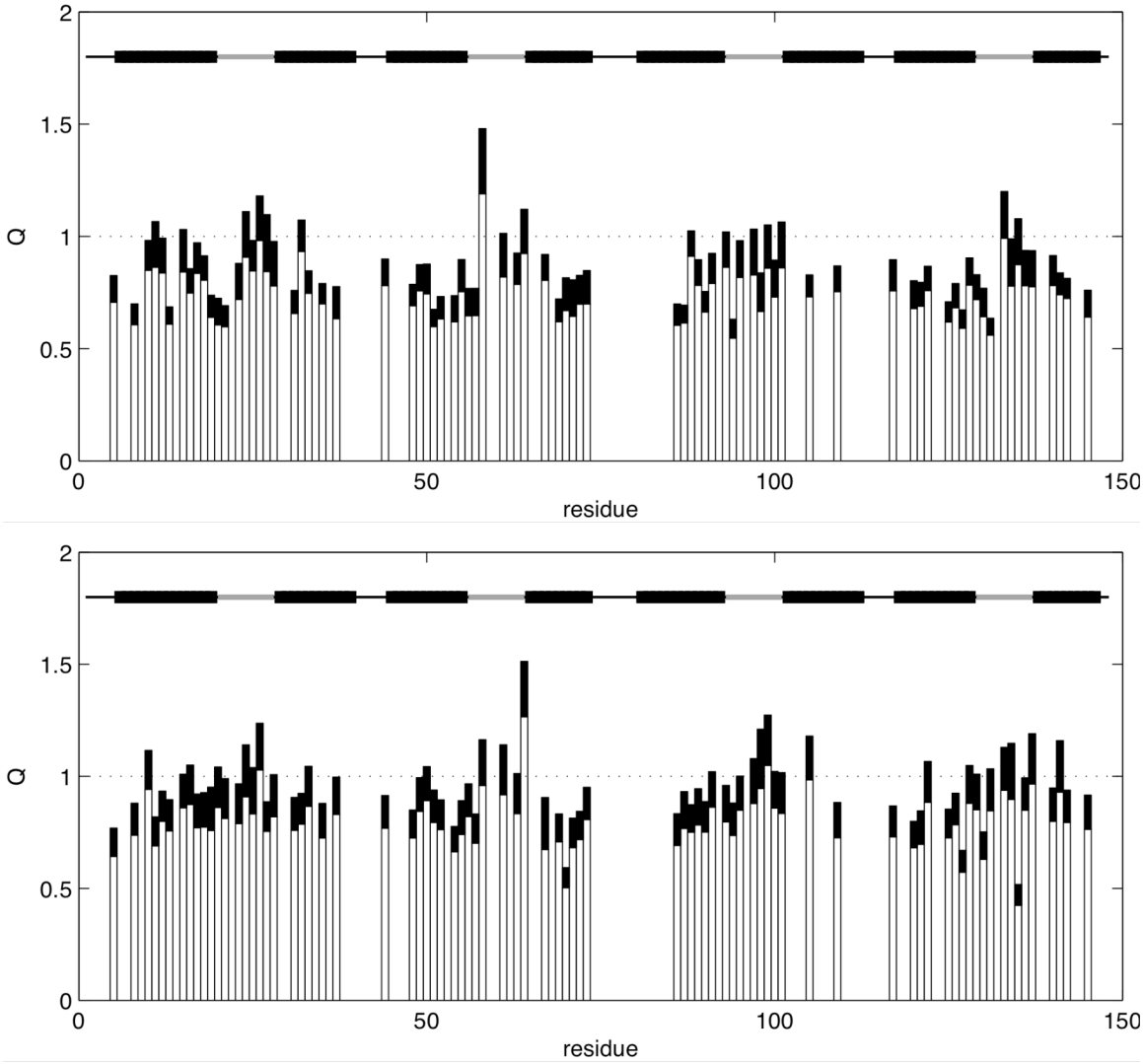

Figure 4.14. Q values for free CaM (top) and CaM/smMLCKp (bottom). Black bars indicate one standard deviation of Q value. Average *Q* values are: $<Q_{NT}>$ = 0.88 +/- 0.02 and $<Q_{CT}>$ = 0.89 +/- 0.02 for smMLCKp-bound calmodulin. (N-terminus = residues 1-76; C-terminus = residues 82-148).

In a non-covalent network it is also necessary to consider dynamic changes in interproton distances. If we consider the coordinates used to compute the theoretical *K*-rates to represent the dynamic average, we must consider two possibilities. Translational fluctuations of a neighboring proton along the virtual vector connecting it with the amide proton will foreshorten the apparent interproton distance, since r^{-6} averaging will cause the apparent increase of K rates by fluctuations on the 'shortened' side of the average position to outweigh the decrease of K rates by fluctuations on the 'lengthened' side of the average position. This leads to higher *K*-rates and can increase *Q* values above 1.

Translational fluctuations perpendicular to the virtual connection vector will make the interproton distance longer, reducing *Q*. In this case, *Q* is further reduced by the apparent angular motion of the interproton vector.

It is therefore possible to consider values of *Q* different from 1 as indicative of motion, but a necessary caveat is that the amplitude of the motion does not necessarily change monotonously as *Q* moves away from 1. Indeed, *Q* values at or near 1 could hide motions that counteract each others' effects on relaxation. For a detailed discussion of the *Q* parameter, readers are referred to the work of Post.[48]

4.4 Calmodulin discussion

Most *Q* values (Fig. 4.14) are smaller than 1 for both free and bound calmodulin. Average *Q* values are ($<Q_{NTerm}>$ = 0.82 +/- 0.02 and $<Q_{CTerm}>$ = 0.80 +/- 0.02) for free calmodulin and ($<Q_{NTerm}>$ = 0.88 +/- 0.02 and $<Q_{CTerm}>$ = 0.89 +/- 0.02) for smMLCKp-bound calmodulin (Table. The N-terminus comprises residues 1-76 and the C-terminus comprises residues 82-148.

	N-terminal domain <Q>	**C-terminal domain <Q>**
Free CaM	0.82 +/- 0.02	0.80 +/- 0.02
smMLCKp-bound CaM	0.88 +/- 0.02	0.89 +/- 0.02

Table 4.2. Mean Q values for N- and C-terminal domains of free and smMLCKp-bound Ca^{2+}-saturated calmodulin.

A paired t-test of *Q* values in free and smMLCKp-bound calmodulin indicates that the means of these values differ with a p-value of 0.001. The 95% confidence interval for the change in the mean *Q* is 0.03-0.11. For the N-terminal domain alone, the paired t-test p-value is 0.05 and the 95% confidence interval is 0.00-0.12, while for the C-terminal domain alone the paired t-test p-value is 0.01 and the 95% confidence interval is 0.02-0.14. These results clearly indicate global quenching of proton-proton dynamics in Ca^{2+}-saturated calmodulin upon smMLCKp binding. Figure 4.15 illustrates ΔQ between

free and smMLCKp-bound calmodulin as Q(bound) – Q(free). Significant and extensive changes in Q value occur throughout the molecule upon peptide binding. Our experimentally measured *Q* values agree well with values found in earlier theoretical and experimental work.[45; 48; 49]

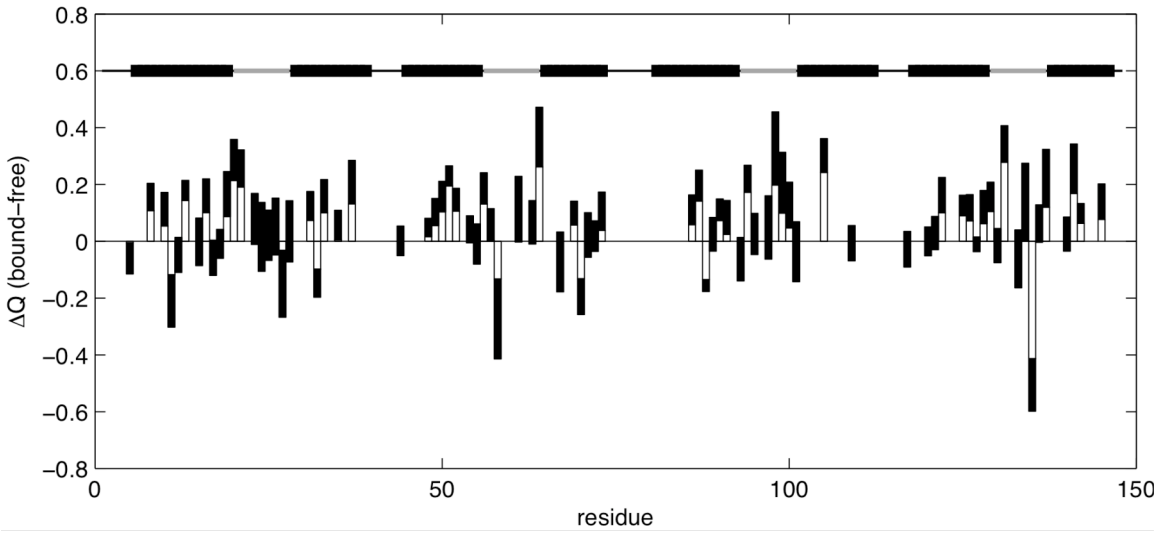

Figure 4.15. Δ*Q* between free and smMLCKp-bound calmodulin (Q(bound) – Q(free)).
Average Δ*Q* is +0.07. (N-terminus = residues 1-76; C-terminus = residues 82-148).

The distribution of *Q* values for bound and free states is shown in Figures 4.16. As discussed previously, mean and median *Q* values are both considerably less than 1. The shift in average *Q* between bound and free calmodulin constitutes a shift in the total distribution.

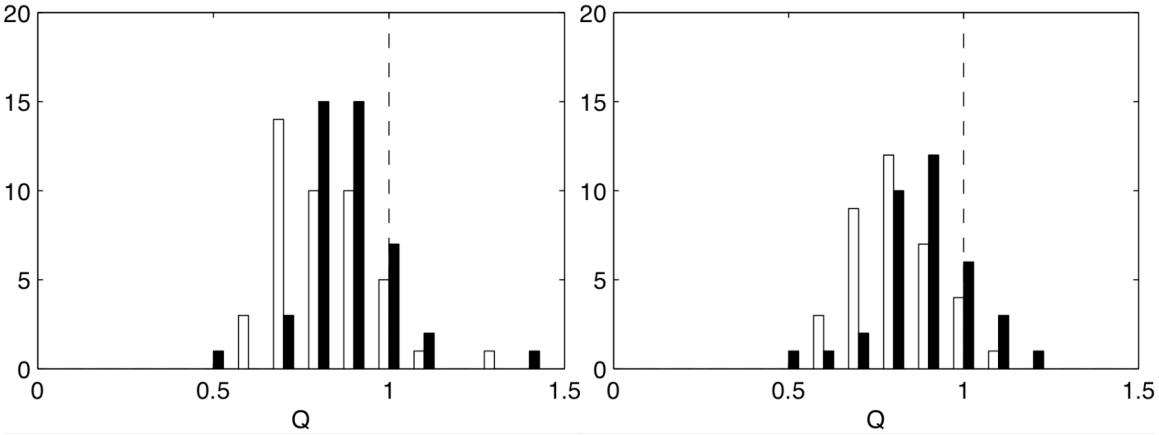

Figure 4.16. Distribution of *Q* values for free calmodulin (open bars) and smMLCKp-bound calmodulin N-terminal domain (closed bars). Left: N-terminus (residues 1-76) Right: C-terminus (residues 82-148).

Q values are dependent on accurate measurements of the rotational diffusion tensor, since the theoretical *K*-rates are constructed from measurement of the diffusion tensor. In this work, we obtained axially symmetric rotational diffusion tensors and associated site-specific t_c values from the ratio of the ^{15}N CSA / $^{15}N - {}^{1}H_N$ dipolar cross correlation ratios η_{xy} / η_z (see Chapter 2). We obtain values that are on average 10% smaller than the ^{15}N R_2/R_1 ratio-based τ_c values reported for this protein in the literature.[45] Without evaluating the relative merits of these methods, we note that the use of the literature values for t_c would cause <*Q*> to decrease by 10%.

<*Q*> is significantly less than one throughout both molecules, strongly implying the presence of sub-nanosecond angular motion of the proton environment of the amide protons. Since the effects of local motion on *Q* are complicated (see the arguments contained in Post[48] and Abseher et al.[50], as well as the Results section), we are interested in prying deeper into the statistics of *Q* factors in order to see if we can truly assign more motion to free calmodulin (<*Q*> = 0.81 +/- 0.02) than to smMLCKp-bound calmodulin (<*Q*> = 0.88 +/- 0.02).

Figure 4.17 compiles Q data according to the percentage contribution of different types of protons driving the amide proton flip rates.

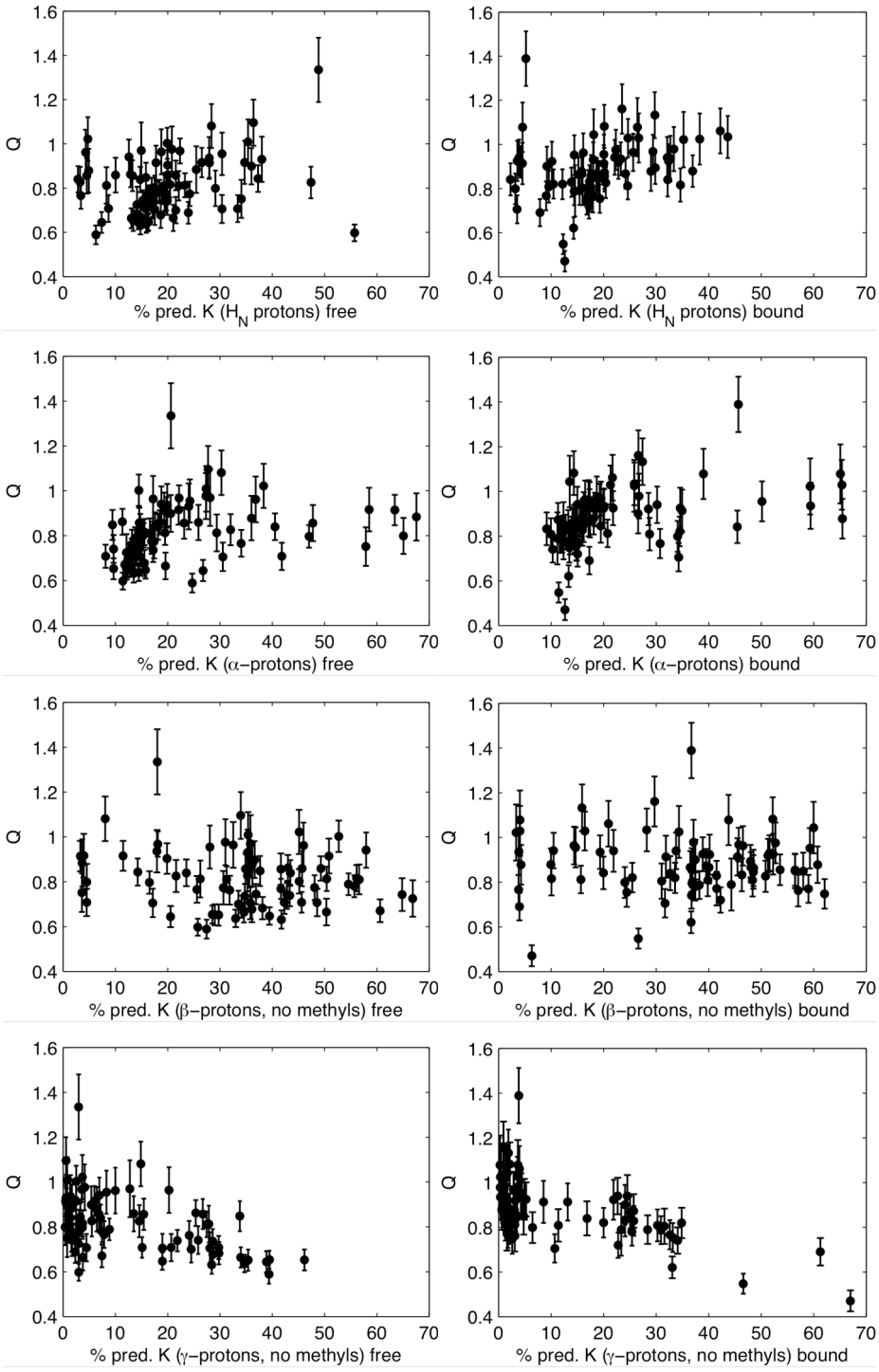

Figure 4.17. Structural statistics of Q.
The horizontal axis specifies the percentage of the contribution to theoretical *K* values as expected from the PDB coordinates of the protein for the indicated type of neighbor proton. The vertical axis gives the *Q* value associated with the proton. Left, CaM free; right, CaM/smMLCKp. From top to bottom: HN, HA, HB (no methyls) , HG (no methyls).

For example, protons which have 40% of their K rate contributed by Hα protons will be at 40% on the x-axis of the Hα plot. Dominance of Hα protons (mostly those on the same residue) in the K rate is associated with higher values of Q, indicating little motion of the protein backbone with regard to itself. This result is in accordance with ^{15}N relaxation measurements for calmodulin.[16] Dominance of sidechain γ protons (predominantly located on the residue's own sidechain) is associated with *Q* values tending to 0.5, indicating motions of very significant amplitude in these moieties. Dominance of H_β does not show any particular trend, likely reflecting their intermediate position between the H_α and H_γ's which have opposite trends. The statistics for the amide protons in the bound state shows similar trends as for the Hα protons. This would also indicate that the dynamics of the amide-amide proton vector are rather restricted.

Overall, trends in Fig. 4.17 are to be expected if Q, like an order parameter, does anticorrelate monotonically with the extent of motion. Before a quantitative relation can be established, more work is required. A comparison of experimental Q values with calculations of interproton disorder derived from computational molecular dynamics could prove particularly interesting.

If we take the observations above to indicate that <*Q*> is at the very least a *monotonic* function that anticorrelates with local sidechain motion, we may conclude that the protein relaxation network as seen by the amide proton is significantly more dynamic than that as felt by the amide nitrogen. This conforms to the common observation that all dynamical measures, such as S^2CH_3-detected dynamics[16; 115], backbone RDC-detected dynamics[116; 117], backbone ^{13}CO-detected dynamics[118], and cross-correlated dynamics[39] such as backbone ^{13}C-^{15}N tensor reorientation detected dynamics[119], indicate more motion than classical ^{15}N-detected dynamics. It also supports results from molecular dynamics calculations, which show large amplitude sidechain motions even in the protein interior[18].

If we assume that <*Q*> is a monotonic function that anticorrelates with local sidechain motion, we may associate the difference in *Q* values (Figure 4.15) and shift in the *Q* distribution between free and smMLCKp–bound calmodulin (Figure 4.16) with a change in motion. The difference in *Q* in Figure 4.15 is particularly reminiscent of the

work of Lee et al.[16], which shows a similar change in S^2CH_3 for the same protein upon binding of the same ligand. Surprisingly, and likely by chance, the increase in average Q, $\langle \Delta Q \rangle$ of 0.07 upon ligand binding corresponds exactly to the $\langle \Delta S^2CH_3 \rangle$ of 0.07 as determined by Lee et al.

4.5 Effect of the limited precision of experimental protein structures

In this work, we have used 1EXR and 2O5G X-ray crystal structures (1.0 Å and 1.08 Å resolution respectively). The Cruickshank error propagation procedure formally translates these resolutions into a 0.055 and 0.075 Å point-to-point distance imprecision, contributing ~5.5% and ~7.5% error margins to theoretical K-rates respectively. These errors are only slightly larger than the errors of fit, and are much smaller than the individual differences in K-rates. The salient points of our discussion in terms of structure and dynamics are not affected by the presence of X-ray structural error.

More commonplace structures are of 1.5-2.0 Å resolution and have a point-to-point distance precision of 0.3-0.4 Å. Distance errors of this size would render a calculation of Q essentially meaningless under the assumptions made above, since the error in Q measurements in combination with r^{-6}-dependent relaxation rates would cause large swings in theoretical K-rates.

However, this assertion may be overly harsh. Intuitively, the proton geometry immediately local to a given amide proton should be reasonably well defined by the restrictions placed upon it by just a few bond lengths and angles. Although estimates of a given atom's absolute position in space may be imperfect, the network of chemical bonds near an atom ensures that relative inte-atomic distances and orientations are preserved, and these may be known with high precision even when the overall precision of the structure is lower. This means that one may be able to make a more precise estimate of distances between nearby neighbors than would be expected from the error in the difference between two absolute position estimates.

The Hsc70 NBD studied in Chapter 6 has three crystal structures available: 1HPM at 1.7 Å, 2QWL at 1.75 Å, and 3HSC at 1.93 Å. In Chapter 6, we will examine the

situation of Hsc70 NBD K measurements and reevaluate the Cruickshank error propagation procedure in light of the results available for these two crystal structures.

4.6 Structure-free use of *K*-rates

In the absence of a high-resolution structure, possibilities remain for using *K*-rates in experimental structural biology. The key is the ability to compare two different molecular states.

When corrected for changes in overall rotational correlation times, changes in *K*-rates between states A and B indicate that the local proton network experiences a structural and/or dynamic perturbation during the transition from A to B. In the case of calmodulin, changes in *K* correlate with changes in Q (Fig. 4.18).

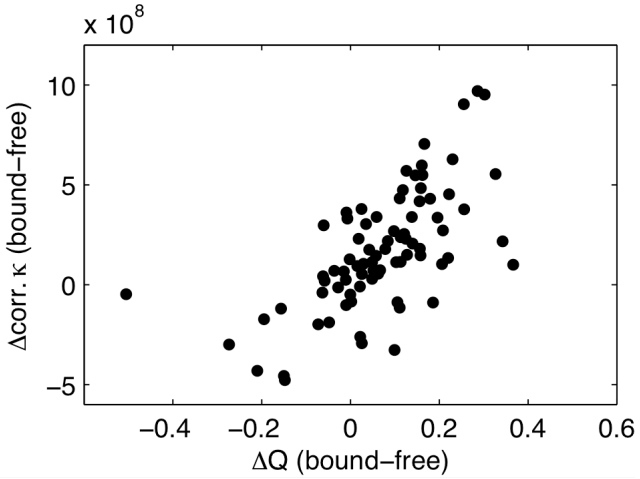

Figure 4.18. Correlation plot of change in corrected *K* *vs*. change in *Q* upon smMLCKp binding to Ca^{2+}-saturated calmodulin.
The value of the correlation coefficient R is 0.64.

Investigating this in more detail, we find that protons within 4 Å of the amide protons, reporting the 10 largest changes are concentrated within the hydrophobic clefts of both domains and in the face of helix 1 packed against the bound peptide (Figure 4.19).

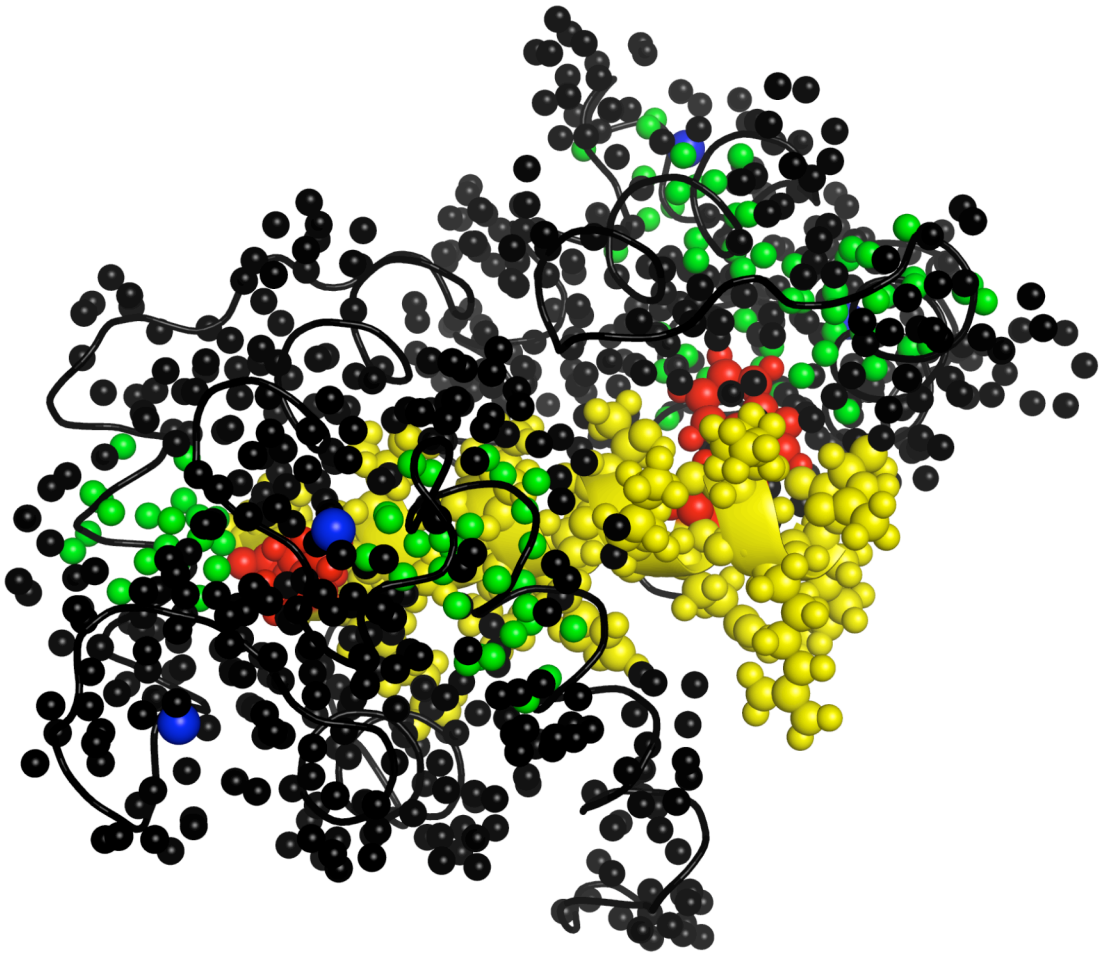

Figure 4.19. Illustration of ten largest corrected *K* changes (by absolute value) upon smMLCKp binding to Ca^{2+}-saturated calmodulin.
smMLCK peptide in yellow, evolutionarily conserved smMLCKp major hydrophobic anchors W5 and L18 in red. All smMLCK atoms are displayed. CaM protons are generally in black with CaM protons within 4 Å of amide protons reporting ten largest corrected *K* changes in green. Protons belonging to residues excluded due to amide proton exchange are not included in the illustration. Sphere size is set to one-half the van der Waals radius for clarity of illustration.

In this example, comparison of *K*-rates can help pinpoint the areas that change most in structure and dynamics.

Quantitatively, the mean N-terminal corrected *K*-rate increases by 5%, while the mean C-terminal corrected *K*-rate increases by 8%. Increased proton density or quenched

proton dynamics may be inferred from these changes, both of which would raise *K*-rates. One phenomenon likely goes hand in glove with the other.[120; 121] The changes in K rates may be compared with increases in mean *Q* by 7% and 10% for the N and C-terminal areas respectively. The overall correlation between site-specific changes in K rate and site specific changes in Q is 0.64 (p-value 1.3e-10). Even in the absence of structure, one would correctly conclude quenching in dynamics for these areas. This example suggests that changes in *K* may provide leads on the location of binding sites, regions of structural plasticity, and other functionally interesting areas in the absence of detailed structural information.

CHAPTER 5: PROTON-PROTON DYNAMICS OF UBIQUITIN

5.1 Ubiquitin introduction

Ubiquitin is a 76-residue, 8.5 kDa regulatory protein present in all eukaryotic cells. Ubiquitin underlies the essential cellular post-translational modification process of protein ubiquitination, in which one or more copies of ubiquitin are covalently attached to target proteins in order to mark them for later recognition by various cellular agents or to alter aspects of their operational state.

The first ubiquitin is attached to its target protein by formation of a peptide bond between a lysine sidechain moiety on the target protein and the C-terminal glycine of ubiquitin. Further ubiquitins may be attached to the first by formation of a peptide bond between a lysine sidechain on the n^{th} ubiquitin and the C-terminal glycine of the $(n+1)^{th}$ ubiquitin. This process is catalyzed by a chain of enzymes consisting of E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin-protein ligases.

Monoubiquitination of various proteins is essential to multiple cellular processes including histone regulation, endocytosis, DNA repair, virus budding, and nuclear export.[66; 122; 123] Polyubiquitination also plays multiple roles in the cell, with the choice of lysine linkage determining the form of the polyubiquitin linker and the resulting cellular signal. K48-linked polyubiquitin chains target proteins for proteasomal degradation, while K63-linked chains adopt extended conformations that mimic or augment the functionality of monoubiquitination. Polyubiquitin chains with varying lysine linkage sites, as well as branched polyubiuquitin chains, have also been reported in vitro.[66]

Ubiquitin has a long history as a subject of NMR experimentation. Beginning with the assignment of ubiquitin in 1987[124; 125] and continuing with the determination of its NH order parameters[126], rotational diffusion anisotropy[127], and ^{15}N chemical shift

anisotropy[128], the excellent resolution and sensitivity afforded by ubiquitin contributed to the development of numerous NMR techniques. Recent work of interest in ubiquitin includes Lange et al.'s determination of recognition dynamics across a broad range of timescales via the study of numerous residual dipolar coupling (RDC) alignment conditions.[70] These recognition dynamics span the space of ubiquitin crystal and co-crystal structures, suggesting that conformational selection explains the molecular recognition dynamics of ubiquitin.

Ubiquitin has been extensively structurally characterized both in solution and as attached to ubiquitinated proteins. Lange et al. were able to make their conclusions regarding recognition dynamics on the basis of 46 crystal structures involving ubiquitin free in solution or ligated to a ubiquitinated protein. Ubiquitin's excellent behavior in the NMR sample tube has led to several NMR structures of ubiquitin free in solution, including 1D3Z[67], 1XQQ[68], 2NR2[69], 2K39[70], and 2KLG[71]. However, there is only one X-ray structure of free ubiquitin deposited in the PDB: the original 1UBQ[129] structure of 1987 with 1.8 Å resolution.

Despite its small size, ubiquitin contains alpha-helical, antiparallel beta-sheet, 3_{10}-helix, and loop secondary structure. This large degree of structural variation is another characteristic that makes ubiquitin an interesting target for NMR spectroscopists. Figure 5.1 illustrates the structure of ubiquitin. The structure portrayed is the first member of the 1D3Z structural ensemble.

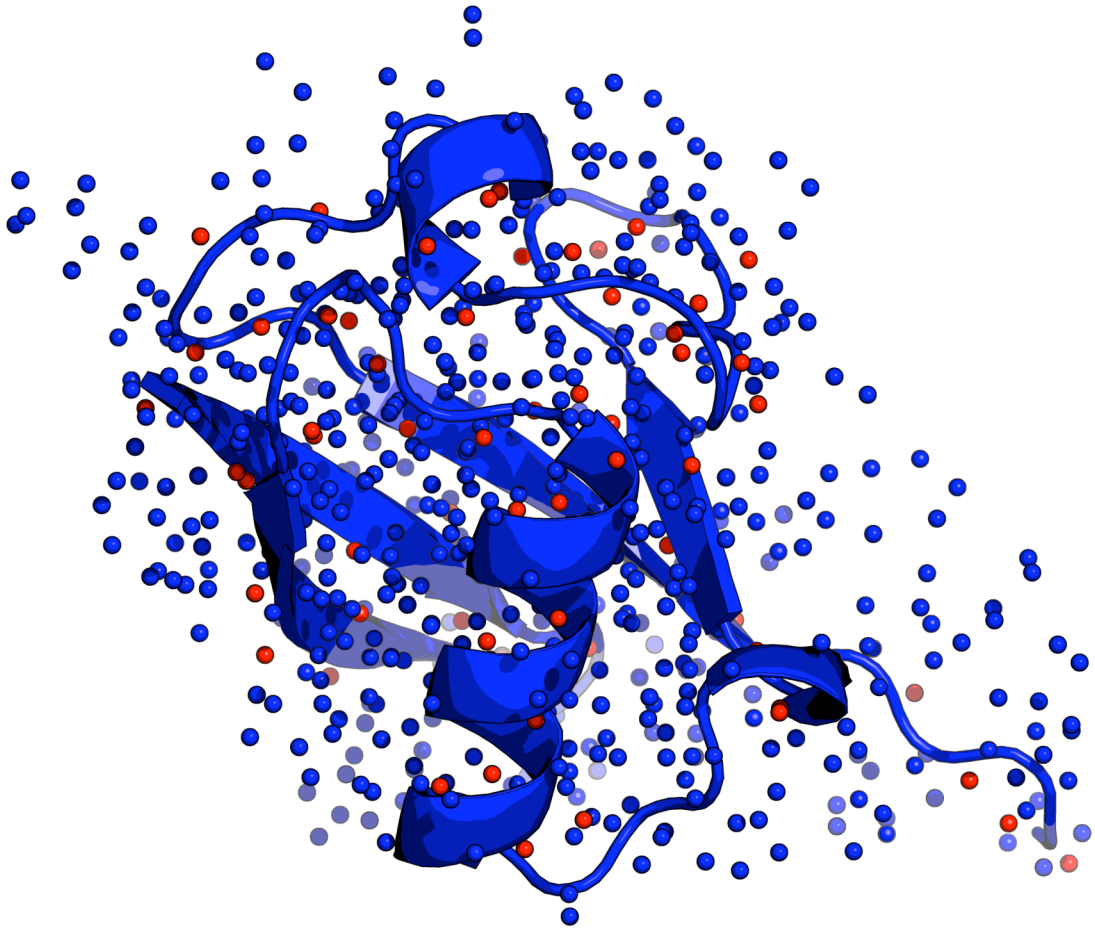

Figure 5.1. Structure of human ubiquitin.
The structure portrayed is the first member of the 1D3Z structural ensemble.[67] Spheres indicate proton locations, with red spheres denoting amide protons and blue spheres denoting all other protons. Sphere size is set to 1/4th of the van der Waals radius for clarity.

The available NMR structures of free ubiquitin have a wide variety of pedigrees. 1D3Z was created in order to test RDC validation of protein structure determination techniques and consists of an ensemble of the 10 lowest-energy conformers. 1XQQ was created as part of a protocol for simultaneous determination of protein structure and dynamics, and consists of an ensemble of 128 conformers, with all conformers submitted to the PDB. 2NR2 was created in a similar vein of inquiry and constitutes an attempt to

refine the techniques used in 1XQQ. It consists of an ensemble of 144 conformers, with all conformers submitted to the PDB. 2K39 is the result of an exhaustive analysis of ubiquitin RDCs across 36 NH, 6 H^NC'/NC', and 11 sidechain methyl RDC data sets with the aim of elucidating conformational dynamics at timescales from pico- to microseconds. It contains 116 conformers. 2KLG is the result of experiments on improving NMR protein structure determination with the addition of paramagnetic relaxation enhancement information, and contains 20 conformers.

This large array of solution structures of ubiquitin allows us an interesting opportunity to compare experimental K rates with K_{rig} rates derived from several different structural methodologies. The focus of this chapter will be on a comparison of NMR structural ensembles with the results of the η_zK experiment. The work done on ubiquitin will also allow further validation of the η_zK methodology in a small protein system.

5.2 Ubiquitin materials and methods

Ubiquitin experiments were carried out on a previously synthesized 2.0 mM sample of ^{15}N-labeled protonated human ubiquitin in 90/10% H_2O/D_2O at pH 4.8 and 300 K. All NMR spectroscopy on ubiquitin is carried out on a Bruker Avance 500 MHz spectrometer equipped with a triple-resonance triple-axis gradient room-temperature probe. All relaxation experiments consist of 2D ^1H_N-^{15}N TROSY-HSQC spectra acquired at several different relaxation delays. The recycle delay is set to 1.0 s.

η_zK experiments are conducted with 2048 complex points in the proton dimension and 120 complex points in the nitrogen dimension, with the carrier set at 4.773 and 117.473 ppm respectively. The experiment is conducted with 17 values of the relaxation delay τ: 20, 30, 40, 60, 60, 80, 90, 100, 120, 120, 140, 150, 160, 180, 180, 200, and 210 ms. η_{xy} experiments are performed at values of N = 1, 2, 3, 4 and 5 for the relaxation delay τ, according to the relation $2\tau = N/J^1_{NH}$.

5.3 Ubiquitin results

Figure 5.2 illustrates a representative η_zK relaxation spectrum for human ubiquitin at 27 C with a spectrometer field of 500 MHz. The spectrum displayed is a II-type (NzHa to NzHb) crossover experiment at a relaxation time τ = 140 ms. This spectrum was acquired in about two hours and fifteen minutes, although significant time savings could be achieved with fewer points in the indirect dimension. The sparsity and cleanliness of the ubiquitin spectrum is evident, even for the diminished intensity of an early crossover-type experiment.

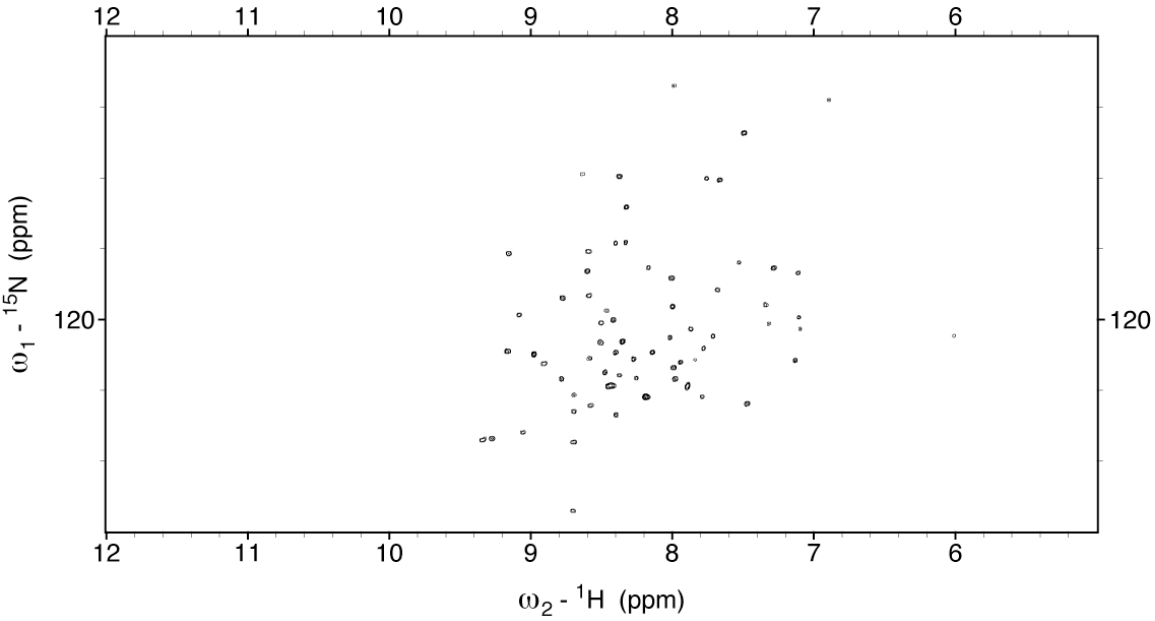

Figure 5.2. Representative η_zK relaxation spectrum for human ubiquitin at 27 C with a spectrometer field of 500 MHz.
The spectrum displayed is a II-type (NzHa to NzHb) crossover experiment at a relaxation time τ = 140 ms.

62 of 73 non-proline ubiquitin residues are assigned and included in the data analysis. These 62 residues are split among 13 alpha-helical residues, 2 3_{10}-helix residues, 23 beta-sheet residues, and 24 loop residues.

Figure 5.3 illustrates representative X(t) and Y(t) fits for residue V5 of ubiquitin, located in a beta-sheet region of ubiquitin. As in calmodulin, the agreement between theory and experiment is excellent. Fit results at ±1 bootstrap ensemble population standard deviation (as detailed in Chapter 3 and illustrated in Figures 5.4 and 5.5) are marked by green lines.

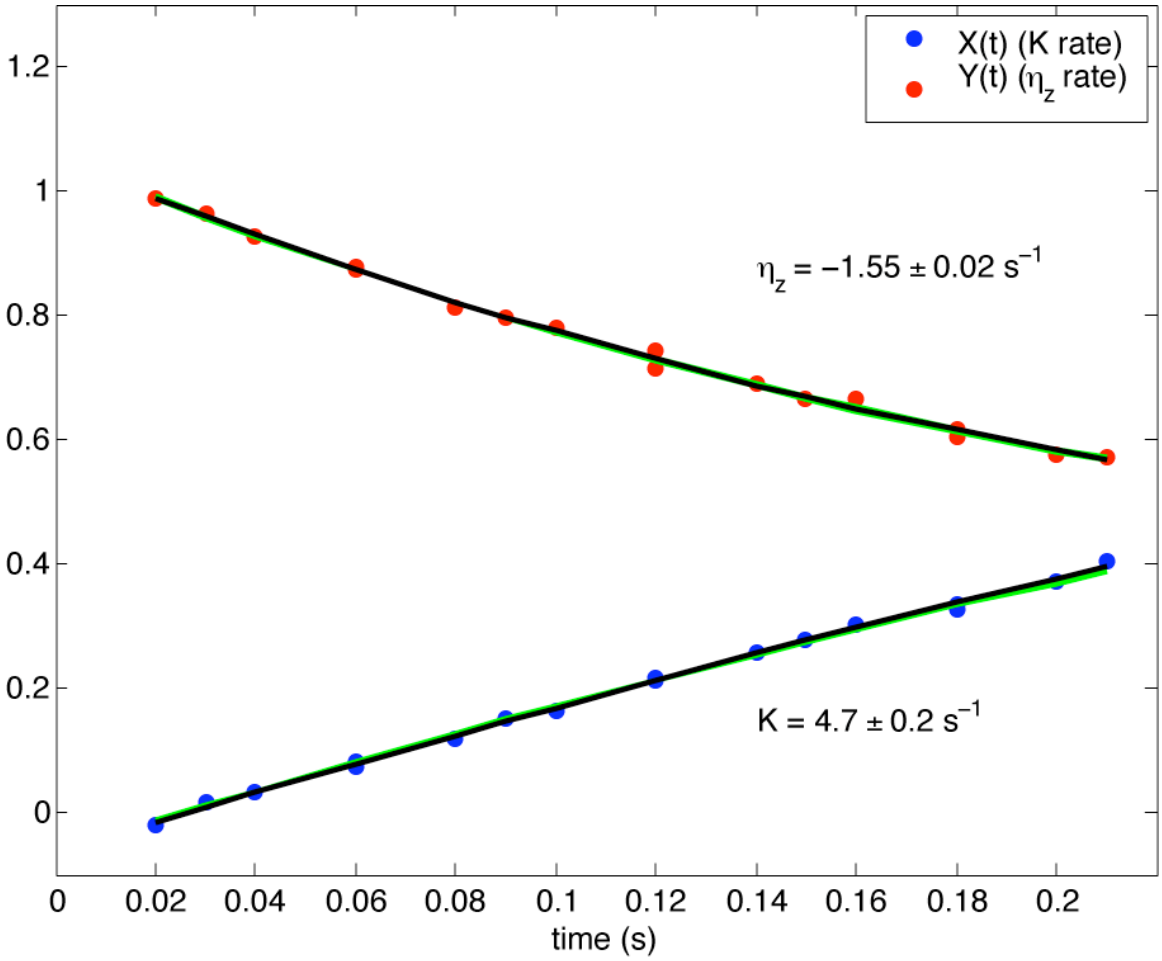

Figure 5.3. Representative X(t) and Y(t) fits for residue V5 of ubiquitin. Fit results at ±1 bootstrap ensemble population standard deviation are marked by green lines.

Figures 5.4 and 5.5 illustrate the distribution of K and η_z fits, respectively, for residue V5 of ubiquitin. This distribution is based on a bootstrap ensemble of 200 permutations of V5's symmetric reconversion data. Similar distributions are prepared for

all residues and the mean and standard deviation of each distribution are used as estimators of the true values of K and η_z.

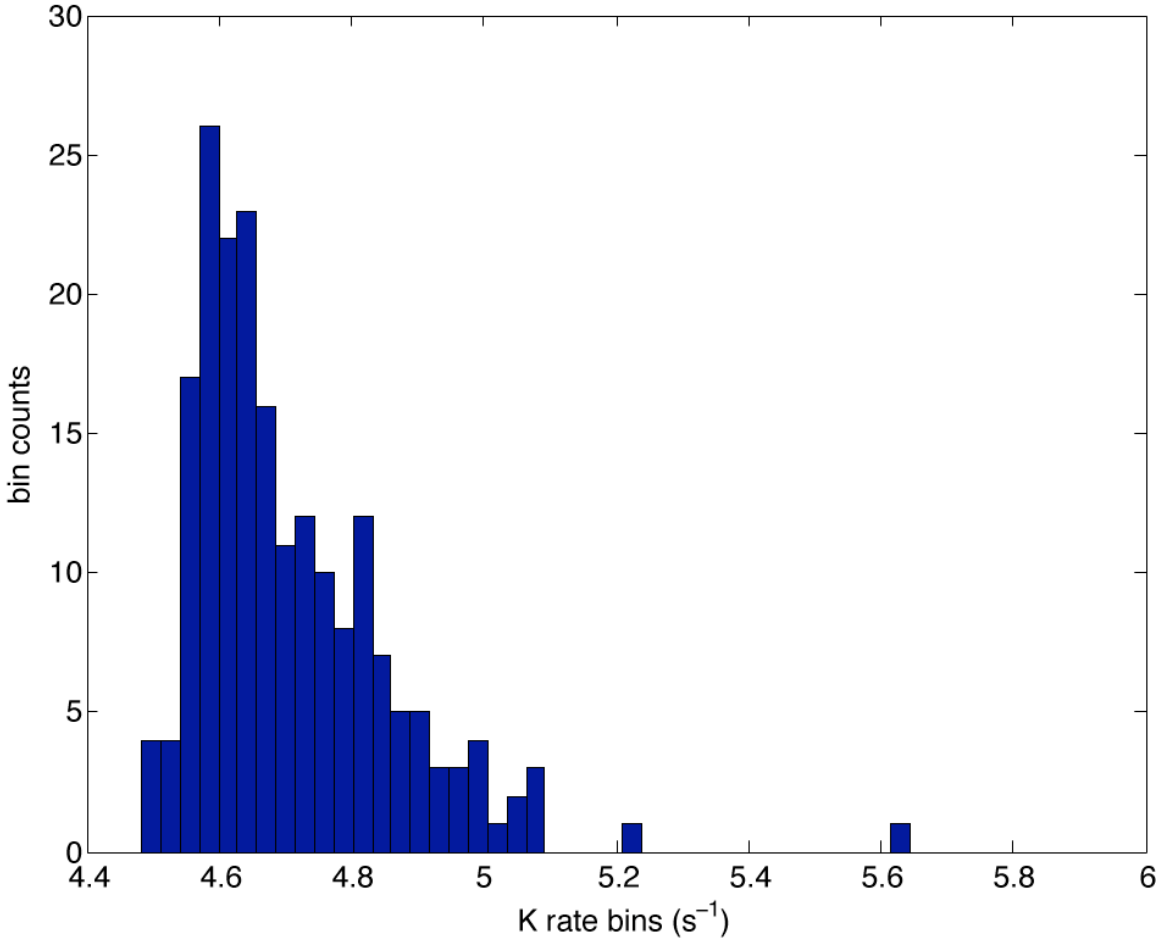

Figure 5.4. Distribution of K fits for residue V5 of ubiquitin.
This distribution is based on a bootstrap ensemble of 200 permutations of V5's symmetric reconversion data.

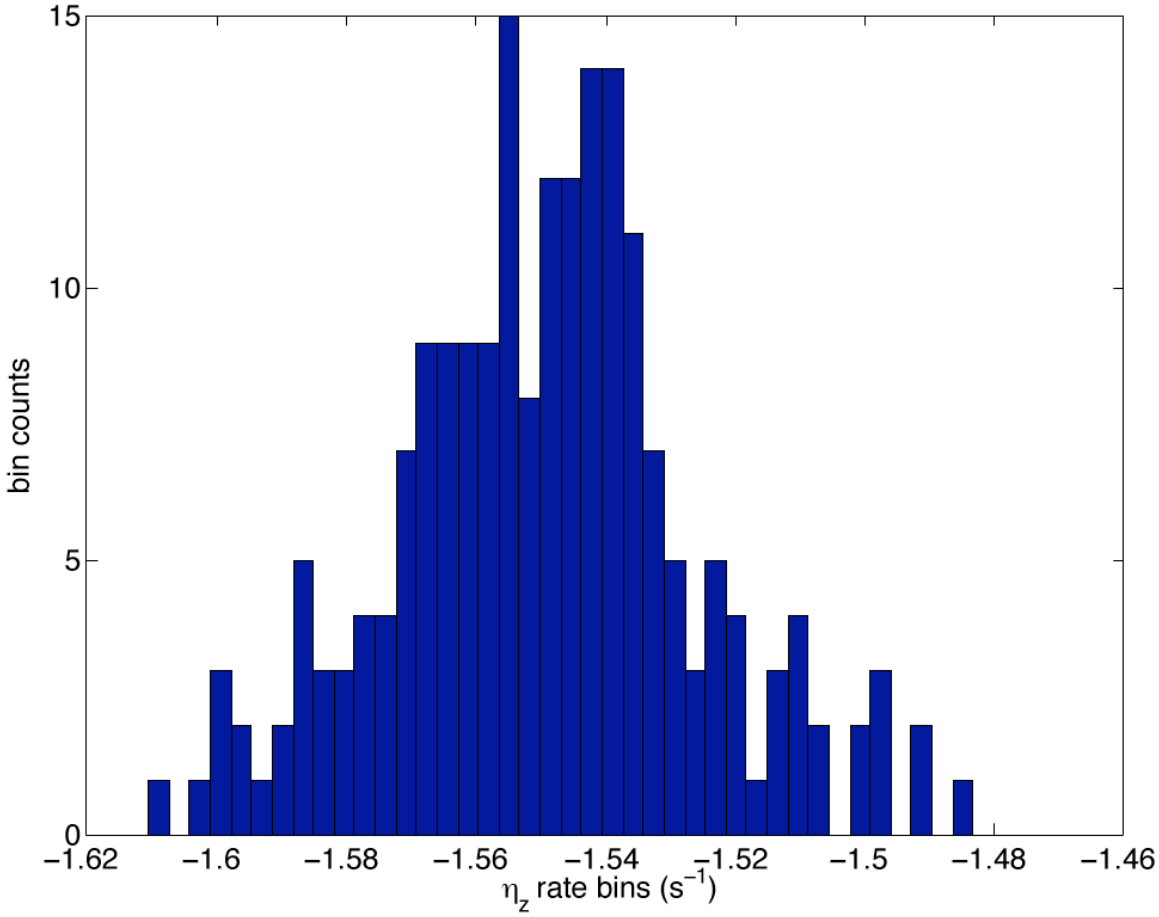

Figure 5.5. Distribution of η_z fits for residue V5 of ubiquitin.
This distribution is based on a bootstrap ensemble of 200 permutations of V5's symmetric reconversion data.

Figure 5.6 illustrates an η_{xy} fit for the Wang/Rance/Palmer η_{xy} experiment[110] on residue V5 of ubiquitin. Fit results at ±1 standard deviation are marked by green lines.

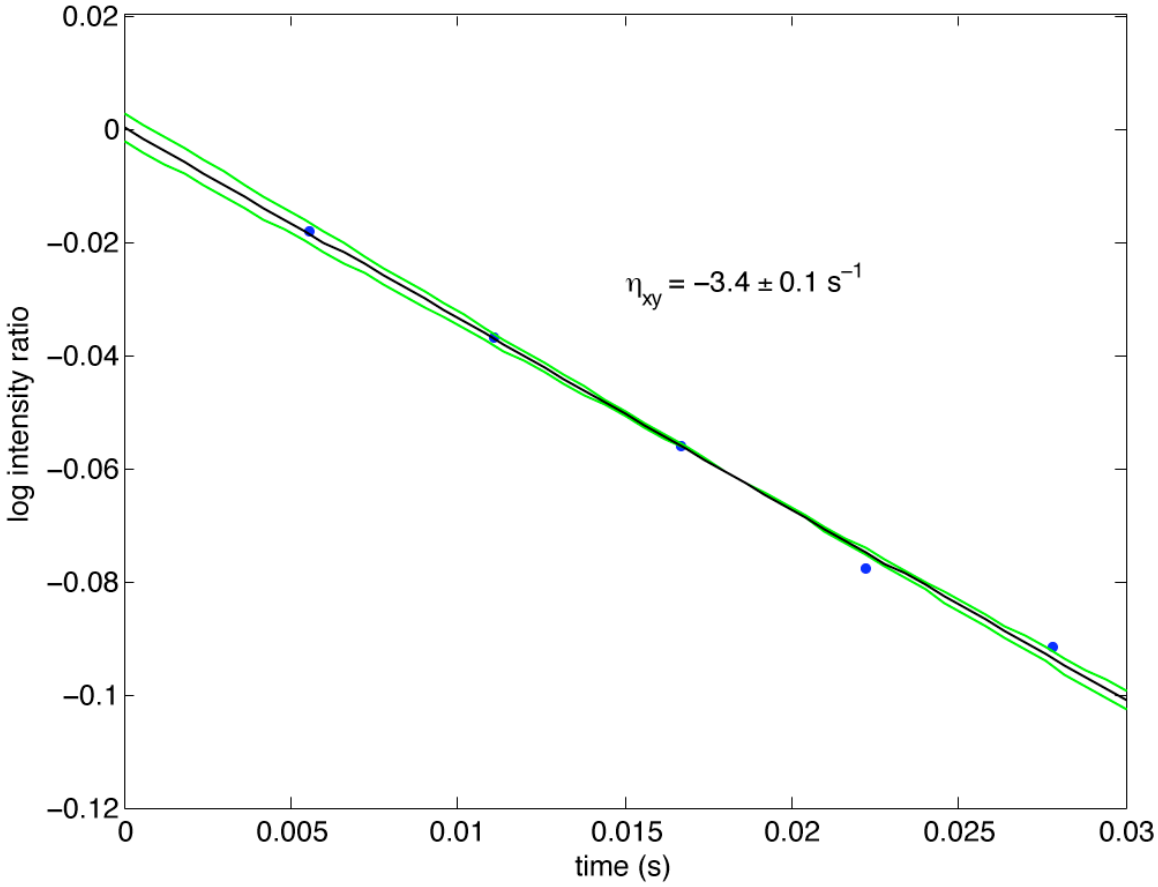

Figure 5.6. Representative η_{xy} fit for residue V5 of ubiquitin.
Fit results at ±1 bootstrap standard deviation are marked by green lines.

η_{xy} and η_z rates are used to calculate D_i values for individual residues, and individual residues are fit to produce an overall diffusion tensor picture for the entire molecule. Figure 5.7 illustrates D_i calculated from the η_{xy}/η_z ratios of ubiquitin as a function of $P_2(\cos(\vartheta_i))$, ϑ_i being the off-angle of the i^{th} ^{15}N-$^{1}H_N$ amide bond vector axis from the symmetry axis of the axially symmetric diffusion tensor resulting from PDBINERTIA. QUADRIC_DIFFUSION is then used to calculate an overall diffusion tensor estimate from the D_i data as a function of theta. The resulting axially symmetric characteristic global diffusion times are listed in Table 5.1.

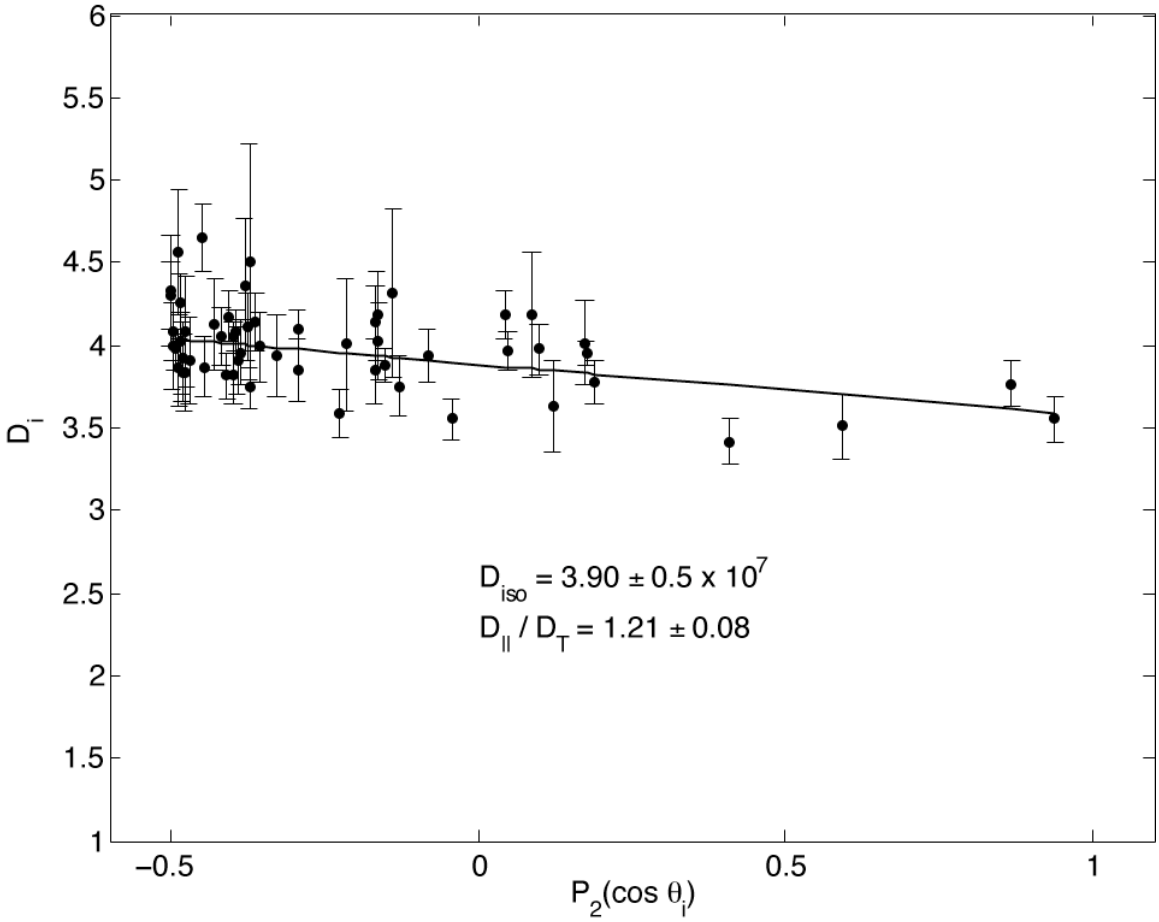

Figure 5.7. D_i vs. $P_2(\cos \vartheta_i)$ fit used to determine rotational diffusion constants for ubiquitin.
Line indicates best fit axially symmetric rotational diffusion tensor from the program QUADRIC_DIFFUSION.

Diso (10^7 s^{-1})	3.90 ± 0.05
$D_{||}/D_T$	1.21 ± 0.08
ϑ (rad)	0.3 ± 0.1
φ (rad)	3.8 ± 0.7

Table 5.1. Ubiquitin axially symmetric diffusion tensor characteristics.
Angles are defined with respect to the PDBINERTIA alignment of the first conformer of the 1D3Z NMR structure.

Figures 5.8 and 5.9 illustrate K and η_z rates, respectively, for ubiquitin. Error bars in each case are derived from 200 Monte Carlo trials as in Figures 5.4A and 4B.

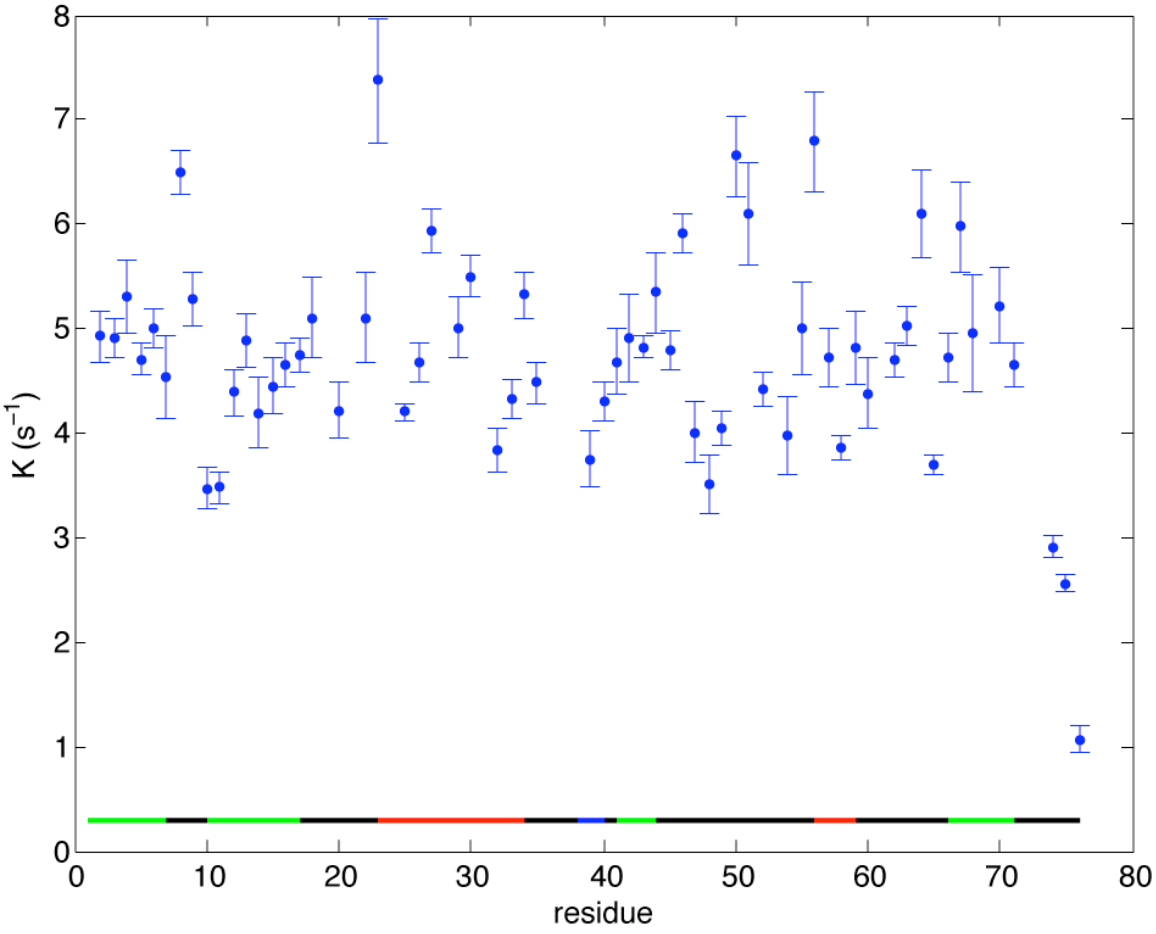

Figure 5.8. K rates for ubiquitin.
Error bars in each case are derived from 200 Monte Carlo trials as in Figures 5.4 and 5.5. The line at the bottom of the graph indicates protein secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, blue indicating 3_{10}-helix structure, and black indicating loop structure.

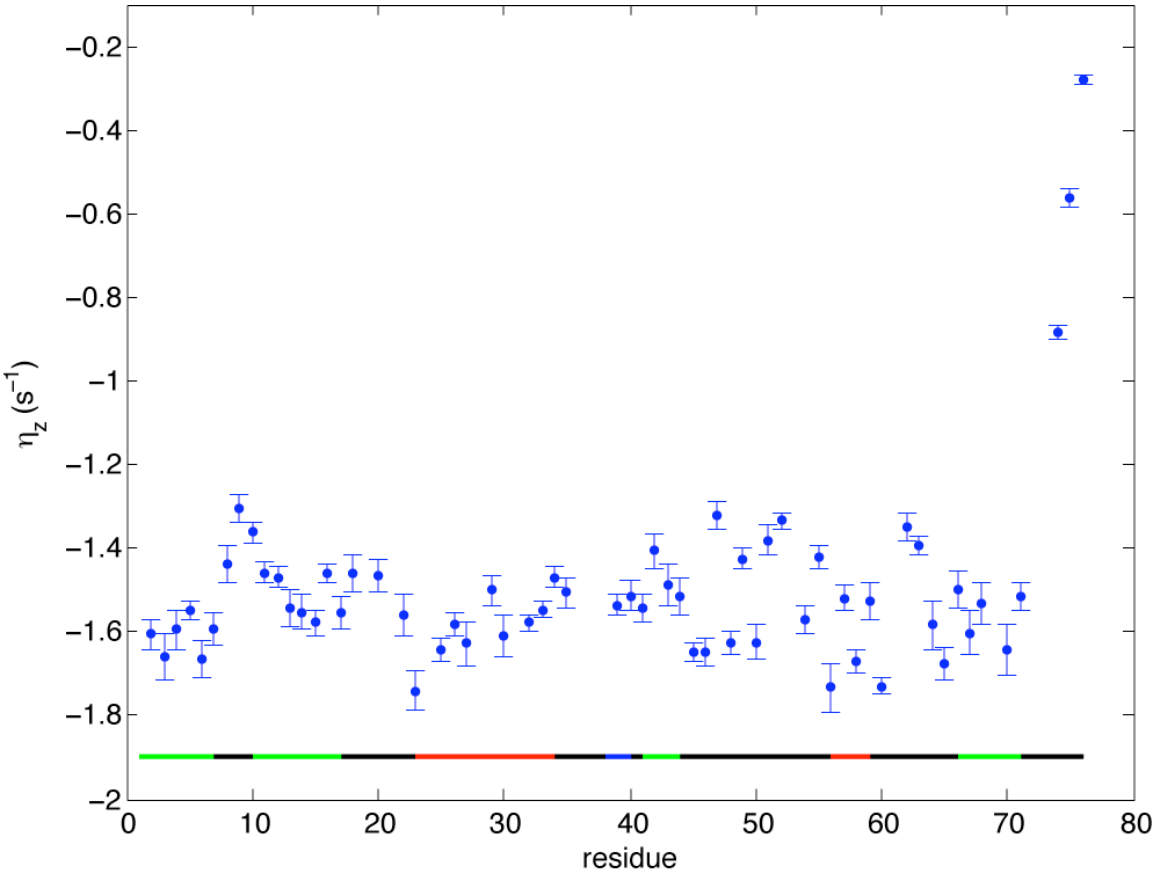

Figure 5.9. η_z rates for ubiquitin.
Error bars in each case are derived from 200 Monte Carlo trials as in Figures 5.4 and 5.5. The line at the bottom of the graph indicates protein secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, blue indicating 3_{10}-helix structure, and black indicating loop structure.

So far, the data presented have followed the course laid out for calmodulin in its free and smMLCKp-bound forms. According to the path previously charted, our next task is interpretation of experimental K rates by comparison with K_{rig} rates calculated from reference structures, and calculation of Q values from the ratio of these two quantities. Calmodulin has obvious choices of reference in the 1EXR and 2O5G ultra-high-resolution X-ray structures. A large number of possible options exist for ubiquitin. Which to choose? Counting only those structures reporting single ubiquitin molecules, there are 6 structures available. These have been described previously: the 1UBQ 1.8 Å X-ray structure plus 5 NMR ensembles: 1D3Z (10 structures), 1XQQ (128 structures), 2K39 (116 structures), 2NR2 (144 structures), and 2KLG (20 structures).

This bewildering array of structural options does not leave us with a clear choice of reference structure. However, the dilemma of which structure to use can be reframed as an opportunity to evaluate the quality of the NMR ensembles using experimental K rates. Although experimental K values will inevitably be altered by dynamics and dynamical variability over the protein sequence will cause this alteration to be uneven, "good" ubiquitin NMR conformers should still be reasonably well-correlated with experimental K values. (Amide proton exchange in the ubiquitin system can be safely neglected, since the sample pH of 4.7 is close to the .) All correlations that follow are calculated from residues 1-70 of ubiquitin, as residues 71-76 appear disordered in solution.

In order to test this assertion of reasonable correlation, we calculate the degree of correlation between experimental K rates and calculated K_{rig} rates for non-exchanging residues of bound and free forms of calmodulin. These correlations are listed in Table 5.2.

	K vs K_{rig} correlation coefficient (ρ)	***p*-value**	**ρ^2**
CaM N free (1EXR)	0.74	7.2e-09	0.55
CaM C free (1EXR)	0.82	2.2e-09	0.67
CaM bound (2O5G)	0.75	3.2e-19	0.56
Ubiquitin (1UBQ)	0.40	0.002	0.16

Table 5.2. CaM free/bound and ubiquitin X-ray structure K vs. K_{rig} correlations.

The ultra-high resolution 1EXR (1.0 Å) and 2O5G (1.08 Å) calmodulin structures exhibit correlation between calculated K_{rig} rates and experimental K rates of 0.7 or better. The lower-resolution 1UBQ (1.8 Å) ubiquitin structure exhibits a weaker 0.40 correlation. Although we should not draw conclusions too aggressively from these results, they support the idea of using experimental K / calculated K_{rig} correlations as rough guides to the quality of structural representation within NMR ensembles.

The next step is to evaluate the 5 available NMR ensembles by calculating K_{rig} rates and K/K_{rig} correlation coefficients for each conformer within the ensembles. Figure 5.10 displays histograms of correlation coefficient counts for conformers within 1D3Z, 1XQQ, 2K39, 2NR2, and 2KLG respectively. The means of these distributions are outlined in Table 5.3.

	Mean K vs K_{rig} correlation coefficient (ρ)	***p*-value of mean**	**Mean ρ^2**
1D3Z (10 conformers)	0.74 ± 0.03	6.4e-10	0.54
1XQQ (128 conformers)	0.40 ± 0.14	0.002	0.18
2K39 (116 conformers)	0.41 ± 0.11	0.001	0.18
2NR2 (144 conformers)	0.36 ± 0.17	0.005	0.16
2KLG (20 conformers)	0.42 ± 0.10	0.001	0.19

Table 5.3. Ubiquitin NMR structure K vs. K_{rig} correlations.

The 1D3Z ensemble is clearly able to contribute the greatest degree of explanation for the variance in the experimental K data. Furthermore, the 1D3Z ensemble provides by far the strongest ensemble-wide consensus on K_{rig} rates, as evidenced by its low standard deviation of correlation coefficient.

Further illustration of this result is provided by Table 5.4, which contains mean Q values (with errors) and standard deviations across all Q values calculated based on all conformers within 1D3Z, 1XQQ, 2K39, 2NR2, and 2KLG respectively. All structures have similar mean Q values. The distribution of Q values for 1D3Z has a significantly smaller standard deviation, indicating a narrower breadth of Q values. Table 5.4 contains mean Q values (with errors) and standard deviations across all Q values.

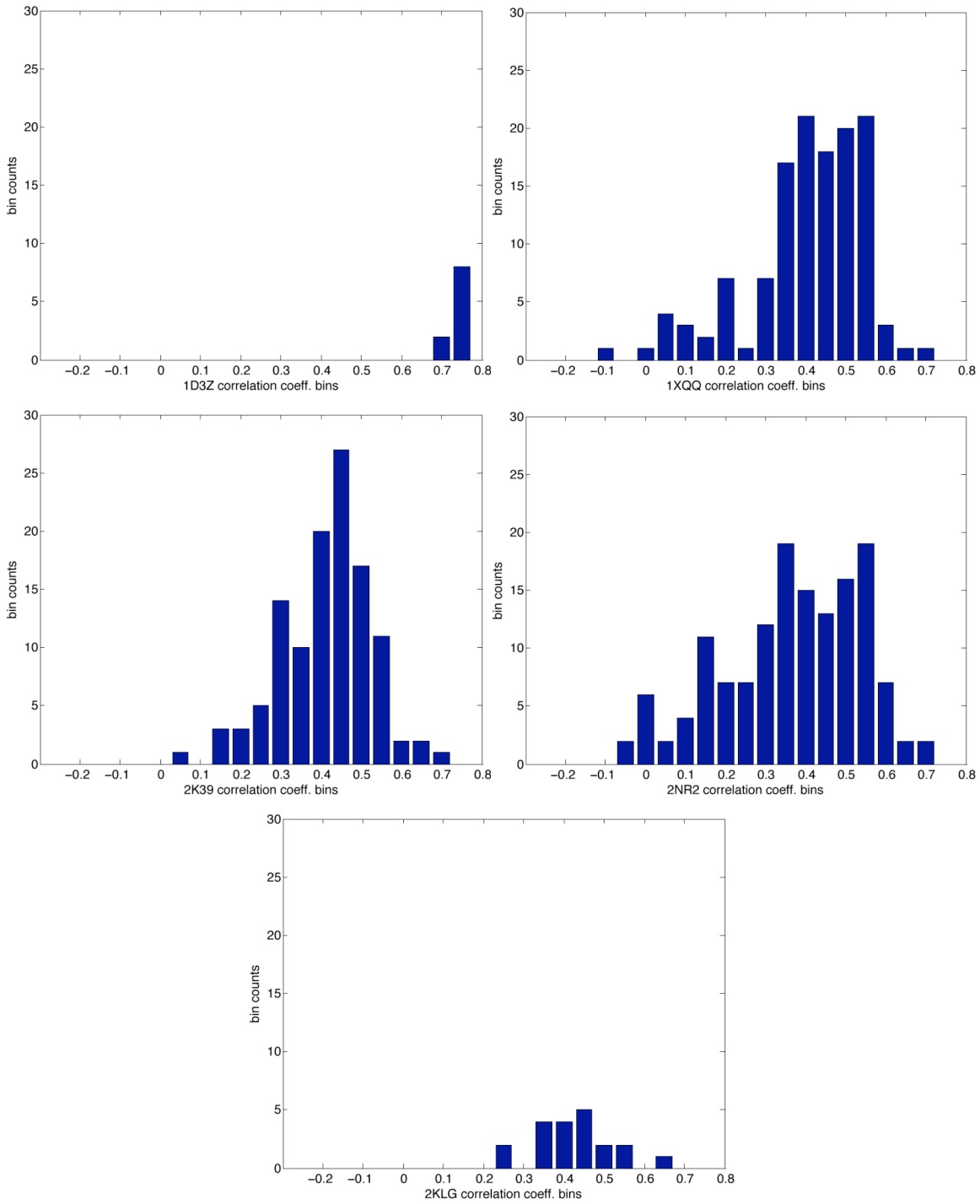

Figure 5.10. Histogram of correlation coefficients between experimental K rates and calculated K_{rig} rates within several ubiquitin NMR structures.
NMR ensembles are , moving from left to right, 1D3Z, 1XQQ, 2K39, 2NR2, and 2KLG.

	Mean Q	Stdev Q
1D3Z (10 conformers)	0.79	0.15
1XQQ (128 conformers)	0.75	0.20
2K39 (116 conformers)	0.78	0.21
2NR2 (144 conformers)	0.75	0.20
2KLG (20 conformers)	0.78	0.22

Table 5.4. Ubiquitin NMR structure mean Q and standard deviation of Q.

The correspondence of K dynamics with NH amide backbone dynamics provides information on the extent to which these phenomena are probing the same types of backbone flexibility. Figure 5.11 displays Q values from the 1D3Z ensemble along with NH S^2 amide backbone parameters for ubiquitin as measured by classical ^{15}N relaxation methodology. Lysines are highlighted in red. A reasonable correspondence is seen, with Q dropping with S^2 at the beta turn and sheet from roughly residues 6-15, near K63, and at the C-terminus. Q senses motion in several areas that do not experience particular amide backbone disorder, particularly in the solvent-exposed face of the amide helix region from residues 23-34. The correlation between Q and S^2 is present but not strong, with a value of 0.32 ± 0.04 (mean *p*-value 0.024) across all conformers of 1D3Z.

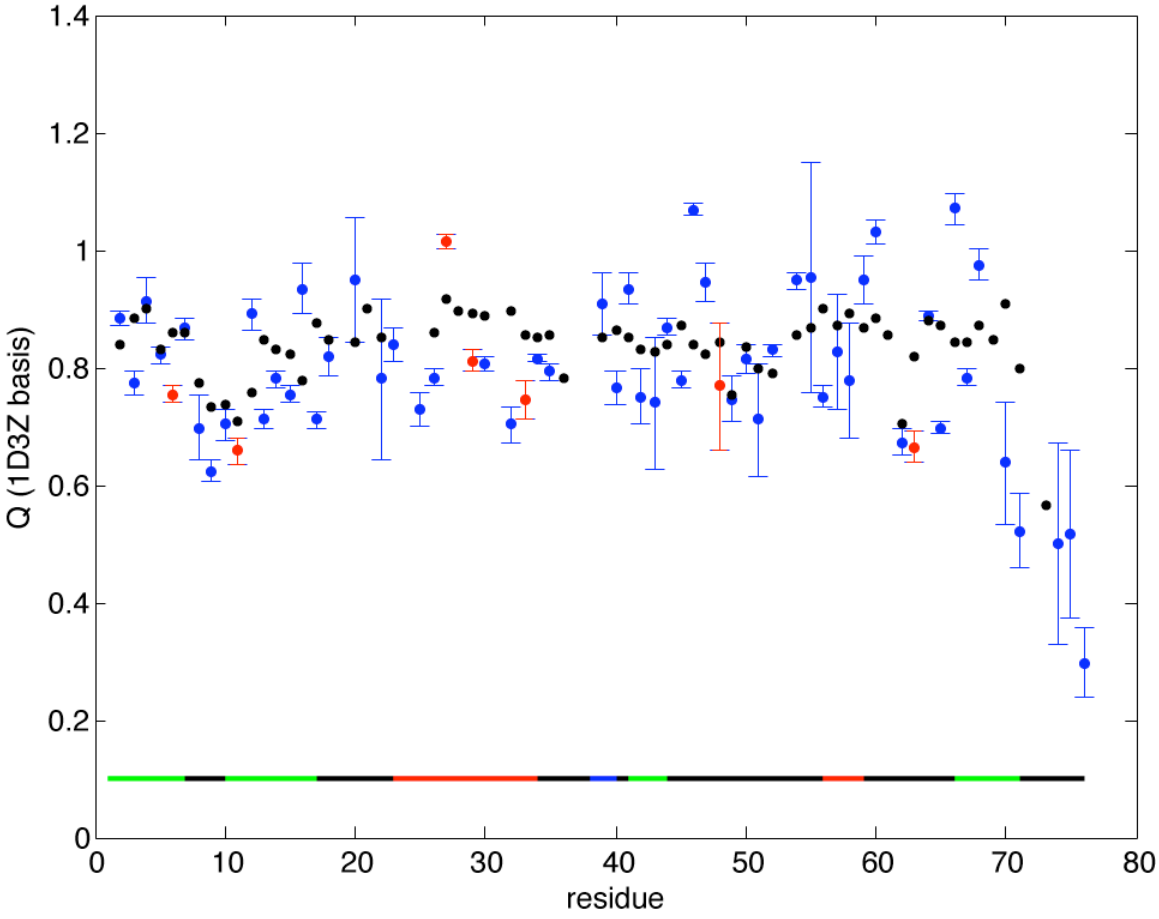

Figure 5.11. Q values from the 1D3Z ensemble (blue) along with NH S^2 amide backbone parameters for ubiquitin as measured by classical ^{15}N relaxation methodology (black).
Lysines are highlighted in red. The line at the bottom of the graph indicates protein secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, blue indicating 3_{10}-helix structure, and black indicating loop structure.

5.4 Ubiquitin discussion

NMR structure determination by homonuclear proton-proton NOEs[43] is a process that operates based on a set of constrained upper bounds on interproton distances. NOE crosspeaks are divided into strong, medium, and weak intensities, with the strength of crosspeak intensity inversely proportional to the distance between the protons represented

by the NOE crosspeaks. The structure is refined based on the constraints provided by these crosspeak intensities, plus information from scalar couplings, hydrogen bonding as determined by amide proton exchange experiments, and other covalent constraints such as the location of disulfide and metal linkages, where known.

The NMR ensembles examined in this chapter are generated from a variety of different technical approaches to improving the accuracy and impact of NOE structure determination in the modern era of NMR. These approaches include additional restraints derived from residual dipolar coupling measurements (1D3Z, 2K39), use of molecular dynamics and NMR dynamics information (1XQQ, 2K39, 2NR2), and use of soluble paramagnetic relaxation enhancers (2KLG).

The K rates derived from the η_zK experiment consist of measurements of the sum total of homonuclear proton NOE relaxation rates experienced by each amide proton. These rates depend on proton-proton distance and are modulated by interproton dynamics on the ps-ns timescale, and can be calculated from any protonated protein structure.

The summation of amide proton-neighbor proton NOE relaxation rates into experimental K rates means that protein structure cannot practically be derived de novo from K experiments, since there is no way to differentiate among the contributions of individual neighbor protons to the total K rate.

We can, however, go in the opposite direction. We can calculate theoretical K_{rig} rates from any protein structure conformer we generate in the course of NOE structural refinement and compare this theoretical K_{rig} to experimentally measured K rates. At the most basic level, we can look for metrics such as strong values of correlation between K and K_{rig} rates. These indicate that the structure proposed by the protein structure conformer does a good job of explaining the experimental K rates. Poorer values of correlation indicate either structural inaccuracy, extraordinary situations of dynamic variability, or significant unidentified amide proton exchange.

At this basic level, K_{rig} rates can also be used to measure the coherency of an NMR structural ensemble. NOE structural ensembles can therefore be evaluated in terms of their mutual consensus on K_{rig} rates as well as the agreement of this K_{rig} consensus with the pattern of K rates exhibited by experiment. A good structural ensemble should

meet both criteria. The figures also show that experimental K rates are always smaller than the predicted ones, indicating structural overpacking, significant dynamics, or both.

At a higher level of refinement, experimental K rates provide interesting candidates for an additional constraint on NOE structure refinement. Figure 5.11 compares experimental K rates with K_{rig} rates calculated from all conformers of 1D3Z, 1XQQ, 2K39, 2NR2, and 2KLG respectively. This figure suggests that this constraint will be particularly useful, since K rates have a measurement error significantly smaller than the variation in most NOE structural ensembles.

Use of K rates as NOE refinement constraints must be undertaken with caution due to the inevitable presence of proton-proton dynamics and their modulation of K rates. The dynamic-ensemble refinement (DER) technique of Lindorff-Larsen et al., which combines MD simulations with simulated annealing-based minimization[68; 130; 131], provides a natural starting point for this work. Proton-proton dynamics have not been extensively studied in the context of molecular dynamics simulations and the existing literature on the subject is sparse due to a lack of experimental information.[45; 48; 49; 50] Refinement of molecular dynamics programs using K rates promises enough of an advance in our understanding of proteins to motivate further work in this regard.

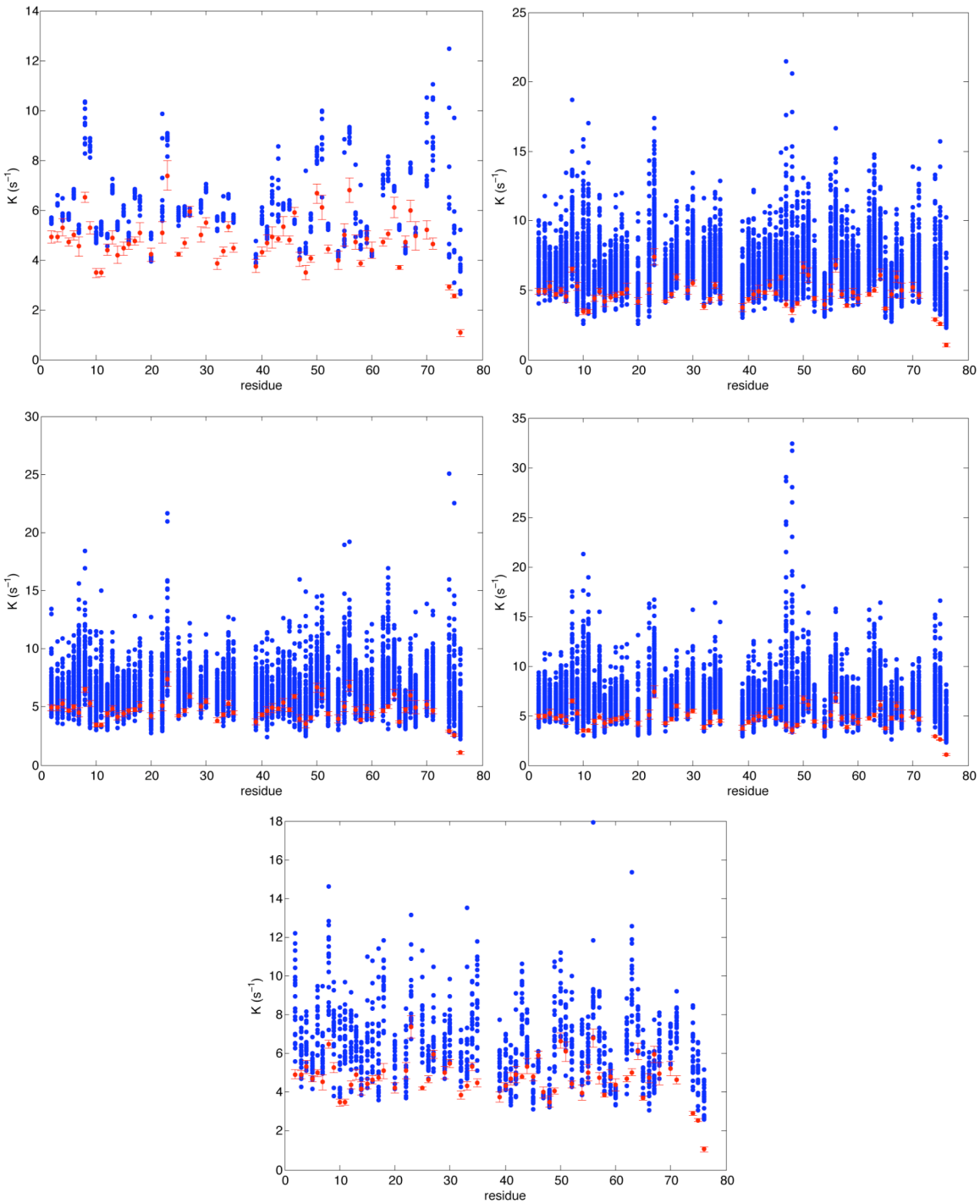

Figure 5.12. Experimental K rates compared with K_{rig} rates calculated from all conformers of several NMR structures.
Experimental K rates are in red, with K_{rig} rates in blue. NMR ensembles are , moving from left to right, 1D3Z, 1XQQ, 2K39, 2NR2, and 2KLG.

CHAPTER 6: PROTON-PROTON DYNAMICS OF HSC70 NUCLEOTIDE-BINDING DOMAIN

6.1 Hsc70 NBD introduction

The Hsp70 family of molecular chaperones is present in all organisms. Constitutively expressed for numerous cellular functions and integral to the cell stress response, Hsp70s are literally vital to life on earth. Members of the Hsp70 protein family are implicated in numerous disease conditions, including cancer. A thriving literature (695 publications and 28 review articles in 2009 alone) exists on all aspects of the Hsp70 family and related folding processes, with multiple avenues of investigation open into Hsp70 pathology, regulation, and participation in proteomic networks.

Our chosen exemplar of the Hsp70 family is the mammalian heat shock cognate protein Hsc70 (UniProt/SwissProt P19120). Originally identified as an ATPase with clathrin-uncoating ATPase activity[132], this constitutively-expressed protein is present in the cytosol under normal cellular conditions and is particularly strongly expressed in growing cells, where it comprises about 1% of the protein present.[133] Early efforts relied on purification of Hsc70 from large quantities of fresh calf brains[132]; thankfully, modern recombinant protein synthesis techniques (*vide infra*) allow Hsc70 to be prepared in a much less grisly manner.

Like many of its cousins within the Hsp70 family, Hsc70 acts as a constitutive chaperone, binding to unfolded hydrophobic regions of polypeptides and stabilizing them against aggregation in the unfolded state. Since unfolded polypeptides cannot refold while Hsc70 is attached to them, Hsc70's binding to its polypeptide substrates must be switched between high-affinity and low-affinity states in order to cycle Hsc70 on and off in a timely fashion.

This task is accomplished by allosteric coordination between the 44-kD N-terminal

nucleotide binding domain (NBD) and the 27-kD substrate binding domain (SBD) of Hsc70. The nucleotide state of the NBD regulates the affinity displayed by the SBD: when the NBD has ATP bound, the SBD is placed into a low-affinity 'open' state capable of binding protein. Substrate binding at the SBD impels the NBD to hydrolyze ATP to ADP and Pi (a process greatly accelerated by the presence of J protein cochaperones). In the ADP.Pi state, the NBD causes the SBD to enter a high-affinity 'closed' state which is tightly bound to its hydrophobic polypeptide substrate. The NBD nucleotide state must be reset to ATP in order to place SBD back into its low-affinity state for dissociation from polypeptide. This reset process is slow in isolation. Nucleotide exchange factors such as Bag accelerate the rate of nucleotide release and NBD state reset.

The work described in this chapter focuses on a protein construct comprising the bovine Hsc70 NBD. This construct encompasses residues 1-386 of the 650-residue complete Hsc70 protein. The presence of a 44 kD protease digestion fragment with ATPase activity was identified in 1987.[134] The NBD domain was subsequently crystallized in 1988[135] and its structure was published at 1.93 Å resolution as 3HSC in 1990.[73] The NBD retains the catalytic characteristics (K_m and k_{cat}) of ATP hydrolysis displayed by the full-length protein in the peptide-bound state, indicating that it is a functionally intact piece of allosteric machinery.

The Hsc70 NBD construct consists of four subdomains surrounding a central nucleotide binding cleft in an actin-like fold.[136] These subdomains are referred to as IB, IA, IIA and IIB, with the N- and C-terminal (leading to the hydrophobic linker region and SBD) of the construct both located in the IA domain. Extensive efforts at site-directed mutagenesis have been carried out in order to characterize the essential residues for ATP hydrolysis and allosteric communication.[74; 137; 138; 139; 140]

The major biological focus of the Zuiderweg laboratory consists of research into the structural biology of Hsp70 regulation. This research laid the groundwork of purification protocols and backbone assignments required for the efforts described in this chapter. However, the overarching goal of this thesis work has been the development of experimental NMR relaxation methodology capable of being applied to large proteins on the size scale of the Hsc70 NBD. The remainder of this chapter will therefore focus on Hsc70 NBD primarily as a physical protein object, although the system's biological

context will be considered where necessary and relevant.

6.2 NMR spectroscopy of the Hsc70 NBD

The 44-kD Hsc70 NBD is a very large protein by the standards of common NMR relaxation studies, although it is average-sized in terms of the proteome (average human protein size 52 kD, average *E. coli* protein size 35 kD). Quantitative NMR relaxation studies require excellent resolution and signal-to-noise (S/N). Our decisions and methodology will be guided by a quest to achieve comfortable levels of both in the face of the large-protein relaxation regime.

The large-protein relaxation regime is characterized by rapid relaxation processes which have deleterious effects on both resolution and S/N. Relaxation affects S/N by rapidly returning prepared (non-equilibrium) magnetization to its equilibrium state, causing signal attenuation. Relaxation affects resolution by causing severe line broadening during periods where magnetization is in the transverse plane, such as during coherence transfer, chemical shift labeling, and any transverse relaxation elements in the pulse sequence.

Homonuclear proton-proton relaxation processes operating at zero-frequency spectral densities are particularly effective in the large-protein relaxation regime. One reason for this effectiveness is due to the previously detailed direct scaling of zero-frequency proton-proton relaxation with the global rotational correlation time of the protein. The other reason for this effectiveness is due to the increased number and density of protons in larger proteins. All other things being equal, larger proteins are likely to have a smaller ratio of surface area to volume than smaller proteins, leading to more of the backbone being buried within the body of the protein. In the body of the protein, the backbone is exposed to a dense network of relaxing protons from all directions. In contrast, at the surface of the protein, the backbone is exposed to a dense network of relaxing protons only on its "protein side", while receiving comparatively less relaxation from its "solvent side".

In small proteins such as calmodulin and ubiquitin, zero-frequency proton-proton

relaxation processes are less effective at driving rapid relaxation that negatively affects S/N and resolution. The same cannot be said of the Hsc70 NBD, as is graphically demonstrated by the example η_zK relaxation spectrum in Figure 6.1. This spectrum of 0.3 mM protonated ^{15}N-labeled Hsc70 NBD exhibits poor S/N and resolution, with many residues indistinguishable from one another even after a large number of scans. Since this spectrum is a first time point from an auto-relaxation experiment (I-type), subsequent spectra will be inferior. Spectra of this quality are not sound bases for quantitative NMR.

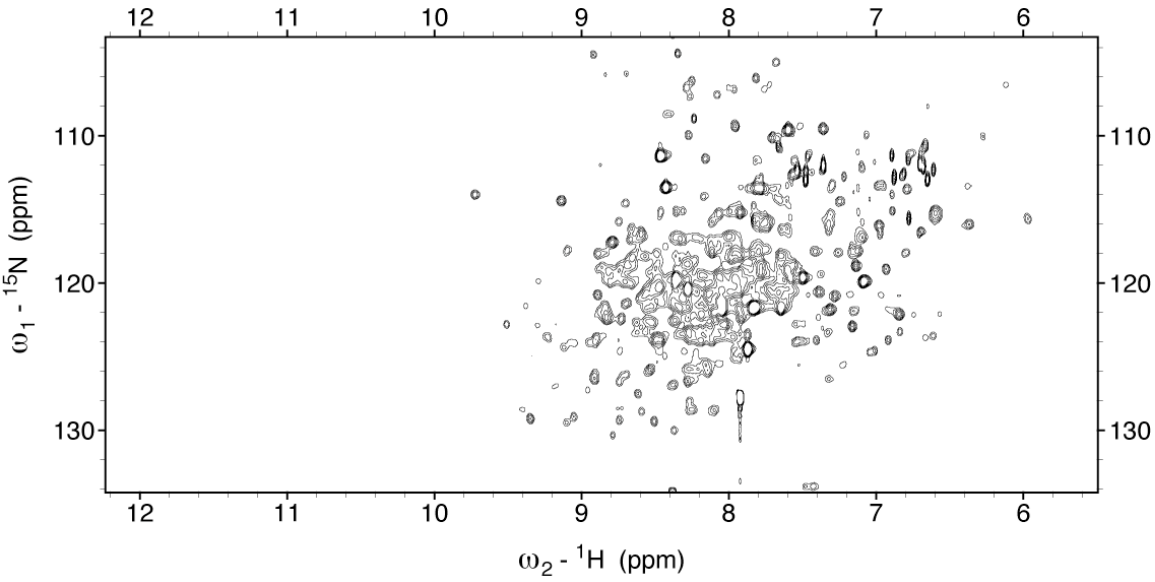

Figure 6.1. Representative spectrum of 0.3 mM protonated ^{15}N-labeled Hsc70 NBD spectra for an I-type η_zK experiment (NzHa to NzHa) at 10 ms relaxation time. Spectrum acquired with 2048 x 60 complex points, with the carrier set at 4.725 and 118.751 ppm respectively. 128 transients are collected, with a recycle delay of 0.8 s.

Simple consumption of additional experimental time will not remedy the problem. Though additional scans can add S/N by coadding signal at a faster rate than noise, additional scans cannot improve resolution, which is determined by the system and spectrometer. Additional scans in a poorly resolved spectrum will result in intense, broad peaks.

This situation dictates protein deuteration. Deuteration has a long history in NMR as the standard response to problems of poor resolution and S/N caused by zero-quantum proton-proton relaxation. Protons with spin-1/2 and Larmor frequency ω_H are replaced

by deuterons with spin-1 and Larmor frequency ω_D, effectively eliminating them from the proton relaxation network and neutering their ability to take part in proton-proton relaxation. The resulting greatly diminished relaxation behavior yields greatly improved spectra.

Figures 6.2-6.5 contain example η_zK relaxation spectra of 0.4 mM ^{2}H,^{13}C,^{15}N-labeled Hsc70 NBD in which all exchangeable sites (protons bound to N, O and S atoms) have been back-exchanged with protons, a level of deuteration referred to as perdeuteration. The spectra illustrated are representative of Hsc70 NBD spectra for I, II, III and IV-type experiments at 400 ms relaxation time. Each spectrum is acquired in about two and a half hours. The quality of relaxation spectra is inevitably inferior since they (by definition) decay from an initial point of maximum intensity, or at least suffer from relaxation of signal intensity even while undergoing a cross-relaxation buildup process. The impressive spectral quality of mid-relaxation data for this 44 kD protein is a testament to the effectiveness of TROSY-HSQC spectroscopy in carefully perdeuterated proteins.

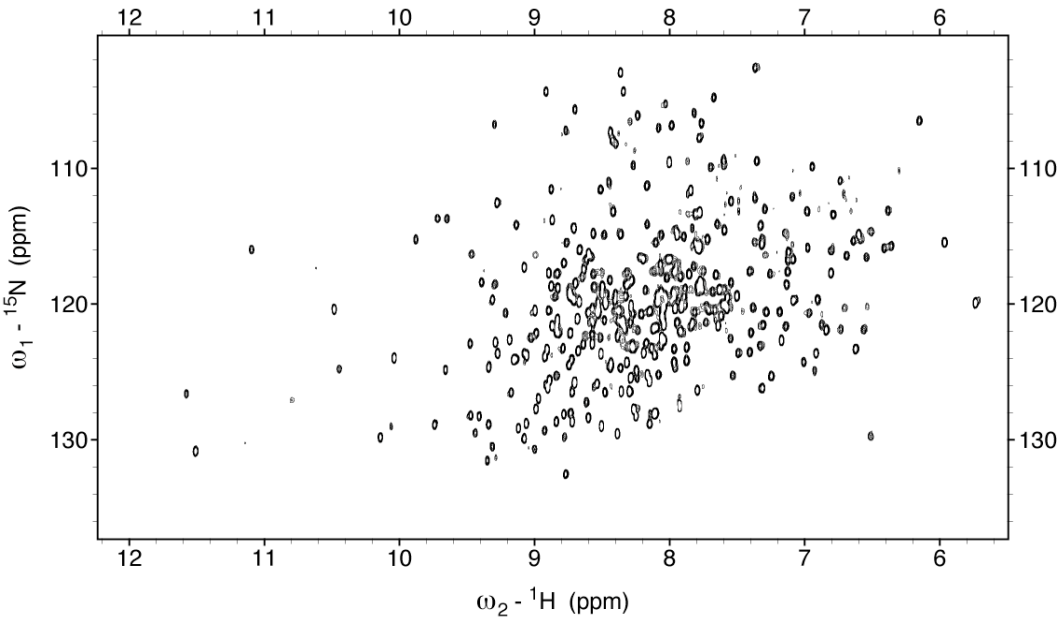

Figure 6.2. Representative Hsc70 NBD I-type (NzHa to NzHa) spectra at 400 ms relaxation time.
This spectrum is acquired with 2048 x 150 complex points, with the carrier set at 4.725 and 118.751 ppm respectively. 16 transients are collected, with a recycle delay of 2.25 s. Collection takes about two and a half hours.

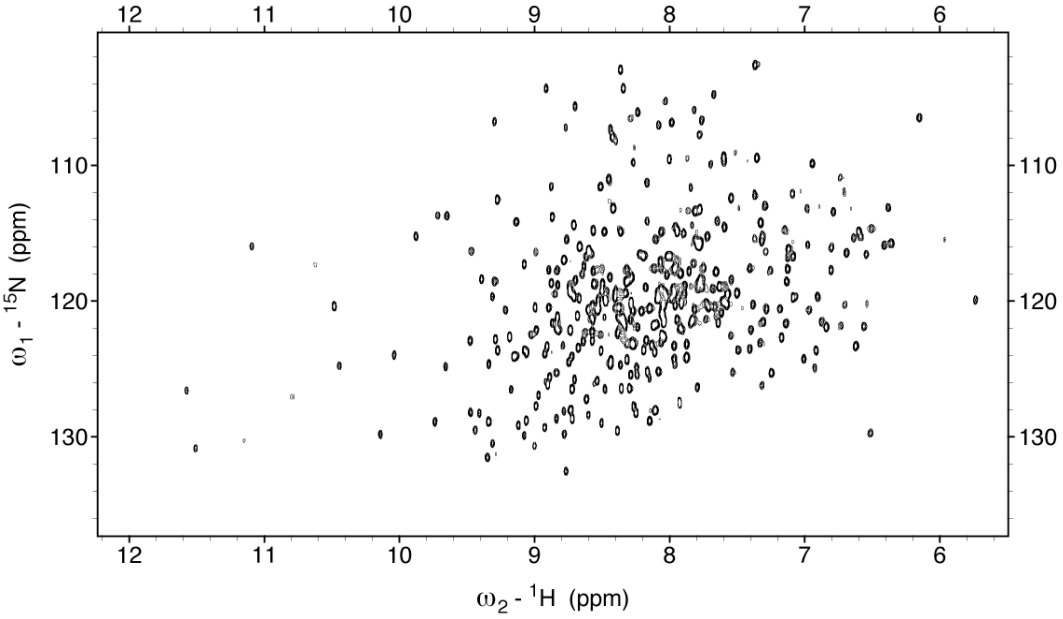

Figure 6.3. Representative Hsc70 NBD II-type (NzHa to NzHb) spectra at 400 ms relaxation time.
This spectrum is acquired with 2048 x 150 complex points, with the carrier set at 4.725 and 118.751 ppm respectively. 16 transients are collected, with a recycle delay of 2.25 s. Collection takes about two and a half hours.

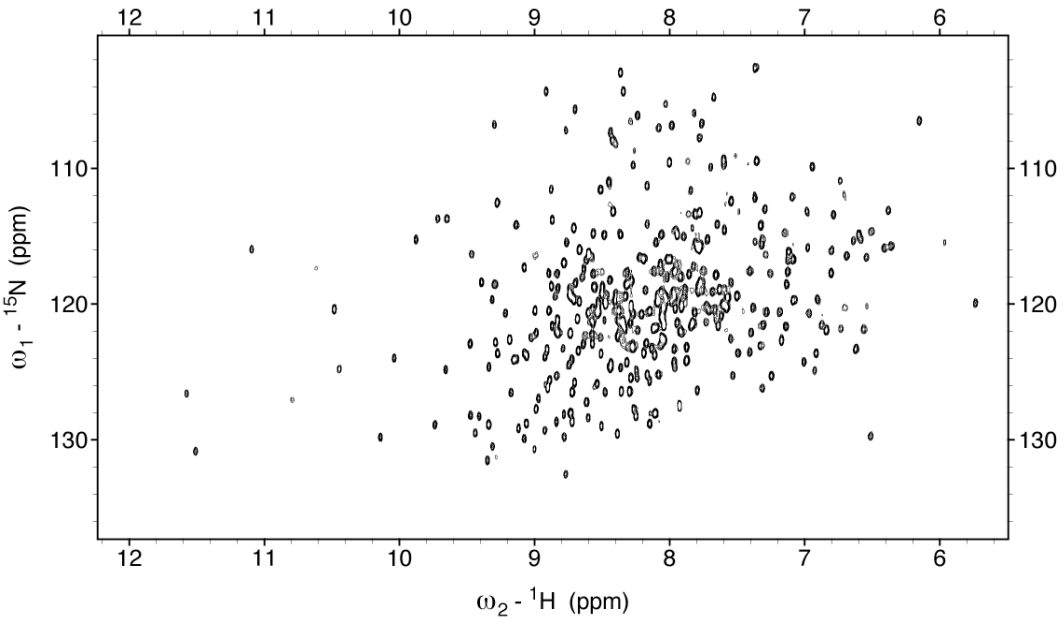

Figure 6.4. Representative Hsc70 NBD III-type (NzHb to NzHa) spectra at 400 ms relaxation time.
This spectrum is acquired with 2048 x 150 complex points, with the carrier set at 4.725 and 118.751 ppm respectively. 16 transients are collected, with a recycle delay of 2.25 s. Collection takes about two and a half hours.

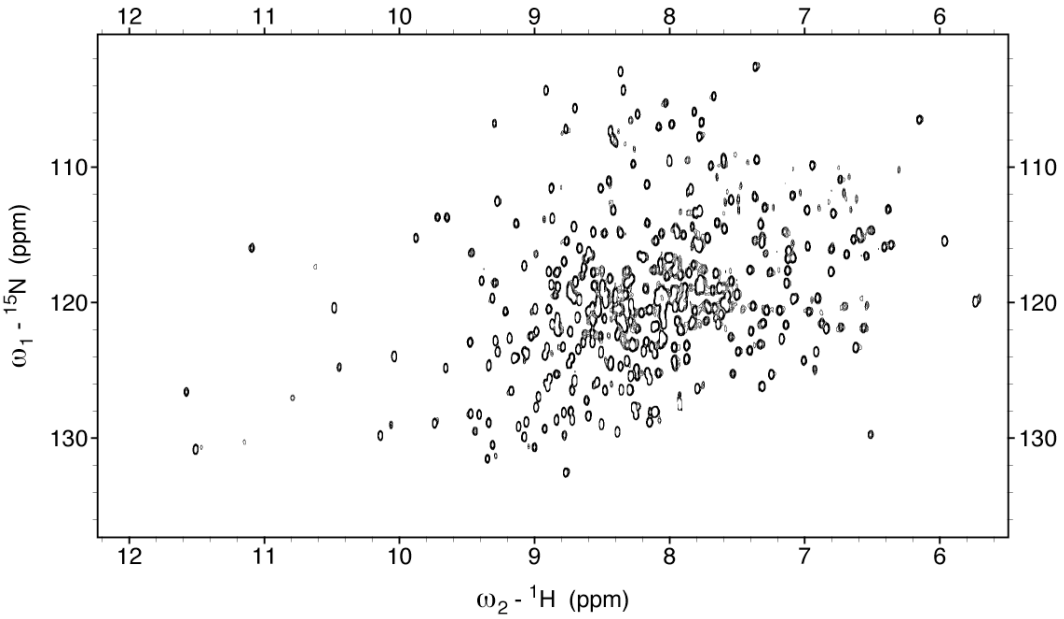

Figure 6.5. Representative Hsc70 NBD IV-type (NzHb to NzHb) spectra at 400 ms relaxation time.
This spectrum is acquired with 2048 x 150 complex points, with the carrier set at 4.725 and 118.751 ppm respectively. 16 transients are collected with a recycle delay of 2.25 s. Collection takes about two and a half hours.

The resolution-driven requirement for deuterated samples raises the obvious question: if we are studying proton-proton relaxation, and we remove most of the protons, what is left to observe in terms of relaxation phenomena? The obliteration of the proton network in the spectrum displayed above leaves only exchangeable protons to relax each other. If less drastic deuteration measures are undertaken, such as growth in 50% deuterated media, the experimentalist immediately encounters problems with the vagaries of bacterial biosynthesis. Deuterons and protons are shuffled around different metabolites in a highly amino-acid-dependent manner, leading to extreme difficulties in predicting where the protons end up. The result is that K_{rig} rates become impractical to calculate, because one has no idea which protons are actually protons and which protons are deuterons. For this reason, nonselective partial deuteration is not appropriate for $\eta_z K$ experiments.

A more promising avenue of approach lies in selective protonation of selected amino acids via the introduction of protonated amino acids into a deuterated medium. This approach is expensive and may require preparation of several different samples, each with different amino acids selected for protonation. However, protonation will be well-localized in such systems, and it will be practical to calculate K_{rig} rates for selectively-protonated proteins. These experiments may be of great interest in the future for the determination of order parameters between selectively protonated residues' amide protons and their sidechains, or between neighbor residues' amide protons and a neighboring selectively protonated residue's side chain.

The work described in this chapter is aimed at validation of the $\eta_z K$ experimental procedure under large protein conditions. For this reason, we focus on the simplest case of protein perdeuteration and subsequent back-exchange, leaving us with an Hsc70 NBD that is completely deuterated except for exchangeable proton sites. Zero-quantum proton-proton relaxation in this system is supplied entirely by protons bound to nitrogen, oxygen and sulfur atoms. As will be seen, these interactions are still sufficient to relax the system and provide interesting data about the backbone's motion with respect to itself.

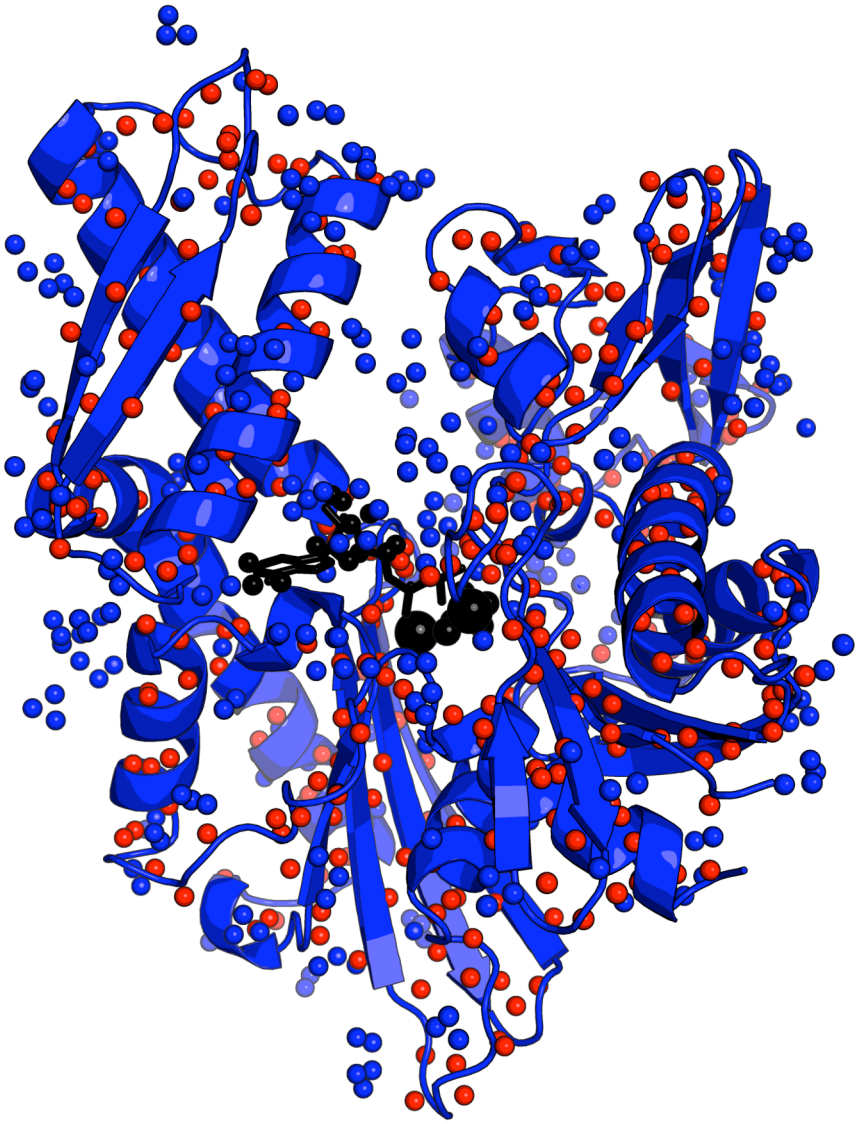

Figure 6.6. Structure of bovine Hsc70 NBD.
The structure portrayed is 1HPM (1.75 Å).[137] Small spheres indicate proton locations, with red spheres denoting amide protons, blue spheres denoting all other exchangeable protons, and black spheres denoting protons attached to ADP. Large black spheres denote bound K^+, Mg^{2+}, and PO_4^-. Sphere scale is set to half the van der Waals radius for purposes of clarity.

6.3 Hsc70 NBD materials and methods: protein synthesis

The $^{2}H,^{13}C,^{15}N$-labeled, proton back-exchanged *Bos taurus* Hsc70 NBD construct (*Bos taurus* Hsc70 1-386) is overexpressed, purified, amide proton back-exchanged, and

prepared for NMR spectroscopy according to protocols developed by A. Kurochkin and Y. Zhang.

Novagen BL21(DE3) *E. coli* cells are transformed with a pET T7 expression system plasmid containing wild-type *Bos taurus* Hsc70 NBD (*Bos taurus* Hsc70 1-386) and ampicillin resistance genes. The plasmid was kindly donated by D. B. McKay (Stanford University). Transformed cells are grown according to Novagen-provided protocols, plated on agar medium (pH 7.0) containing ampicillin (2 mg/L), and incubated for ~12-16 hr at 37 C.

Protein perdeuteration requires growth in perdeuterated media. Perdeuterated conditions present a challenge to cell growth. Transfer through several stages of preparatory media before the final inoculation of deuterated media allows selection of cell phenotypes with good deuterated growth characteristics, as well as buildup of sufficient inoculation biomass to ensure entry of log phase in deuterated media within a reasonable time.

A starter culture of 2 mL LB+ampicillin (2 mg/L ampicillin) is inoculated with freshly transformed and plated BL21 (DE3) cells and grown at 37 C and 250 rpm for 3-4 hours. 50 uL of this starter culture is then used to inoculate a flask containing 25 mL LB+ampicillin (1:500 dilution) in order to select the fastest-growing phenotype from the starter culture. The 25 mL flask is grown at 37 C and 250 rpm until it reaches OD 0.3-0.4 at 600 nm (Ultrospec 10, Amersham Biosciences). The 25 mL culture is then spun down (2000 RPM for 15 minutes). The supernatant is discarded and the cell precipitate is used to inoculate a 50 mL culture of H_2O M9 minimal media (pH 7.2) with 2.5 g/L glucose, 1 g/L NH_4Cl, 5 mg thiamine and 50 mg/L ampicillin. The 50 mL M9 H2O culture is grown at 37 C and 250 rpm until it reaches OD 0.4-0.5 at 600 nm. It is then spun down (2000 RPM for 15 minutes). The supernatant is again discarded and the cell precipitate is used to inoculate a 1 L culture of triple-labeled D_2O M9 minimal media containing 2.5 g/L $^{2}H,^{13}C$-labeled glucose, 1 g/L $^{15}NH_4Cl$, 5 mg thiamine and 50 mg/L ampicillin pH 7.2 (pD 7.6). In order to increase the deuterium fraction of the media, D_2O M9 minimal media is prepared using heptahydrous $MgSO_4$ that has been dissolved in D_2O, lyophilized and redissolved in D_2O in order to replace coordinated H_2O molecules with coordinated D_2O molecules. The 1 L D2O culture is grown at 37 C and 250 rpm and induced with

500 uM IPTG at OD 0.6 (600 nM). Saturation is observed at OD ~1.2 (600 nm) with a characteristic doubling time of ~2 hours 45 minutes and 3 doubling times' worth of growth after induction.

The time course of the growth of 1 L D_2O M9 culture is detailed in Table 6.1, with the first OD data point taken 12 hours after inoculation.

Elapsed time since 1 L inoculation (min)	OD (600 nm)
730	0.31
800	0.42
840	0.49
875	0.57*
1040	0.92
1130	1.04
1170	1.07
1235	1.08
1410	1.17†

Table 6.1. Optical density (600 nm) measurements of E. coli growth in 1 L D_2O M9 culture.
*** indicates time of induction by addition of IPTG (*lac* repressor system)**
† indicates termination of cell growth and spin down of culture

Upon reaching saturation the 1 L D_2O culture is spun down (5000 RPM for 30 min), the supernatant is discarded, and the cell precipitate is resuspended in 20-30 mL 50 mM Tris·HCl/2 mM EDTA (pH 8.0). An appropriate number of cOmplete, Mini protease inhibitor cocktail tablets (Roche) are added to hinder proteolysis during the purification process. The cells may optionally be incubated with lysozyme (0.3 mg/mL) on ice for 30 min in order to digest the outer cell membrane. Following the optional lysozyme digestion step, the cells are microfluidized and the resulting crude lysate is spun at 30,000 g for 20 minutes. The resulting supernatant, which contains the soluble protein fraction, is immediately loaded onto a DEAE52 anion exchange column (Whatman) which acts as a negative filter. The Hsc70 NBD is a basic protein and passes directly through the negatively charged column along with a handful of minor protein fractions. The remainder of the soluble protein fraction binds to the column. The flowthrough from the DEAE52 column load is collected, along with several washes in

buffer. This flowthrough contains the Hsc70 NBD and the minor protein fractions mentioned previously.

Following the anion exchange column step, the column load flowthrough containing the desired protein fraction is dialyzed against 20 mM HEPES/25 mM KCl/10 mM EDTA (pH 7.0) in order to chelate all available Mg^{2+} ions. Mg^{2+} is required for nucleotide binding to the Hsc70 NBD nucleotide-binding cleft. The addition of this high concentration of EDTA empties the nucleotide binding pocket by scavenging available Mg^{2+} and interfering with nucleotide binding, allowing ATP-affinity chromatography to proceed more effectively. Upon completion of dialysis 25 mM MgCl2 is added to the solution and the solution is spun at 10000g for 20 min. Mg^{2+} bound to EDTA precipitates out (2 Mg^{2+} per EDTA molecule) and the resulting buffer composition is 20 mM HEPES/25 mM KCl/5 mM MgCl2 (pH 7.0). The supernatant is then loaded on an ATP-agarose affinity column (Sigma), washed with high salt (1 M NaCl) to elute non-specifically bound proteins, and eluted into 15 6 mL fractions with a buffer containing 3 mM ADP and 3 mM KP_i (pH 7). The fractions are analyzed for their protein contents with SDS-PAGE gel electrophoresis and those fractions containing Hsc70 NBD (44 kD) are combined. 0.05% NaN_3 may be added at this point to act as a bacteriostat. A typical yield from this process is 15-20 mg purified Hsc70 NBD.

Following purification and immediately prior to buffer exchange and concentration, the protein must be unfolded and refolded in H_2O in order to ensure uniform amide protonation via exchange with water protons. This process is known as amide proton back-exchange. It is accomplished by introducing the desired amount of purified protein (buffer conditions as previous) into 6 M guanidinium HCl (GuHCl). This unfolds the protein, causing all amide deuterons remaining from growth in D_2O media to immediately exchange with water protons.

The GuHCl-unfolded protein solution is then concentrated to 10 mg/mL and separated into many small aliquots of 20 uL each (Eppendorf tubes work well here). The protein is then refolded by rapid dilution in NMR buffer containing 20 mM KP_i, 25 mM KCl, 5 mM MgCl, 10 mM ADP, 0.05% NaN_3, and 5% D2O by volume (pH 7.2). The rapid dilution process is accomplished by pipetting 1 mL of NMR buffer into each Eppendorf tube in a single smooth motion. Each tube is then gently pipetted several

times to mix and all Eppendorfs are combined in a single volume. This combined volume is allowed to stand at room temperature for one hour.

The purified protein is then throughly buffer-exchanged with additional NMR buffer to remove residual GuHCl and concentrated by filtration through a Centricon-15 filter with a 10 kD molecular weight cutoff (Amicon). Hsc70 NBD concentrations of 0.4 mM in 500 uL (assayed by BCA and NMR signal intensity) have been achieved without major precipitation of protein and are stable for months.

6.4 Hsc70 NBD materials and methods: spectroscopy

All NMR spectroscopy on Hsc70 NBD is carried out on a Varian iNOVA 800 spectrometer equipped with a cryoprobe. All relaxation experiments consist of 2D $^{1}H_{N}$-^{15}N TROSY-HSQC spectra acquired at several different relaxation delays. Perdeuteration leads to long proton T_1 times due to the paucity of proton-proton longitudinal relaxation sources. The recycle delay is consequently set to 2.25 s.

η_zK experiments are conducted with 2048 complex points in the proton dimension and 150 complex points in the nitrogen dimension, with the carrier set at 4.725 and 118.751 ppm respectively. The experiment is conducted with 11 values of the relaxation delay τ: 10, 10, 50, 100, 150, 200, 250, 250, 300, 350, 400, and 500 ms. Perdeuteration removes most of the proton relaxation network. Although the remaining K rates are boosted by Hsc70 NBD's slow global tumbling, this boost is not enough to offset the depletion of the proton network, and total K rates run at low speeds. The relaxation delay is therefore longer for η_zK experiments performed on Hsc70 NBD than for the proteins previously detailed in this work.

η_{xy} experiments are performed at values of N = 0, 1, and 2 for the relaxation delay τ, according to the relation $2\tau = N/J^{1}_{NH}$. Since other protons do not contribute to the interaction generating the η_{xy} relaxation rate, perdeuteration does not affect η_{xy} rates, and the relaxation delays (values of N) used for Hsc70 NBD are shorter (smaller) than those used in η_{xy} experiments on other proteins.

Simple estimations of the presence of amide proton exchange are performed by

conducting sets of two modified TROSY experiments with different pre-experimental delays. In the first part of the set, water magnetization is aligned along the z-axis prior to the initiation of the TROSY pulse sequence. In the second member of the set, water magnetization is aligned along the -z axis (inverted) prior to the initiation of the TROSY pulse sequence. Inverted water magnetization is transferred to amide protons by amide proton exchange. The corresponding change in TROSY signal intensity upon measurement indicates the presence of amide proton exchange. Perdeuteration of non-amide protons prevents inverted water magnetization from transferring throughout the molecule by NOE processes.

6.5 Hsc70 NBD results

As of this writing, 172 of 386 Hsc70 NBD residues are assigned to peaks in the NMR relaxation spectrum. This criterion is based not only on the raw presence of assignments. Although 260 unique residues are assigned, only 172 of these residues are judged sufficiently isolated and distinct from other peaks in Hsc70 NBD to be included in relaxation analysis. These 172 residues are split among 64 alpha-helical residues, 61 beta-sheet residues, and 47 loop residues, with secondary structure judged from the 2QWL X-ray structure.

Of these 260 assignments, only 103 are considered certain. The remainder may contain errors stemming from earlier assignment work in the Zuiderweg group. The Hsc70 NBD sample used for the relaxation experiments described herein has also served as the subject for assignment experiments aimed at correcting these errors and expanding the list of Hsc70 assignments. Work on this project is ongoing.

Figure 6.7 illustrates the time course of relaxation for I, II, III and IV-type experiments on residue E27 of Hsc70 NBD, a residue located in a beta-sheet region within domain IA. Even without symmetric reconversion multiplication and division in place, the characteristic behavior of η_zK experiments in large proteins is evident. Slow K rates cause auto-relaxation decay curves (relaxed primarily by K processes) to reach their 1/e point around 700 ms. Slow η_z rates result in slow divergence of I- and IV-type

experiments. Auto-relaxation decay curves display a jagged shape, the result of acquiring spectra over long periods of time in the presence of spectrometer instabilities, changes in shimming, &c. Because of the internal self-referencing of the symmetric reconversion techniques used in the η_zK experiment, this jaggedness does not result in deleterious effects on the transformed X(t) and Y(t) data.

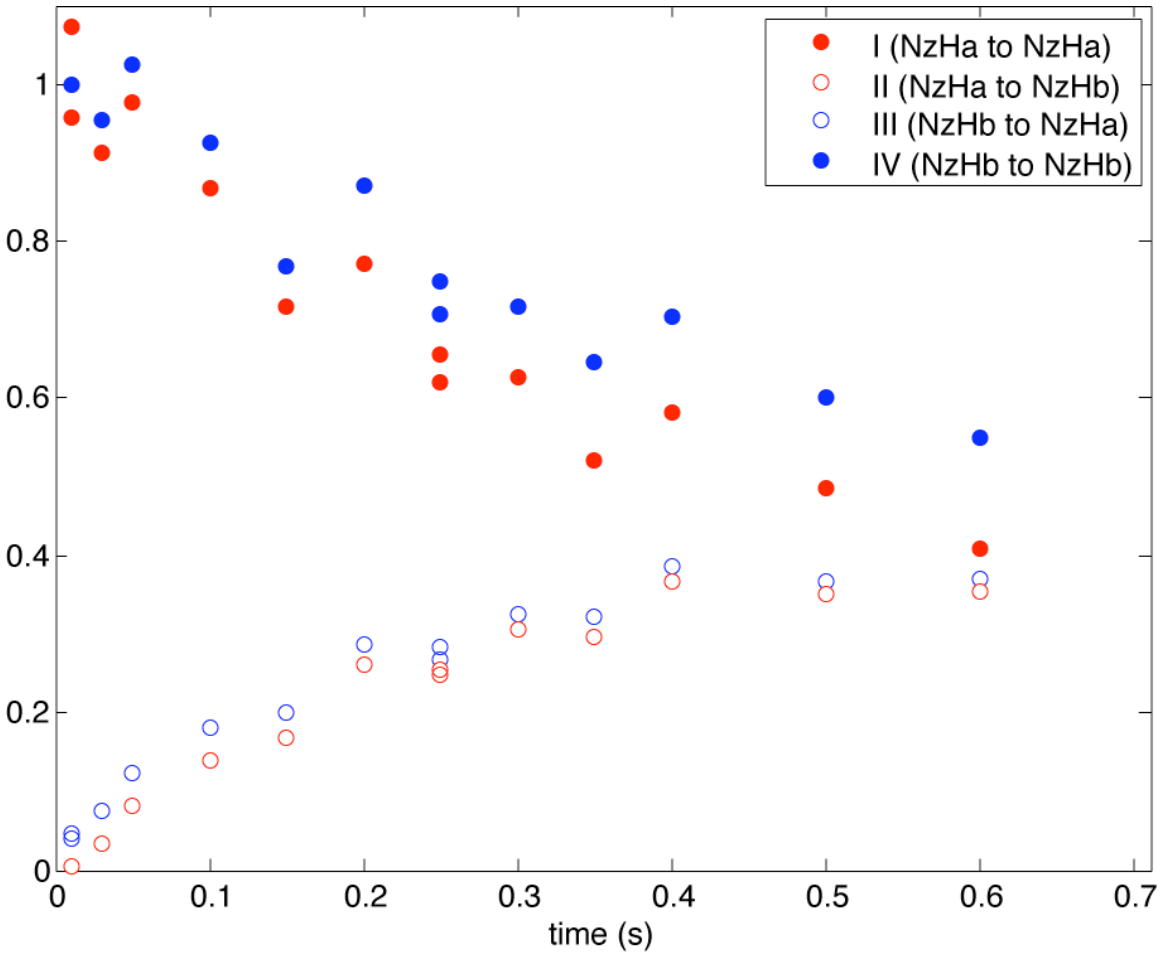

Figure 6.7. Representative time course of relaxation for I, II, III and IV-type experiments on residue E27 of Hsc70 NBD.
E27 is located in a beta-sheet region in subdomain IA.

Figure 6.8 illustrates an η_zK fit to symmetric reconversion data for residue E27 of Hsc70 NBD. Fit results at ±1 bootstrap ensemble population standard deviation (as detailed in Chapter 2 and illustrated in Figures 6.9 and 6.10) are marked by green lines.

The slow η_z and K rates characteristic of slow rotational diffusion in heavily deuterated proteins are immediately evident.

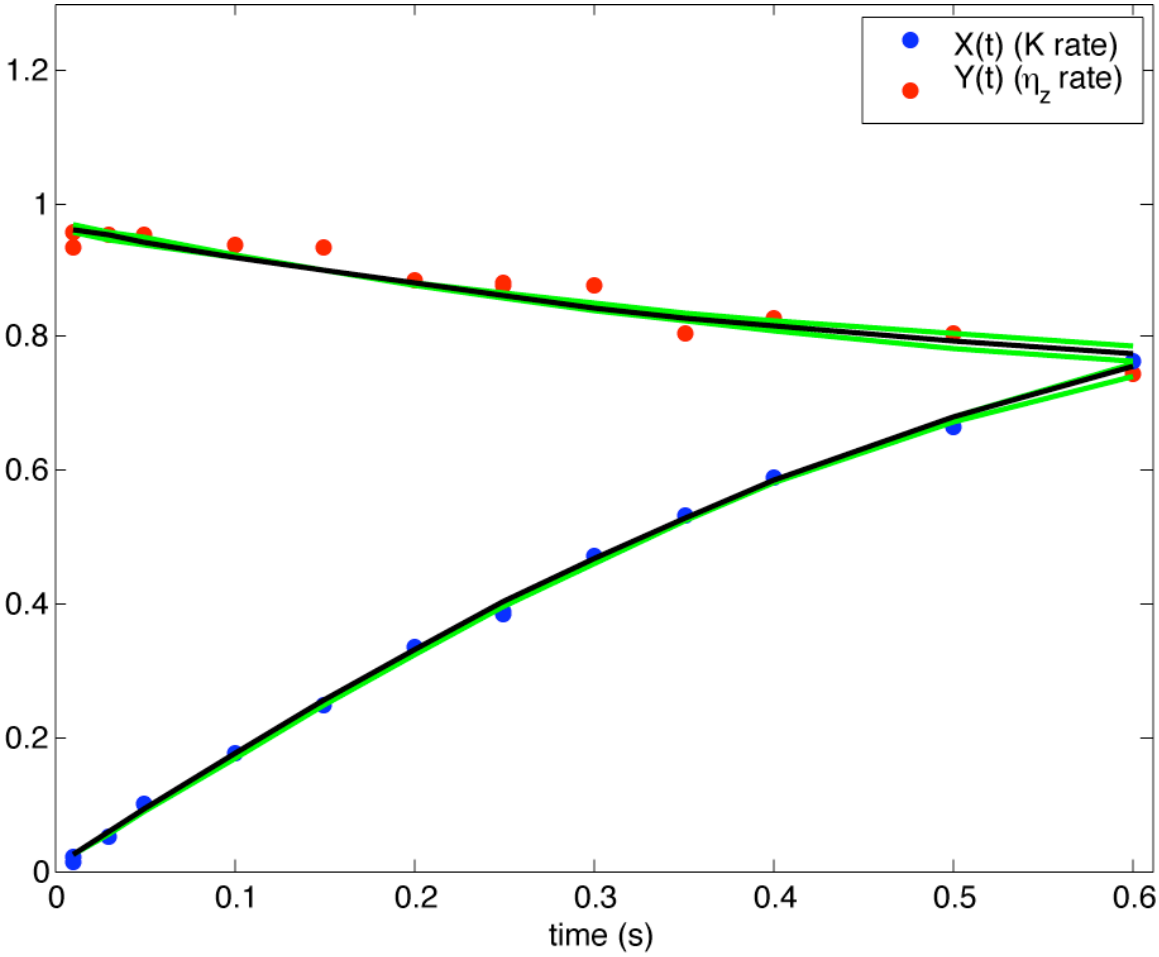

Figure 6.8. η_zK fit to symmetric reconversion data for residue E27 of Hsc70 NBD. Fit results at ±1 bootstrap ensemble population standard deviation (as detailed in Chapter 2 and illustrated in Figures 6.9 and 6.10) are marked by green lines.

Figures 6.9 and 6.10 illustrate the distribution of K and η_z fits, respectively, for residue E27 of Hsc70 NBD. This distribution is based on a bootstrap ensemble of 100 permutations of E27's symmetric reconversion data. Similar distributions are prepared for all residues and the mean and standard deviation of each distribution are used as estimators of the true values of K and η_z. An alternative is choosing confidence intervals based on the percentiles of the distribution, e.g. using the 14th and 86th members of an ensemble containing 100 members as a "1 standard deviation" error estimate.

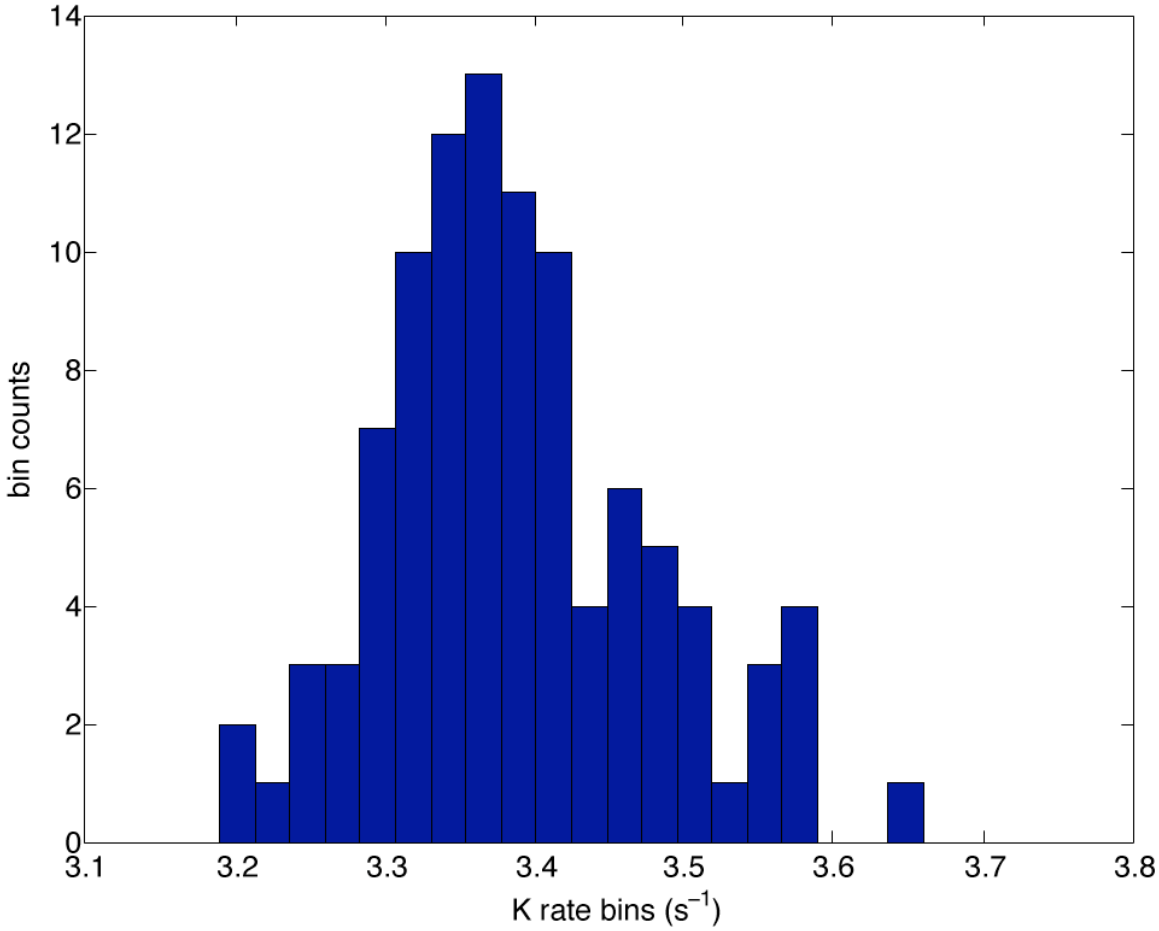

Figure 6.9. Bootstrap distribution of K fits for residue E27 of Hsc70 NBD. This distribution is based on fits to a bootstrap ensemble of 100 permutations of E27's symmetric reconversion data, constructed by random sampling with replacement.

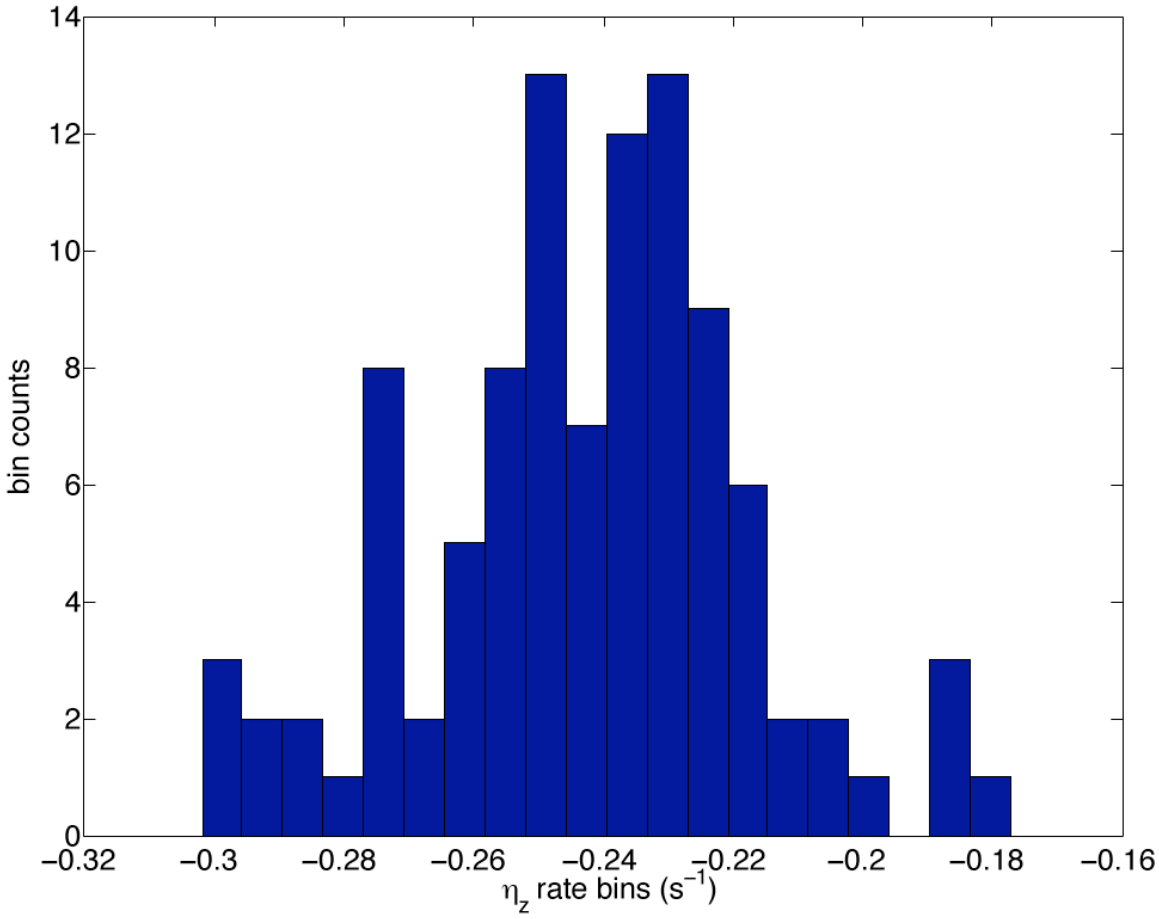

Figure 6.10. Bootstrap distribution of η_z fits for residue E27 of Hsc70 NBD. This distribution is based on fits to a bootstrap ensemble of 100 permutations of E27's symmetric reconversion data, constructed by random sampling with replacement.

Figure 6.11 illustrates an η_{xy} fit for the Wang/Rance/Palmer η_{xy} experiment on residue E27 of Hsc70 NBD. Fit results at ±1 standard deviation are marked by green lines. The high rate of η_{xy} relaxation in large systems is demonstrated by the considerably faster rate of relaxation demonstrated in Hsc70 NBD vice smaller systems such as calmodulin and ubiquitin.

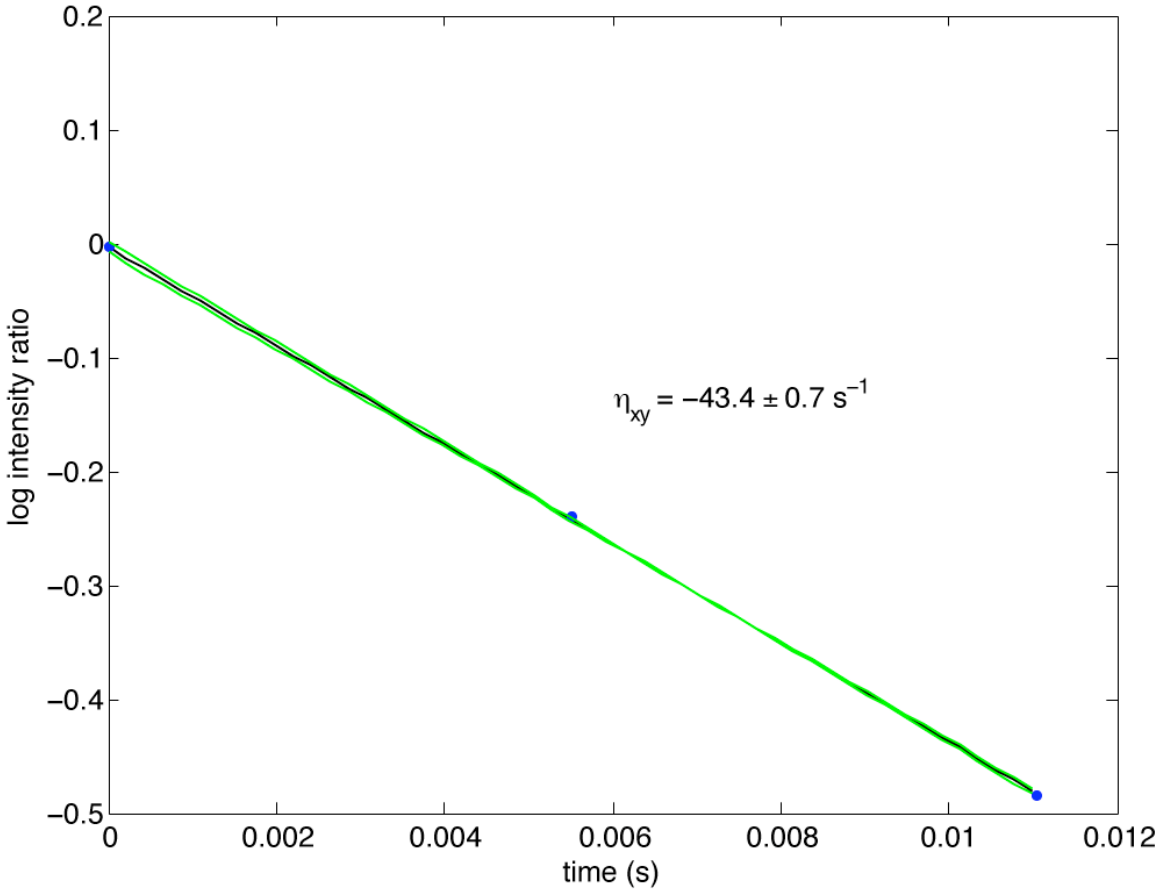

Figure 6.11. η_{xy} fit for the Wang/Rance/Palmer η_{xy} experiment on residue E27 of Hsc70 NBD.
Fit results at ±1 standard deviation are marked by green lines.

η_{xy} and η_z rates are used to calculate D_i values for individual residues, and individual residues are averaged, or fit, into an overall diffusion tensor picture for the entire molecule. Given that there are 172 residues in the η_zK data set, the error of conclusions drawn for the molecule as a whole is significantly reduced below that of the error of conclusions for individual residues. The large errors in η_{xy} and η_z rates are therefore manageable within the experimental framework: as long as a reasonably precise estimate of the overall diffusion behavior can be made, the measurements of η_{xy} and η_z are "good enough".

η_{xy} and η_z rates are used to calculate D_i values for individual residues, and individual residues are fit to produce an overall diffusion tensor picture for the entire molecule. Figure 6.12 illustrates D_i calculated from the η_{xy}/η_z ratios of Hsc70 NBD as a

function of $P_2(\cos(\vartheta_i))$, ϑ_i being the off-angle of the i^{th} ^{15}N-$^{1}H_N$ amide bond vector axis from the symmetry axis of the axially symmetric diffusion tensor resulting from PDBINERTIA. QUADRIC_DIFFUSION is then used to calculate an overall diffusion tensor estimate from the D_i data as a function of theta. The resulting axially symmetric characteristic global diffusion times are listed in Table 6.2.

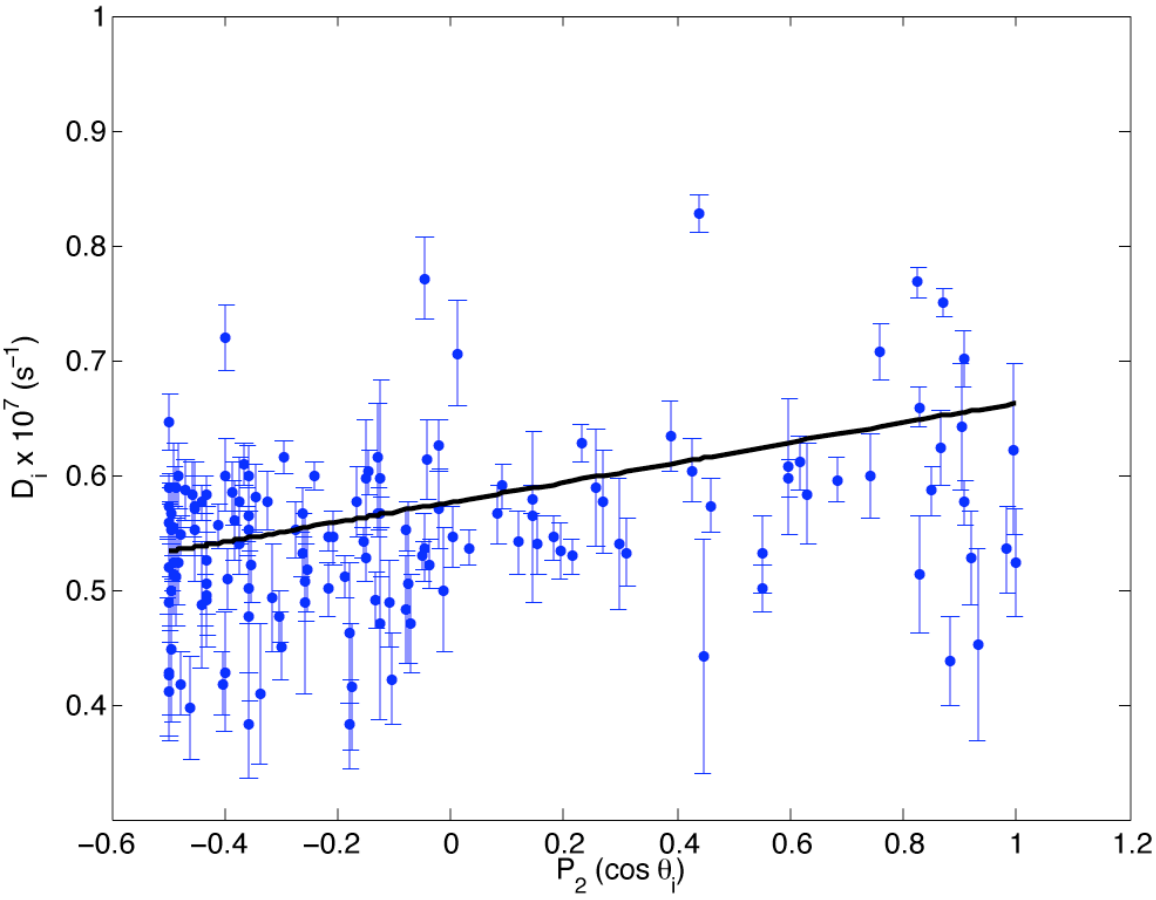

Figure 6.12. D_i vs. $P_2(\cos \vartheta_i)$ fit used to determine rotational diffusion constants for Hsc70 NBD.
Line indicates best fit axially symmetric rotational diffusion tensor from the program QUADRIC_DIFFUSION.

Diso (10^7 s^{-1})	0.58 ± 0.01
$D_{\|}/D_T$	0.61 ± 0.10
ϑ (rad)	1.02 ± 0.15
φ (rad)	1.31 ± 0.15

Table 6.2. Hsc70 NBD axially symmetric diffusion tensor characteristics. Angles are defined with respect to the PDBINERTIA alignment of the protonated 2QWL X-ray structure.

Figures 6.13, 6.14 and 6.15 illustrate K and η_z rates, respectively, for Hsc70 NBD. Figures 6.13 and 6.14 indicate that K rates are strongly affected by secondary structure, with lower rates detected in beta sheets than in alpha helices. Residues experiencing amide proton water exchange are indicated in black in Figures 6.13 and 6.15; these residues include 24, 37, 44, 60, 75, 92, 97,98, 108, 141, 149, 150, 189, 190, 191, 192, 214, 219, 222, 224, 226, 251, 252, 255, 265, 279, 281, 283, 289, 296, 301, 302, 313, 329, 364, 382, 383, and 386.

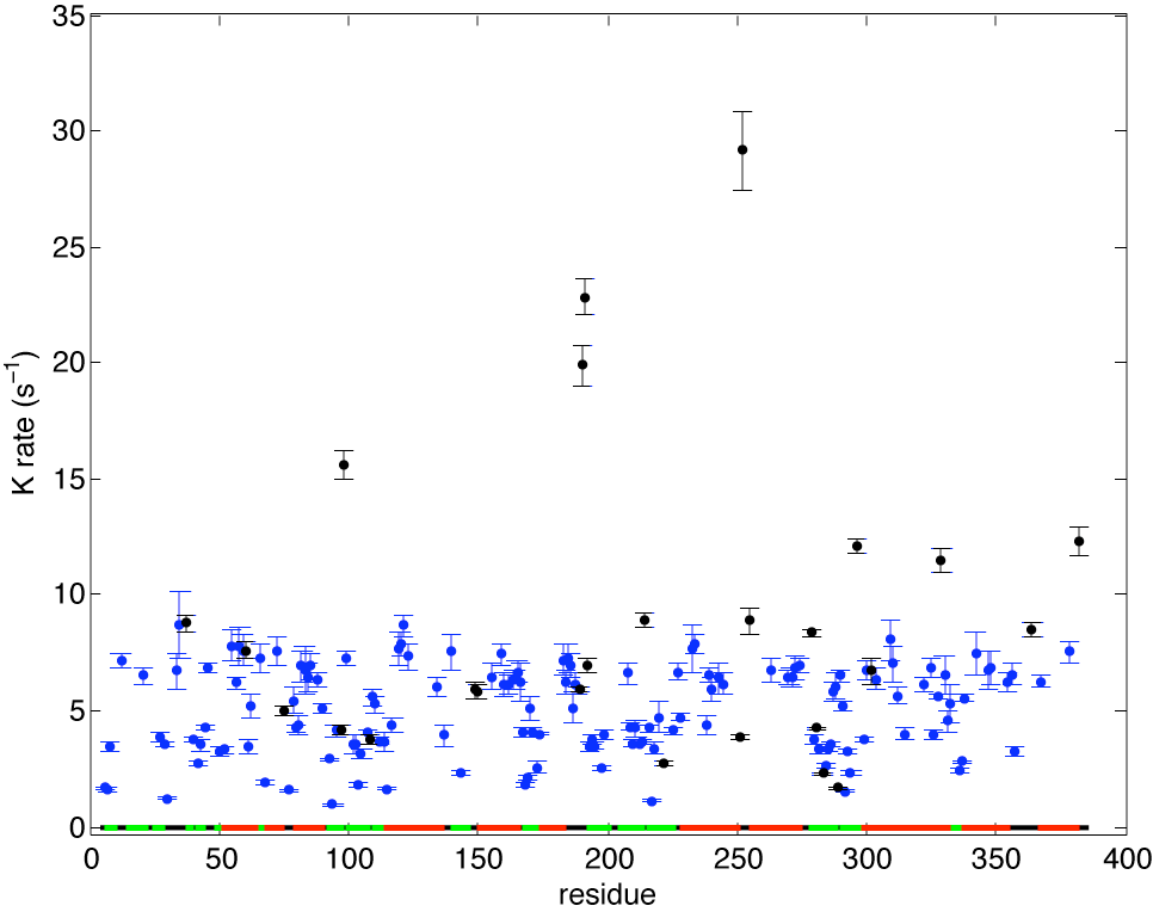

Figure 6.13. K rates for the Hsc70 NBD.
Error bars indicate ±1 standard deviation as measured from bootstrap distributions of 100 permutations of symmetric reconversion data as in Figure 6.9. Residues experiencing amide proton exchange are indicated in black. The line at the bottom of the graph indicates protein secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

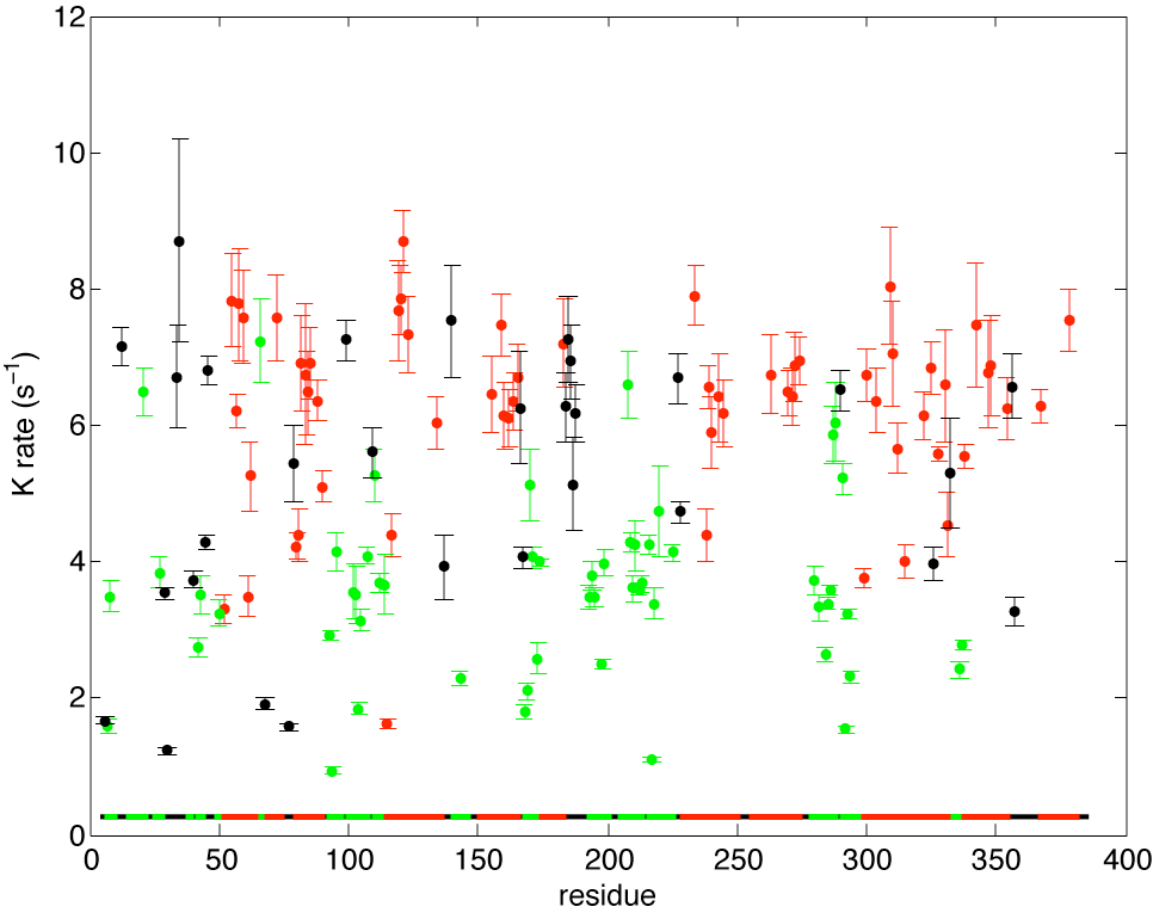

Figure 6.14. K rates for the Hsc70 NBD (amide proton exchanging residues not included).
Error bars indicate ±1 standard deviation as measured from bootstrap distributions of 100 permutations of symmetric reconversion data as in Figure 6.9. The color of data points indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

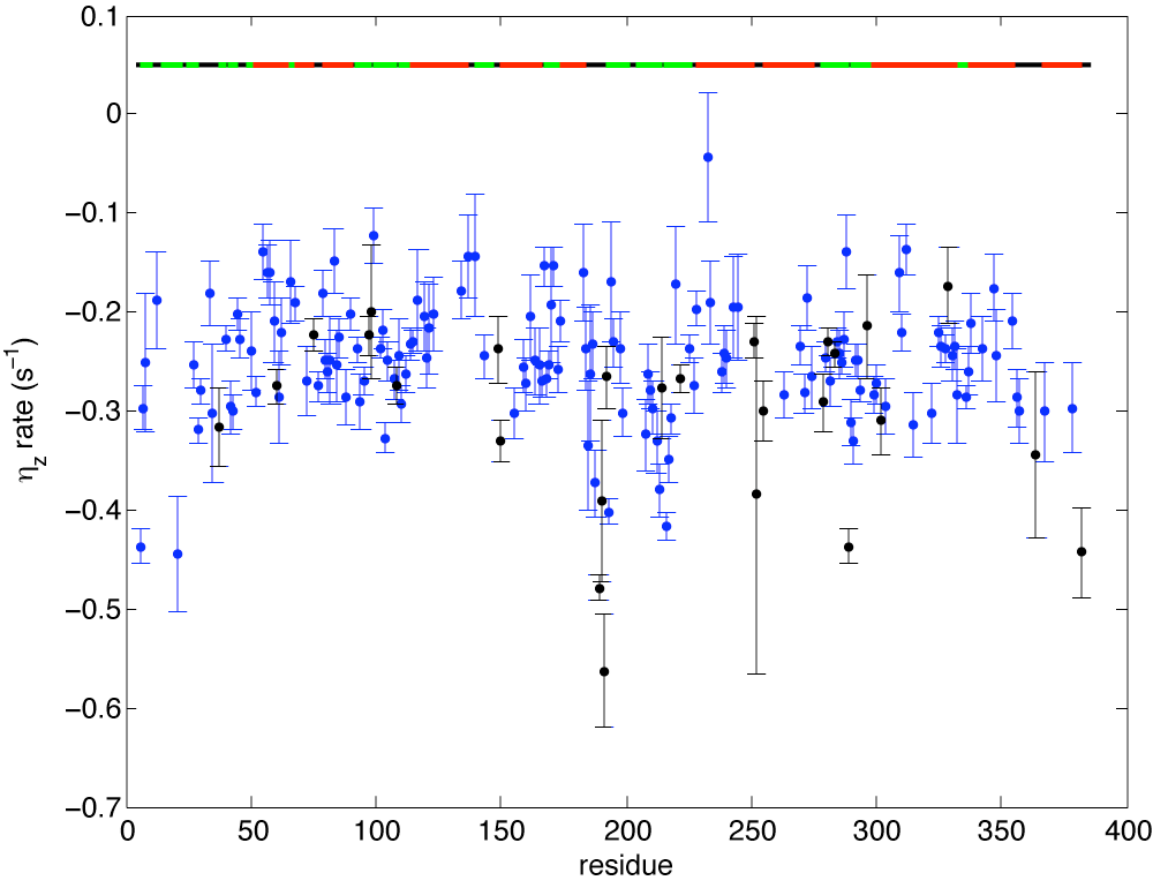

Figure 6.15. η_z rates for the Hsc70 NBD.
Error bars indicate ±1 standard deviation as measured from bootstrap distributions of 100 permutations of symmetric reconversion data as in Figure 6.10. Residues experiencing amide proton exchange are indicated in black. The line at the top of the graph indicates protein secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

Figures 6.16 and 6.17 examine means of estimating K_{rig} rates and their errors in Hsc70 NBD. In Figure 6.16, K_{rig} rates are calculated based on the 2QWL, 1HPM and 3HSC X-ray structures, with the spread of K_{rig} rates indicated by bars stretching from highest to lowest K_{rig} rate.. The K_{rig} rates derived from these three independent crystal structures display small or moderate variations in the large majority of cases. Several residues have large variations in K_{rig} rate, indicating fundamental disagreement among the crystal structures.

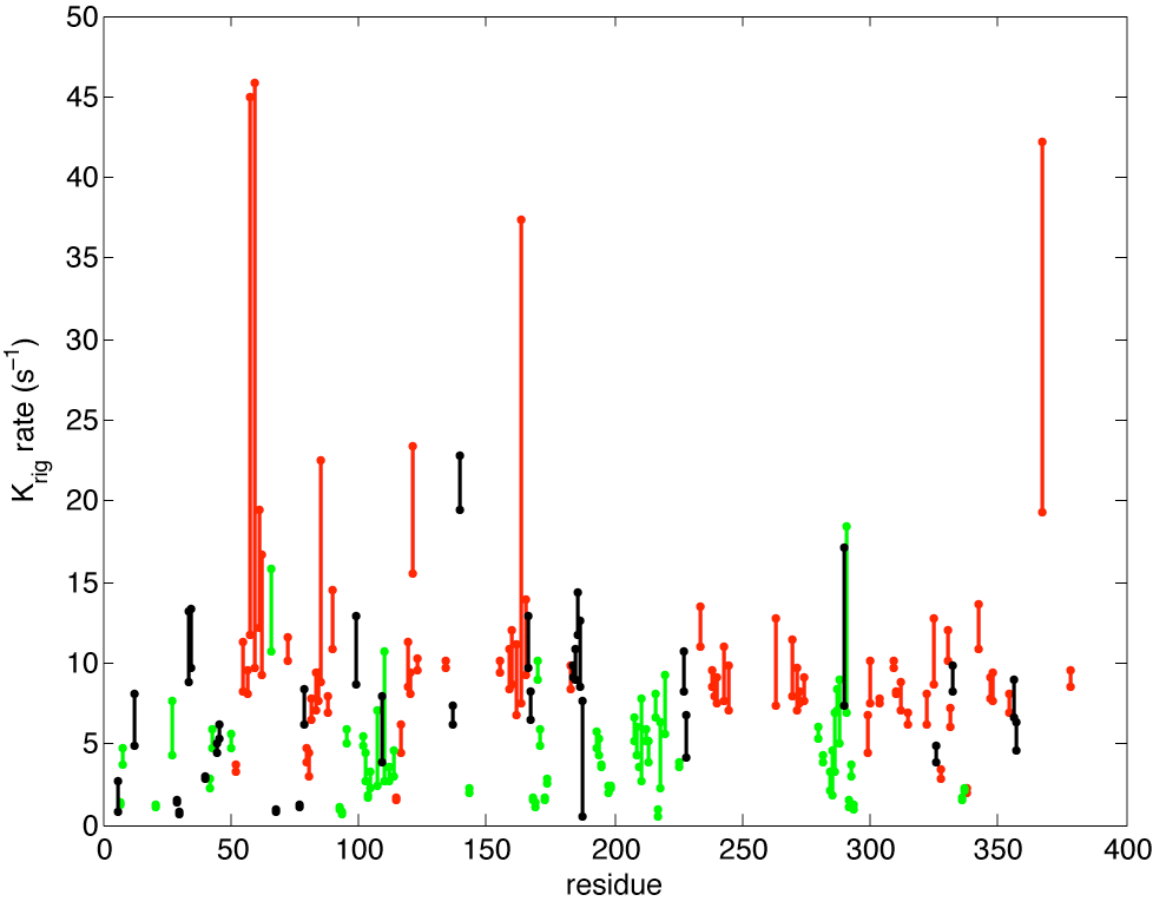

Figure 6.16. K_{rig} values derived from multiple structures of Hsc70 NBD. Vertical lines indicate spread between minimum and maximum values of K_{rig} rates from three independent crystal structures of Hsc70 NBD in ADP.Pi form (PDB identifiers 2QWL, 1HPM and 3HSC). Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL, 1HPM and 3HSC crystal structures with all exchangeable protons incorporated. Vertical line color indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

Figure 6.17 illustrates the results of applying the Cruickshank error estimation method to the 2QWL X-ray structure of the Hsc70 NBD. A 100-member Monte Carlo ensemble of K_{rig} rates is constructed by applying normally distributed random errors to all interproton distances in 2QWL and summing the resulting perturbed $K_{rig,i}$ rates. The standard deviations of these normal errors are generated by the Cruickshank X-ray structure error approximation detailed in Chapter 2. The resulting distributions of K_{rig} rates for each residue are plotted as blue dots in Figure 6.17. Structural K_{rig} spans from

Figure 6.16 are overlaid on Cruickshank-based distributions for purposes of comparison. It is immediately apparent that the Cruickshank error estimation method presents a much larger estimate of the interproton distance error than that actually observed from Figure 6.16's comparison of several independent X-ray structures.

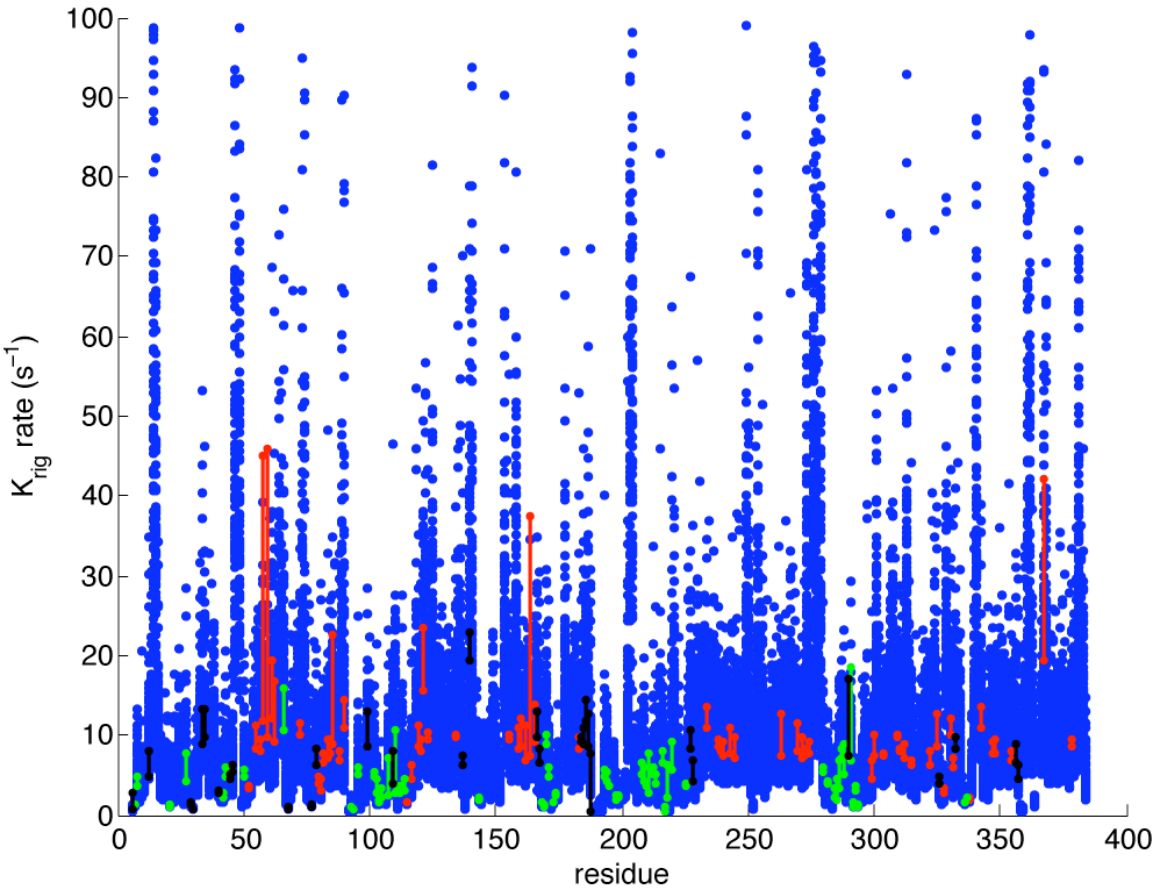

Figure 6.17. Comparison of Cruickshank-calculated errors in K_{rig} with multiple-structure-based errors in K_{rig} for Hsc70 NBD.
Blue dots indicate members of a K_{rig} Cruickshank error ensemble calculated from 100 Monte Carlo perturbations of interproton distance estimates derived from the 2QWL X-ray structure with a Gaussian error model having a standard deviation equal to the uncorrelated interproton distance error (√2 * the position error). Vertical lines indicate spread between minimum and maximum values of K_{rig} rates from three independent crystal structures of Hsc70 NBD in ADP.Pi form (PDB identifiers 2QWL, 1HPM and 3HSC) as in Figure 6.16. Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL, 1HPM and 3HSC crystal structures with all exchangeable protons incorporated. Vertical line color indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

This fact calls into question the validity of the simple Cruickshank error estimate we have developed, suggesting that it severely underestimates the precision with which K_{rig} rates can be extracted from crystal structures. (The following should not be taken as an indictment of Cruickshank's work in its original context.)

The reason for the underestimation likely lies in the fact that the Cruickshank error estimate is a point-to-point distance error estimate between two uncorrelated positions in the molecule. Although the positions of two widely separated protons in a protein may be considered approximately uncorrelated, the positions of nearby neighbors are correlated by the presence of protein primary and secondary structure. An extreme example is the peptide backbone C'-Cα bond length, which is known to be 1.51-1.52 Å: the Cruickshank error estimate from 2QWL will place an 0.26 Å error on this bond length even though the distance of a CO from its Ca neighbor is known within 0.01 Å.

Exchangeable proton-amide proton interactions are nonbonded and are estimated from computational protonation of X-ray structures. This undoubtedly increases the error of interproton distance estimates far above that seen for estimates of the distance between bonded atoms. Nonetheless, amide protons participate in a large number of regular secondary structure hydrogen bonding interactions. Furthermore, thanks to the r^{-6} dependence of K rates, the largest contributors to the $K_{rig,i}$ sum are those exchangeable protons nearest to the ith (sensor) amide proton. These neighbor protons are more likely to share secondary structure with the sensor amide proton, resulting in a more tightly constrained distance error estimate influenced in part by more accurate knowledge of bonded interactions.

In the long term, MD simulations should provide a powerful tool for investigating the issue of interproton distance error. The multiple bonded and nonbonded interactions encompassed by MD force fields should help quantify the degree of imprecision in distance estimates for the local nonbonded interactions that make up K rates.

In the short term, we can take refuge in the fact that three different independent crystal structure estimates of K_{rig} are available. For the remainder of this chapter, the three crystal structures of Hsc70 NBD in its ADP.Pi form will be used as benchmarks for

the evaluation of K data. The level of error in these benchmarks will be reflected in the divergence (or concordance) of their K_{rig} rates.

With the issue of error in K_{rig} dealt with, I turn my attention to the correspondence between K and K_{rig}. Figure 6.18 compares experimental K rates (blue dots/trendline) with K_{rig} data from Figure 6.16. The correspondence between K and K_{rig} is qualitatively excellent, especially considering the high degree of residue-to-residue variability in K_{rig}. K rates are seen to follow the trend of K_{rig} rates in remarkable detail. Figure 6.19 presents the same data as displayed in Figure 6.18, with the y-axis limited to $K_{rig} = 20\ s^{-1}$ in order to display experimental K rates in greater detail.

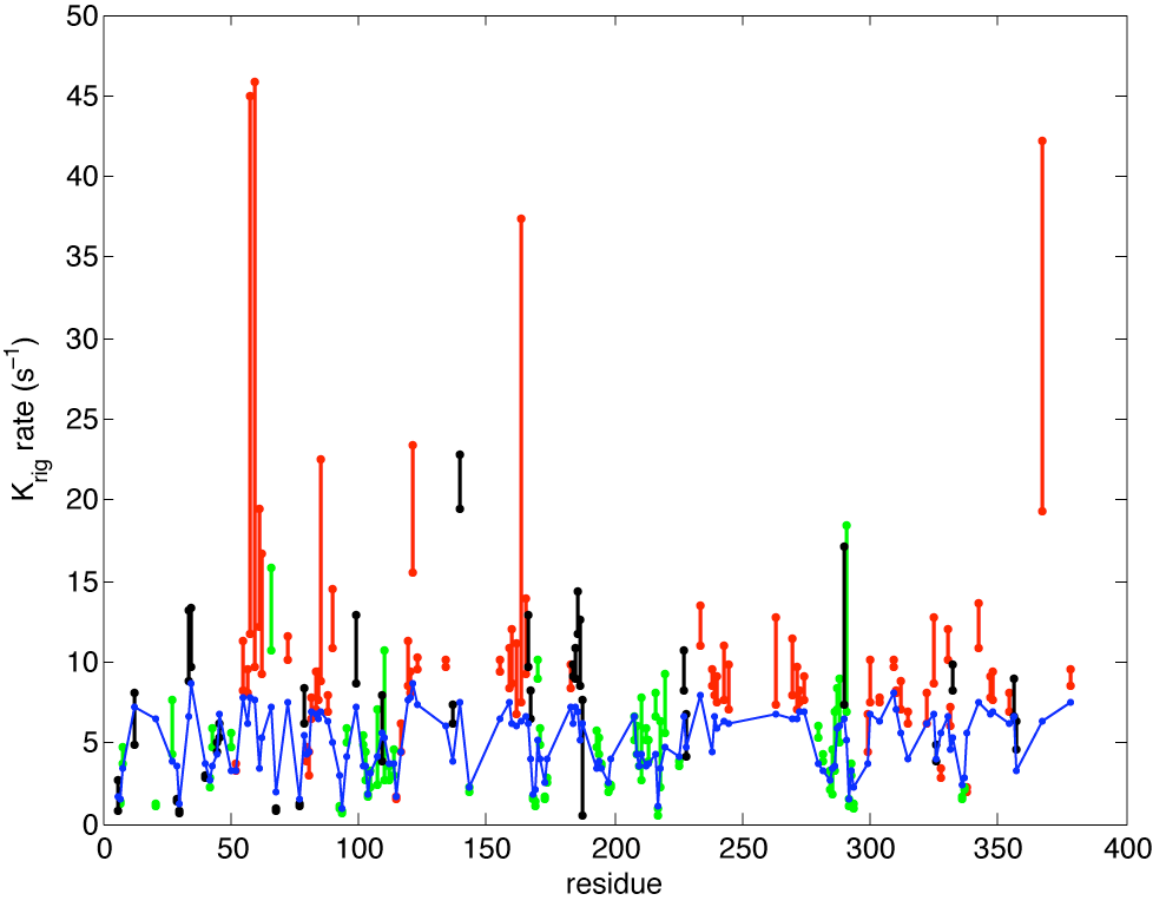

Figure 6.18. Comparison of K and K_{rig} values for Hsc70 NBD.
Vertical lines indicate spread between minimum and maximum values of K_{rig} rates from three independent crystal structures of Hsc70 NBD in ADP.Pi form (PDB identifiers 2QWL, 1HPM and 3HSC). Blue lines indicate experimental K rates measured as described in the text. Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL,

1HPM and 3HSC crystal structures with all exchangeable protons incorporated. Vertical line color indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

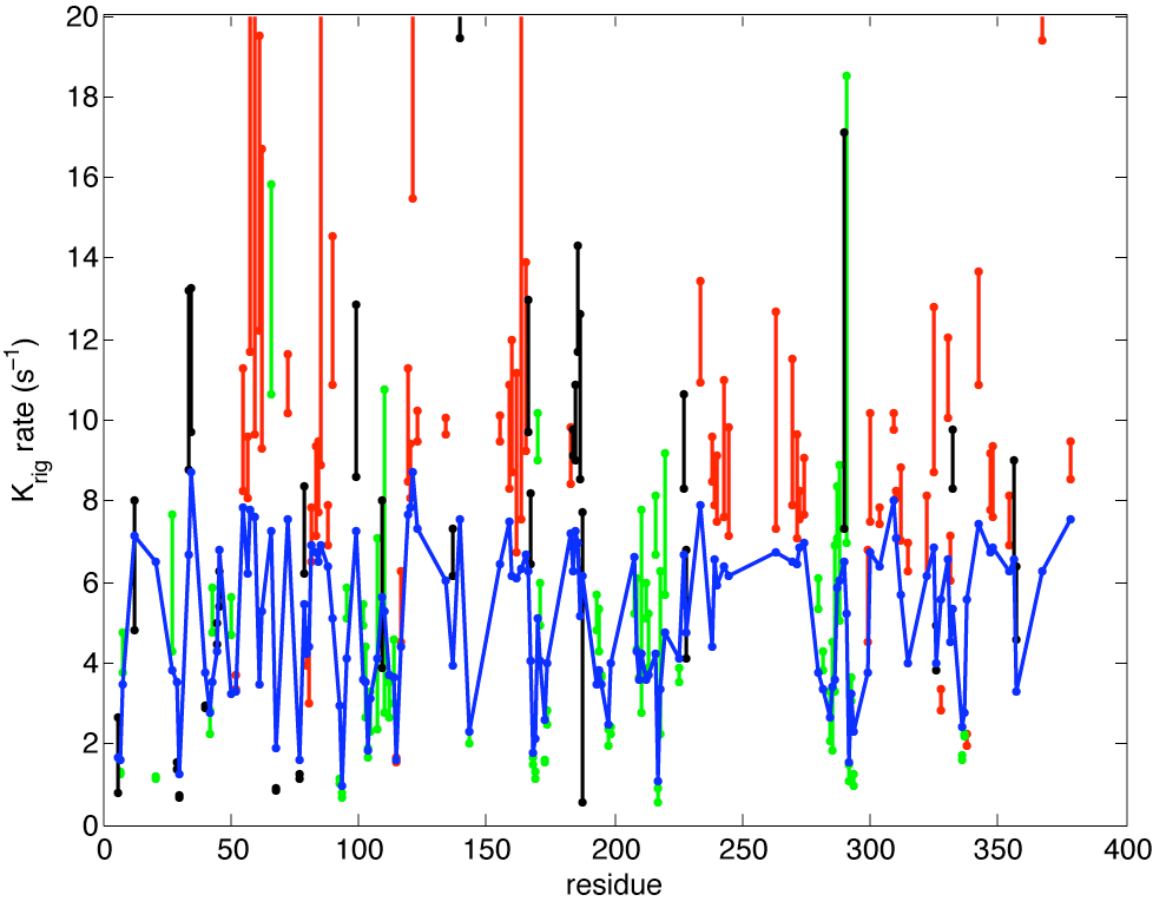

Figure 6.19. Comparison of K and K_{rig} values for Hsc70 NBD (detail of lower K rates).
Data same as shown in Figure 6.17, expanded to display detail below K = 20 s^{-1}. Vertical lines indicate spread between minimum and maximum values of K_{rig} rates from three independent crystal structures of Hsc70 NBD in ADP.Pi form (PDB identifiers 2QWL, 1HPM and 3HSC). Blue lines indicate experimental K rates measured as described in the text. Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL, 1HPM and 3HSC crystal structures with all exchangeable protons incorporated. Vertical line color indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure.

Figure 6.20 contains a correlation plot of experimental K rates vs. multiple-structure K_{rig} rates. β-sheet regions of the protein are seen to have significantly lower K rates than α-helical and loop structures, as expected from β-sheets' lower proton density.

Figure 6.21 contains the same information as Figure 6.19 with K error information included as vertical bars.

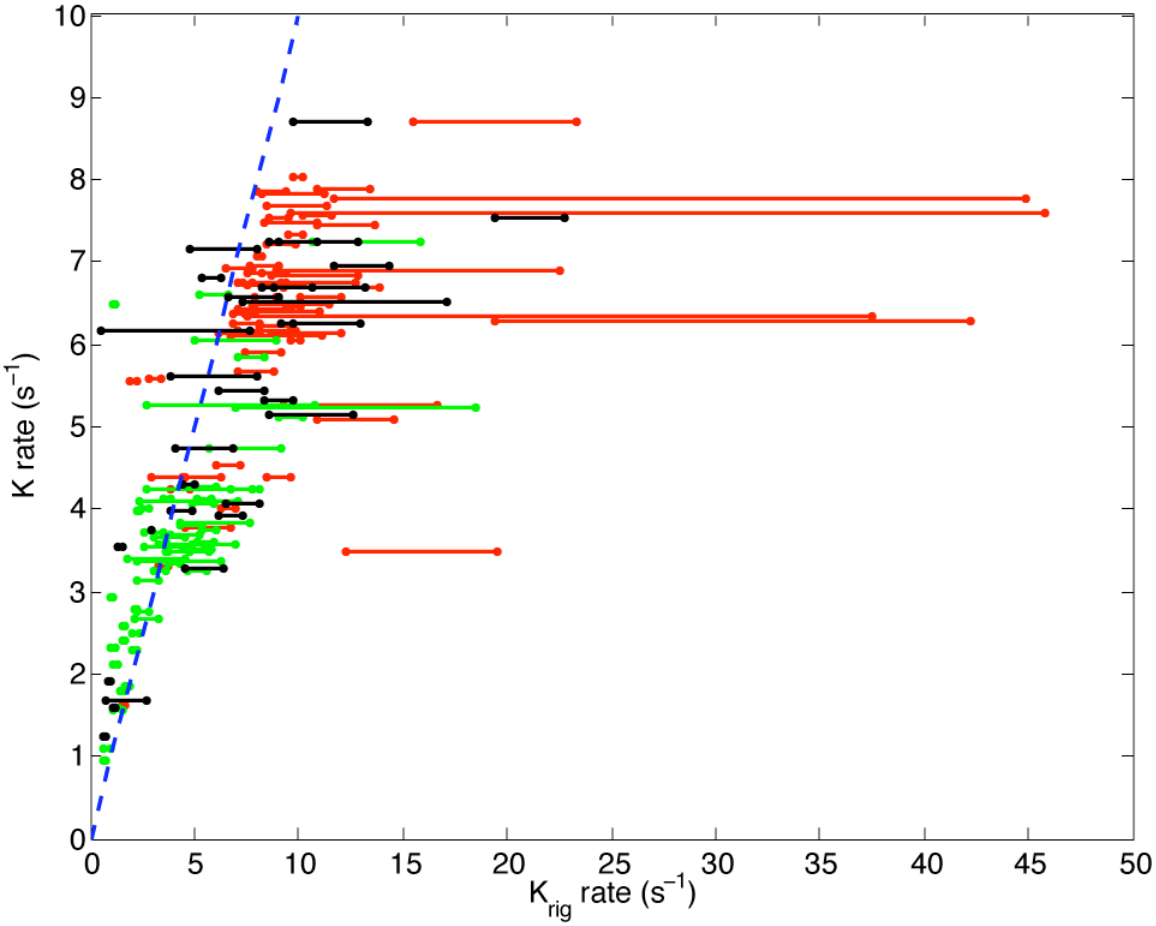

Figure 6.20. Correlation plot of K and K_{rig} values for Hsc70 NBD. Horizontal lines indicate spread between minimum and maximum values of K_{rig} rates from three independent crystal structures of Hsc70 NBD in ADP.Pi form (PDB identifiers 2QWL, 1HPM and 3HSC), while vertical position indicates experimentally measured value of K. Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL, 1HPM and 3HSC crystal structures with all exchangeable protons incorporated. Line color indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure. Blue dotted line indicates K = K_{rig} (slope 1).

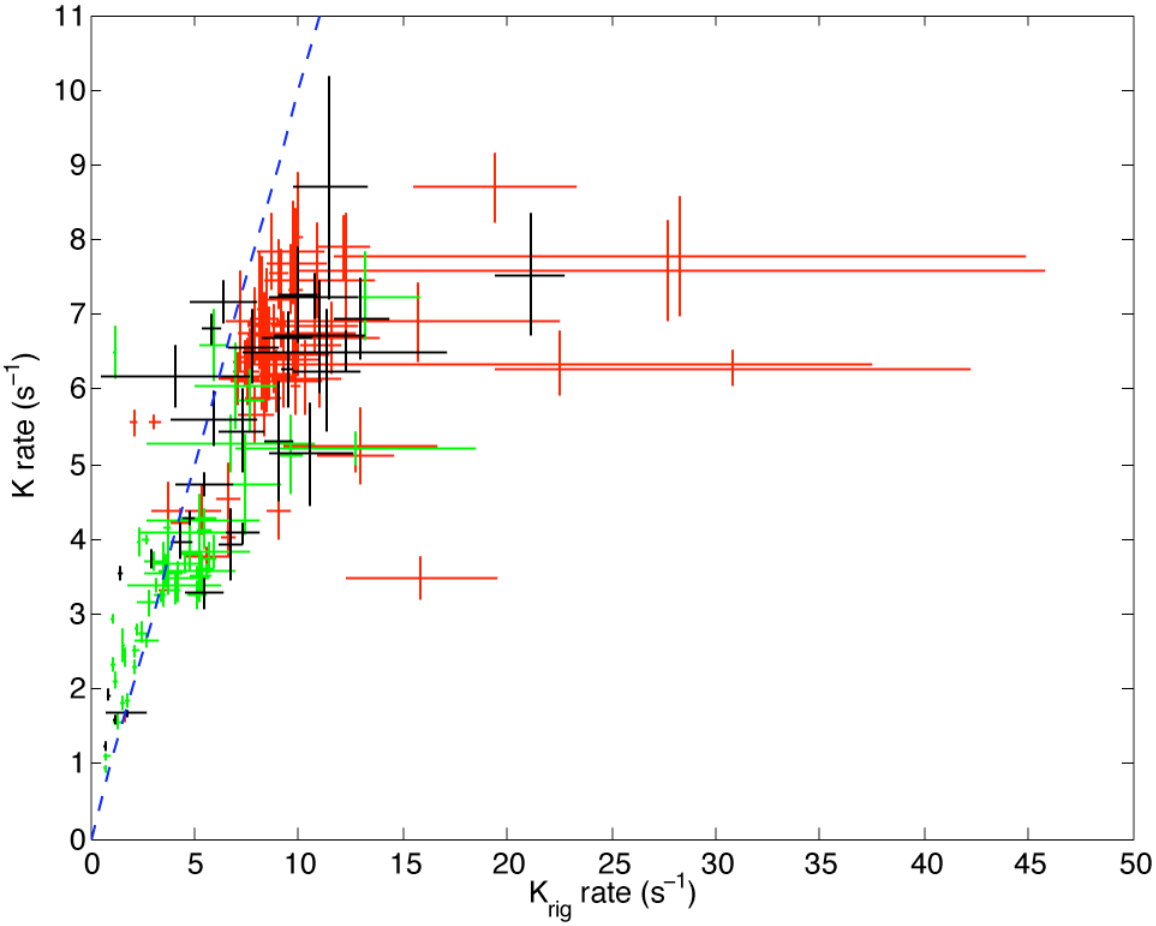

Figure 6.21. Correlation plot of K and K_{rig} values for Hsc70 NBD (experimental error in K indicated).
Horizontal lines indicate spread between minimum and maximum values of K_{rig} rates from three independent crystal structures of Hsc70 NBD in ADP.Pi form (PDB identifiers 2QWL, 1HPM and 3HSC), while vertical lines indicate one standard deviation error in experimentally measured value of K. Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL, 1HPM and 3HSC crystal structures with all exchangeable protons incorporated. Line color indicates secondary structure, with green indicating beta-sheet structure, red indicating alpha-helix structure, and black indicating loop structure. Blue dotted line indicates K = K_{rig} (slope 1).

The K/K_{rig} ratio Q is seen to be somewhat higher than 1 for low values of K_{rig} below about 3.5 s^{-1} and then to drop below 1 on average as K_{rig} rates increase. The residues with $K_{rig} < 3.5$ s^{-1} constitute a distinct population of largely β-sheet residues with strong degrees of structural agreement regarding their K_{rig} rates. The overall effect is that the ceiling of the K vs. K_{rig} correlation does not follow a line of slope K = K_{rig} and K-intercept = 0 with drops in K due to dynamics as is expected. Rather, the line followed

by the ceiling of the data has a positive K intercept around 0.7-1.0 s^{-1}. There are several possible explanations for the discrepancy: underestimation of the overall global tumbling time causing artificially low K_{rig} rates; structural problems causing artificially low K_{rig} rates; and constitutive low rates of amide proton exchange causing artificially high K rates.

Underestimation of the overall global tumbling time seems an unlikely explanation, since a very large underestimation by a factor of 50-100% would be required to explain the observed ratios between K and K_{rig} at these low K_{rig} values. This degree of underestimation would in turn cause higher β-sheet and α-helical K_{rig} rates to be much too large, leading to a K vs. K_{rig} slope considerably less than 1.

Structural problems seem more possible as an explanation, given that residues are more likely to report high Q values if they are experiencing significant contributions to their K_{rig} from amide protons which themselves report very low K_{rig} rates. This suggests that the interacting amide protons are slightly closer to each other than structure indicates, leading to K rates exceeding (very low) K_{rig} rates and consequently high Q values. This phenomenon is illustrated in Figure 6.22, with Q values plotted as a function of the percentage contribution to K_{rig} rates from amide protons which themselves report K_{rig} rates lower than 3.5 s^{-1} plotted on the x-axis. (For instance, if a proton received 50% of its total K_{rig} rate from amide protons which reported K_{rig} rates less than 3.5 s^{-1}, it would be plotted at 50% in Figure 6.22.) This explanation requires further investigation in the context of X-ray electron density and molecular dynamics data in order to evaluate its credibility.

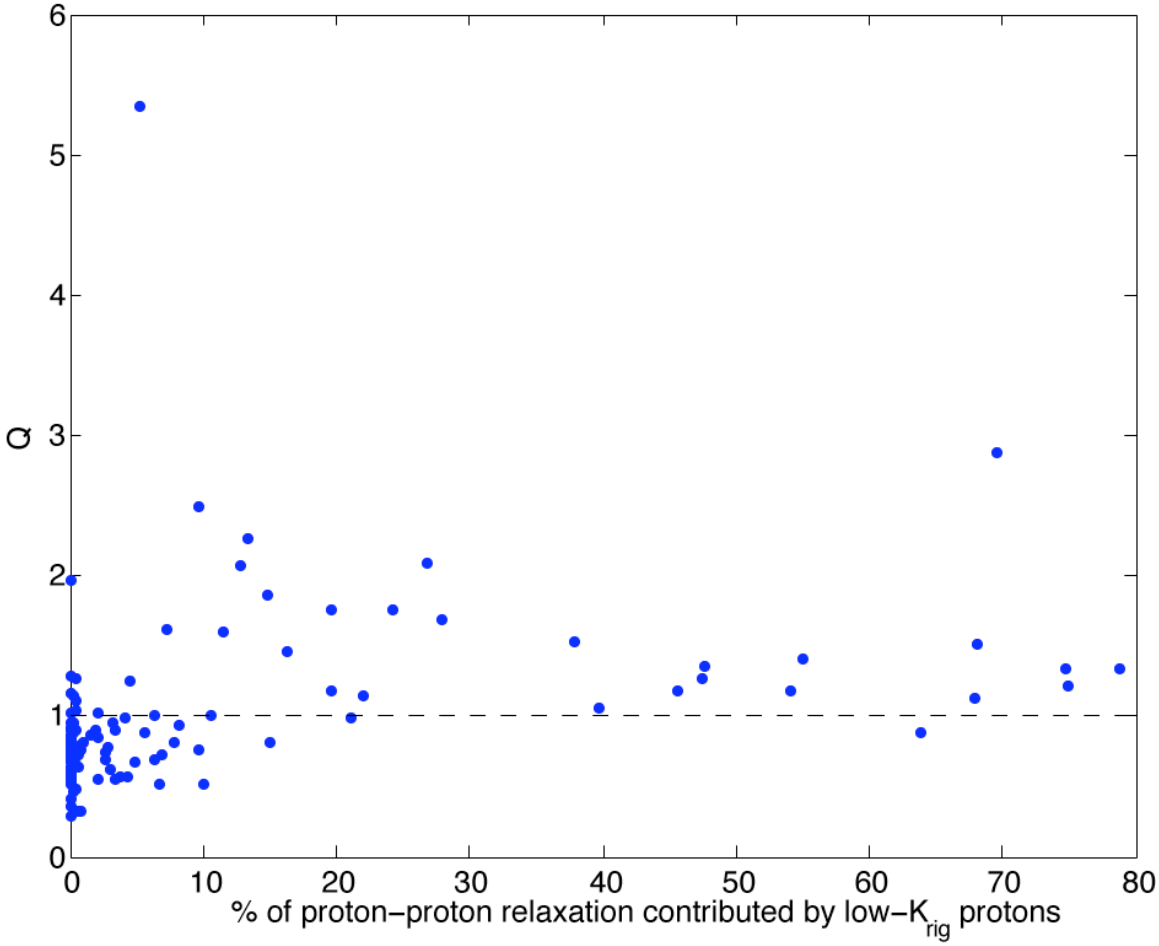

Figure 6.22. Q as a function of percentage of K_{rig} rate contributed by amide protons reporting low K_{rig} rates.
Residues experiencing amide proton exchange are excluded. K_{rig} values are calculated based on MolProbity protonation of 2QWL with all exchangeable protons incorporated.

Low rates of amide proton exchange seem unlikely as a source of excess K rates, given that amide proton exchange experiments do not detect significant transfer of water magnetization to most of the low-K_{rig} peaks during inversion experiments.

In order to quantitate the correlations described in the preceding figures, Table 6.2 displays correlation coefficients R among K_{rig} and K values in Hsc70 NBD. All three crystal structures demonstrate strong, albeit imperfect, correlation with each other. The correlation of experimental K data with the K_{rig} rates calculated from the 2QWL structure is among the strongest observed in this work.

	2QWL K_{rig}	1HPM K_{rig}	3HSC K_{rig}	Expt. K
2QWL K_{rig}		0.83	0.65	0.76
1HPM K_{rig}	0.83		0.85	0.66
3HSC K_{rig}	0.65	0.85		0.57
Expt. K	0.76	0.66	0.57	

Table 6.3. Correlation coefficients R between K_{rig} and experimental K rates for Hsc70 NBD structures.

Figure 6.23 is a histogram of Q values based on the data displayed in Figure 6.18. The distribution is strongly peaked at Q = 0.8-0.9, indicating that that the estimations of K_{rig} from the X-ray structure agree well with the experimental data and that the amide proton network is highly ordered. Several outliers exist, with Q values as high as 6, although it should be noted that the culprit (F21) is not among the confidently assigned residues and may well constitute a peak belonging to another residue. The presence of these outliers militates in favor of the use of the geometric mean, rather than the arithmetic mean, in order to characterize the distribution of Hsc70 NBD Q values. The geometric mean of the distribution is $\langle Q \rangle_G = 0.86$. (For comparison, the geometric mean of peptide-bound calmodulin is 0.87, very close to its arithmetic mean of 0.88.)

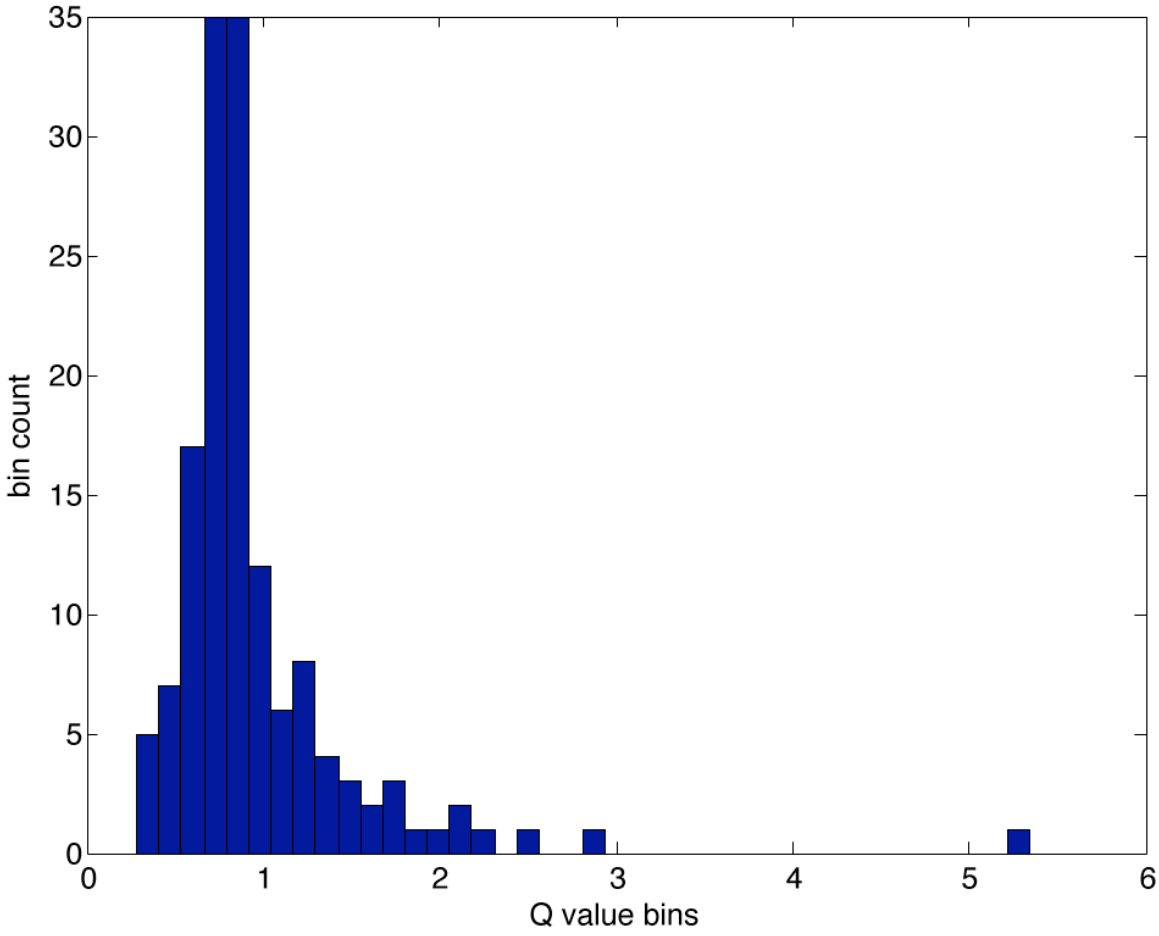

Figure 6.23. Histogram of Q values calculated in Figure 6.15 for 2QWL. Residues experiencing amide proton exchange are excluded.

In summary, studies of the Hsc70 NBD's proton-proton dynamics are an ongoing endeavor. The close agreement of experimental K rate trends with structure-based K_{rig} trends and the successful synthesis of all parts of the experiment (protein perdeuteration, η_zK experiment, η_{xy} experiment, amide proton exchange detection water inversion experiment) in a large protein system augurs well for future use of the η_zK experiment to study proton-proton dynamics in large proteins.

CHAPTER 7: SUMMARY AND FUTURE DIRECTIONS

7.1 Summary

Proteins are dynamic objects that undergo motion over a wide range of timescales from trillionths of a second to minutes and hours. Changes in protein dynamics affect important protein operating parameters such as overall protein stability, ligand affinity, allosteric regulation, and catalytic efficiency.[141; 142; 143; 144; 145; 146; 147; 148; 149; 150] In turn, these protein operating parameters govern biochemical processes that underlie the chemistry of life and a wide range of important technologies from biofuels to cancer therapeutics.[151] These practical applications motivate basic research into the physics of protein dynamics as a means of understanding how protein dynamics affect protein operation.

In this work, I describe the measurement of proton-proton dynamics in proteins by means of NMR measurement of amide proton spin flip rates. To briefly summarize the physics of the η_zK experiment, the local proton environment of the amide proton causes the amide proton spin to flip population state by means of homonuclear dipole-dipole zero-quantum relaxation between the amide proton and each of its proton neighbors. The experimentally measured rate at each amide proton site is the sum of all individual zero-quantum relaxation rates between the amide proton and its proton neighbors. Each of these zero-quantum K rates can be modulated by dynamics along the interproton vector. The result is an inverse-distance-weighted average of dynamic effects over the amide proton's neighbor protons.

As the most numerous atoms in proteins, protons constitute a dense population of dynamic probes. The backbone and sidechain are liberally encrusted with protons, and all amide protons will be in range of one or more highly effective relaxing neighbor protons. The distance weighting provided by the r^{-6} dependence of K rates results in a

focus on the local proton environment, with interactions effectively cut off beyond the 5-6 Å 'NOE range'. The through-space nature of proton-proton dipolar relaxation allows the η_zK experiment to sample motions of all nearby protons, whether they are present in the backbone, on the sidechain, on the same residue, or on distant residues.

The η_zK experiment also measures η_z longitudinal N/NH DD/CSA cross-correlated relaxation rates. These rates, which constitute the (major) difference between NzHa and NzHb auto-relaxation behavior, are incorporated into a protocol for measuring protein rotational diffusion behavior based on the work of Kroenke et al.[109]

We introduce a means of analyzing the proton dynamics of proteins by comparing K rates with K_{rig} rates calculated for the case of a rigid protein derived from high-resolution X-ray structures and protein rotational diffusion tensors. The ratio of K rates to K_{rig} rates, Q, provides an order parameter-like means of measuring the distance-weighted average proton-proton dynamics near amide proton sites.

Measurement of Ca^{2+}-saturated calmodulin Q parameters in free and smMLCKp-bound forms indicates that proton-proton dynamics exist throughout both forms and that significant rigidification of the proton network takes place on peptide binding. These results also reveal the presence of amide proton exchange and its negative effects on the η_zK experiment. Alternative interpretations of K rates available in cases where detailed X-ray structures are not available are discussed.

η_zK data on ubiquitin is used to analyze the agreement between experimental K rates and several published NMR structures of ubiquitin. All structures are found to indicate the presence of proton-proton dynamics in ubiquitin. The level of correlation between K rates and K_{rig} rates is introduced as a measurement of the ability of a proposed protein structure to explain the K rates associated with that protein structure.

The final experimental effort described in this thesis work is the application of the η_zK experiment to the Hsc70 NBD. This 44 kD protein lay beyond the range of practical NMR relaxation spectroscopy prior to the development of the experimental techniques described herein.

Preparation of highly concentrated samples of $^{2}H,^{13}C,^{15}N$-labeled Hsc70 NBD is described, as is the secondary step of reprotonation at exchangeable proton sites. The

η_zK experiment is demonstrated to operate effectively on the Hsc70 NBD. An isotropic rotational diffusion tensor is estimated based on η_{xy}/η_z measurements. The presence of amide proton exchange is measured by means of water inversion experiments.

Experimental K rates are compared with two X-ray structures of Hsc70 NBD in order to estimate levels of interproton disorder associated with the residual protonation available at NH, OH and SH sites. The experimental K rates of Hsc70 NBD residues whose assignments are certain display an excellent correlation with theoretical K rates calculated from both structures. Based on correlation plots between K and K_{rig} rates, the level of disorder is estimated to be lower than that of other proteins involved in this study.

7.2 Theoretical advances

The equidistant virtual neighbor proton approximation underlying the theory of the η_zK experiment constitutes a modest advance in the practical understanding of NMR relaxation. Prior to the development of this approximation[77], no practical theory of proton-proton longitudinal dipolar relaxation was capable of accounting for multiple neighbor protons. Existing theory of AX^n spin systems[78] was so generalized as to be experimentally and theoretically unhelpful. Failure to account for multiple neighbor protons[76; 87; 117; 152] leads to inaccurate equilibration behavior for cross-relaxation magnetization buildup, preventing accurate quantitation of K rates.

The construction of an improved theoretical treatment for proton-proton longitudinal relaxation processes is of interest to theoretical NMR spectroscopy, since these relaxation processes constitute a major phenomenon in protein relaxation that has not previously been modeled in detail. Earlier treatments abstracted these processes into R_1^H proton relaxation rates, unavoidably masking much of their interesting complexity and presenting an inaccurately simplified picture of relaxation behavior. The existence of the equidistant virtual neighbor proton approximation provides a strong theoretical basis for future investigation of relaxation in non-amide AX^n spin systems, opening a new avenue of research for future NMR spectroscopists.

In combination with a judicious (and experimentally convenient) choice of magnetic observables, the symmetric reconversion procedure[56] employed in the η_zK experiment provides effective decoupling of cross-correlated autorelaxation and cross-relaxation in the longitudinal relaxation of the amide NH spin system. This decoupling allows clean extraction of η_z and K rates. Without this decoupling, the process of autocorrelated cross-relaxation (differential auto-relaxation of NzHa and NzHb states towards an overall equilibrium) is interfered with by the much more rapid process of cross-relaxation (zero-quantum transfer of magnetization between NzHa and NzHb). This results in an inelegant muddling of processes.

Two additional benefits are observed from the synthesis of X(t) and Y(t) curves by multiplication and division of the four symmetric reconversion subexperiments. The auto-correlated autorelaxation of the system is suppressed, because all experiments undergo the same auto-correlated autorelaxation 'pseudo-R1' relaxation process as described in Chapter 2. Additionally, the system is constantly self-referenced due to the construction of data points from the relative intensities of closely separated symmetric reconversion subexperiments. This means that changes in peak intensity due to external factors such as spectrometer instability will not interfere with data analysis. This is not the case in unreferenced relaxation experiments such as R_1 or R_2, where the only point of reference for each peak intensity is a (possibly far-distant) initial data point.

The measurement of η_z in the η_zK experiment presents another methodological advance. The previous state of the art in measuring η_z developed by Ferrage et al.[88] is theoretically and practically sound, but is also an HSQC-based experiment that measures differential relaxation of Nz and 2NzHz magnetization. This means that the experiment cannot be rewritten to use a TROSY readout of magnetization without introducing problems with clean measurement of the relative intensities of Nz and 2NzHz states, since both Nz and 2NzHz magnetization states contain the NzHa magnetization measured by TROSY. As a consequence, the η_z experiment of Ferrage et al. cannot be used in larger proteins which suffer unacceptable peak broadening in conventional HSQC experiments.

In contrast, the η_zK experiment is constructed in a basis of NzHa and NzHb magnetization states, which allows measurement of η_z rates with a clean TROSY spectroscopic readout. Kroenke et al.[109] developed a chemical exchange-free method of calculating protein rotational diffusion tensors by measuring η_{xy}/η_z ratios, but no clean η_z experiment was available to provide the denominator for these ratios, and no TROSY implementations were available for either experiment. The η_{xy} experiment developed by Wang et al.[110] provided a TROSY experiment measuring the numerator, and the η_zK experiment provides a TROSY experiment measuring the denominator. Combining these three experimental efforts results in a protocol for measuring the rotational diffusion tensors of large proteins that require TROSY spectroscopy. This constitutes a significant expansion of the range of proteins amenable to rotational diffusion measurements by NMR.

K rates constitute another exciting new horizon opened by the η_zK experiment. The η_zK experiment is the first experiment that allows quantitative measurement of proton-proton relaxation in protein-scale polypeptide systems. High proton density and the through-space character of proton-proton dipolar relaxation result in a wholly new type of relaxation experiment sampling regional dynamics quite different from the bond-vector-linked relaxation experiments currently available. TROSY-based relaxation preparation and observation result in an experiment that can be performed on larger proteins in solution.

The η_zK method has as an inherent advantage over other relaxation experiments in the large-protein environment. The key rate parameter, K, scales with J(0) and increases linearly with molecular weight. Meanwhile, non-zero-frequency spectral density functions decrease in size as molecular weight and magnetic field strength increase, efficiently "cleaning up" the experiment. Zero-quantum K rates are also independent of cross-correlation, CSA, and conformational exchange contributions. This means that they are not affected by the increases in magnetic field needed for the studies of larger proteins.

7.3 Summary of experimental results and advances

7.3.1 Calmodulin

η_z and K rates are measured for free and smMLCKp-bound Ca^{2+}-saturated calmodulin at 35 C and pH 7.0. Measurements of protein rotational diffusion tensors indicate that the η_zK experiment allows diffusion information to be extracted from η_{xy}/η_z ratios. The resulting diffusion tensors are close to literature values and may be more accurate measurements due to the presence of chemical exchange effects in the literature methodology based on R_2 rates.[16] Use of the literature values would further decrease the apparent proton disorder. A strong correlation is observed between experimental K rates and subsequent calculation of theoretical K_{rig} rates from structure and diffusion information for both free and smMLCKp-bound CaM, indicating that structural models do a good job of explaining the variation in experimentally observed K rates. Several extraordinarily large K rates are observed, rates which correlate with the presence of high amide proton exchange rates as measured by Spera et al.[114]

Q values are calculated from ratios of K rates to K_{rig} rates for free and smMLCKp-bound CaM. Average Q values are 0.81 and 0.88 respectively, indicating that pervasive interproton disorder exists in both forms of the protein and that the proton relaxation network significantly rigidifies upon peptide binding. The C-terminal domain rigidifies to a greater extent than the N-terminal domain. This result agrees with methyl dynamics experiments[16] and provides a greater density and universality of dynamics probes, reinforcing the conclusion that protein rigidification is a significant event in calmodulin binding to its target domains.

The fractions of theoretical K_{rig} rates contributed by different types of protons are calculated and compared with the Q values reported by those residues. Q values are found to be positively correlated with the fraction of K_{rig} rates contributed by H_α protons and H_N protons, and negatively correlated with the fraction of K_{rig} rates contributed by H_γ protons. In other words, amide protons which primarily sense interactions with backbone protons report lower values of interproton disorder, and amide protons which primairly sense interactions with sidechain protons report higher values of interproton disorder.

This provides single-experiment confirmation of the 'rigid backbone, dynamic sidechains' model of protein dynamics.

Changes in diffusion-corrected K rates are compared with changes in Q values between free and smMLCKp-bound calmodulin and display good correlation. This suggests that isotropic diffusion estimates combined with measurements of changes in K rate between different forms of the same molecule can be used to estimate alterations in the qualities of the proton network in the absence of detailed structural information. This direction is reinforced by the fact that the largest changes in diffusion-corrected K rates are found at residues in contact with conserved hydrophobic residues on the CaM binding domain peptide.

7.3.2 Ubiquitin

K rates provide novel structural constraints for biomolecules. Intuitively, the proton layout present in protein structure should give rise to theoretical K_{rig} rates that are well-correlated with experimental K rates. Although K and K_{rig} requires 'special pleading' in the form of arguments for exceptional dynamics, amide proton exchange, and other complicated situations. This proposition can be used to compare biomolecular structures and gauge their quality with regards to one another.

Human ubiquitin η_z and K rates are measured at 27 C and pH 4.8. Rotational diffusion tensors extracted from η_{xy}/η_z ratios are found to be slightly higher than literature values, an unsurprising result in light of the fact that the sample in use is at a concentration of 2.0 mM, which may cause some transient aggregation, while the literature sample is at a concentration of 1.2 mM.

K rates are measured and compared with K_{rig} rates derived from several different NMR structures of ubiquitin. The 1D3Z structure is the oldest structure (1998) and is derived from a combination of 2727 NOE, 98 dihedral and 372 RDC restraints. The 1XQQ and 2NR2 structures date from 2005 and 2007 respectively and are based on combination of NOE restraints with molecular dynamics simulations aiming to recreate experimental S^2 order parameters resulting from R_1/R_2/heteronuclear NOE

measurements. The 2KLG structure from 2007 is derived from experiments with structure determination by means of paramagnetic solvent relaxation effects. The 2K39 structure from 2008 is constructed primarily from RDC measurements in numerous alignment media, with NOE measurements added as a secondary constraint on the model.

Each of these NMR structures consists of a large ensemble of conformers. K_{rig} rates are calculated for all conformers and the resulting correlations with experimental K rates are tabulated. Several correlations between experimental K rates and theoretical K_{rig} rates are calculated for the previously discussed CaM X-ray structures as well as for the 1UBQ X-ray structure. Superior levels of correlation are found for the high-resolution CaM X-ray structures dating from 2000 and 2007 (1.0 and 1.08 Å), while lower levels of correlation are found for the lower-resolution 1UBQ structure from 1987 (1.8 Å). This suggests that more accurate structures have higher degrees of K/K_{rig} correlation.

The 1D3Z NMR structure of ubiquitin derived from NOE and RDC restraints is found to provide the best average level of correlation for ubiquitin K data, and to display a significantly tighter distribution of correlation values. K_{rig} rates calculated from other NMR structures display lesser degrees of correlation and much wider correlation distributions. This indicates that 1D3Z provides the most parsimonious explanation for the observed K rates and that the 1D3Z ensemble has the greatest degree of internal consistency among its several conformers. The other four structures' conformers do a considerably poorer job of explaining the observed K rates and display a very wide distribution of theoretical K_{rig} rates.

Based on the previous argument and on the data provided in Chapter 5, the 1D3Z structure is the best structure available for ubiquitin. This raises the question: to what extent should NMR structures be expected to conform with the constraints presented by experimental K rates? Putting aside the problem of amide proton exchange rates, let us assume that the experimental K rates do reflect the true state of proton-proton dynamics in the molecule. (For ubiquitin at pH 4.8, this is an eminently reasonable assumption.) These proton-proton dynamics are fast, taking place over the NMR relaxation timescale of picosecond-nanosecond motions. Slower fluctuations in macromolecular structure (μs-ms timescale and larger) likely shift the fast dynamics of the protein as they take

place. This results in ps-ns dynamics creating K rates smaller than K_{rig} values through fast averaging, and these Q values being further averaged by grosser molecular motion over longer timescales. The observed Q value is that resulting from all stochastic processes taking place over the lifetime of the η_zK experiment (tens to hundreds of milliseconds).

The result of this nested averaging process is that it is not completely straightforward to enforce K rate constraints on NMR structures. It is not enough to simply attempt to construct NMR structures that agree with K rates at the ps-ns timescale. Rather, the agreement must take place over μs-ms timescales, with averaging taking place over all conformations sampled by the molecule during this time period. K rates constrain the NMR ensemble to those structures capable of demonstrating this averaging. This will likely increase the computational effort required to use K rates as an aid to structure determination.

All five structures report similar average Q levels, with the 1D3Z structure having the highest mean Q at 0.79. This value of Q indicates the presence of pervasive interproton disorder in ubiquitin on a scale similar to that perceived in free Ca^{2+}-saturated CaM. This reinforces the existence of interproton disorder in proteins. Reinforcement is provided for this argument by comparing ubiquitin Q values to NH S^2 order parameters, wherein similar features of order and disorder are observed in the structure. Some regions display Q disorder where NH S^2 order is not perceived, suggesting the presence of sidechain dynamics that are not shared by the backbone.

7.3.3 Hsc70 NBD

The Hsc70 nucleotide-binding domain (bovine Hsc70 1-386) is a challenging 44 kDa target for NMR relaxation spectroscopy. Initial preparation of protonated, ^{15}N-labeled Hsc70 NBD resulted in unacceptably broad linewidths and insufficient spectral resolution, underlining the point that spectral quality is the first priority for relaxation work on large proteins. These issues are remedied by the preparation of ^{2}H,^{13}C,^{15}N-labeled Hsc70 NBD with protonation at all exchangeable proton sites, which gives

excellent spectra with sufficient resolution for our purposes. This sample will provide the basis for future research targeting the Hsc70 NBD and an improved assignment of the Hsc70 NBD. Currently, 160 peaks are assigned with complete or some confidence, free from chemical exchange, and sufficiently resolved.

The η_zK experiment works quite well on the Hsc70 NBD. Cross-relaxation and cross-correlated relaxation processes operate as expected and result in X(t) and Y(t) data that can be processed in the same manner as CaM and ubiquitin relaxation data.

The resulting η_z rates are appropriate for a protein of Hsc70 NBD's size and concentration in the NMR sample tube (0.4 mM). Combination of η_z rates with η_{xy} rates measured from the experiment of Wang et. al. results in a rotational isotropic diffusion time measurement of 29.1 ± 0.1 ns, with sufficient detail available to characterize the molecule's considerable anisotropy of rotation as an oblate spheroid. This demonstration of precise estimation of rotational diffusion tensors for large proteins is a significant advance in NMR relaxation spectroscopy.

Proton-proton relaxation rates run at a speed comparable to ubiquitin and calmodulin in the Hsc70 NBD, despite the NBD's much larger size. This is due to the replacement of the CH protons in the sample with deuterons, which do not contribute to proton-proton zero-quantum relaxation. The resulting dimunition of the proton network compensates for the increased rotational correlation time of Hsc70 NBD. Exchangeable protons (those attached to N, O and S atoms) remain in the sample and provide the basis for observed K rates.

Comparison of Hsc70 NBD K rates with K_{rig} rates calculated from the 3HSC, 1HPM, and 2QWL structures of Hsc70 NBD yields a wealth of interesting data. K rates conform well to predictions made by K_{rig} rates calculated from structure. The K rate trendline displays excellent agreement with strong local residue-to-residue variability in Hsc70 NBD K_{rig} rates, providing strong support for the employment of the η_zK experiment in large proteins. Correlations among experimental K rates and K_{rig} rates from all three structures are strong.

The geometric mean of Hsc70 NBD Q values is 0.86, in line with the geometric mean of peptide-bound calmodulin at 0.87. This relatively high value of Q indicates

comparative rigidity in the proton network, a sensible result given the predominantly amide proton composition of the proton network and the constraints placed on amide proton geometry by secondary structure. The geometric mean is used because Q values for Hsc70 NBD include several extremely large values that cannot be explained by the presence of amide proton exchange. These values are associated with low values of K_{rig}, but are not associated with particularly low or high values of K. High values of Q are not associated with major differences in K_{rig} between 1HPM, 2QWL, and 3HSC. Large values of Q are found more frequently in beta sheet and loop residues than in alpha helix residues. Assignment problems likely underlie several high Q values, but not all high Q values lie on peaks with questionable assignments. Further work on the assignment of the Hsc70 NBD will hopefully resolve these discrepancies.

In summary, proton-proton dynamics are present throughout proteins and can be measured by straightforward η_zK NMR experiments within reasonable constraints of time and sample preparation. The analytical complexity of longitudinal proton-proton relaxation processes is greatly simplified by analysis in terms of linear combinations of longitudinal order and the equidistant virtual neighbor proton approximation.

Proton-proton dynamics change in response to binding events and provide further experimental evidence for the 'rigid backbone/mobile sidechain' picture of protein dynamics. η_zK experiments can be used to validate structural conclusions about biomolecules, and to measure the rotational diffusion behavior of larger proteins. The η_zK experiments can be performed on large proteins up to the TROSY limit of 60-70 kDa, and are used to investigate proton-proton dynamics in the 44 kDa Hsc70 NBD.

7.4 Future directions

The η_zK experiment opens up new horizons in structural biology. It provides semi-quantitative constraints on the local geometry of proteins, allowing validation of X-ray structures and NOE structural ensembles. The extreme sensitivity of K rates' r^{-6} distance dependence offers the possibility of considerable improvements in structural accuracy. Research into K trends in the vicinity of various secondary structural elements

and tertiary motifs will allow a more precise understanding of the structure and dynamics that characterize these protein sub-elements.

The η_zK experiment is especially exciting in the context of NMR structural biology, since most samples that yield good NOESY spectra should also yield good η_zK spectra. In the context of NOESY spectroscopy and biomolecular assignment, the time commitment required for an η_zK experiment does not constitute an extraordinary investment. Qualitative restraints provided by NOE crosspeaks can be supplemented by K rates which help enforce appropriate local geometry and maintain proper minimum and maximum packing of protons throughout the structure.

The combination of K rates with NOE information is especially interesting in light of the results reported for ubiquitin, in which K information is used to gauge the precision and accuracy of a number of different strucutres.

NOE crosspeaks might also be used to assist in the interpretation of K rates. If the relative intensities of NOE crosspeaks to an amide proton can be quantified in the initial-rate approximation regime[43], then these intensities can be used to divide up the K rate reported by the amide proton into the subsidiary contributions from its several neighbor protons. This will provide more detailed information regarding the dynamics of different types of protons with regard to the amide proton, and may even be able to provide information on motional models for sidechains.

Selective protein labeling presents a biochemically-oriented route to the problem of partitioning proton-proton dynamics among the numerous contributors to each K rate. Protonated amino acids can be introduced into otherwise perdeuterated growth media, resulting in proteins with CH protons on only one or a few different types of residues.

This should allow for detailed studies of the dynamics of different types of sidechains with respect to backbone amide protons, being essentially a biochemical approach to the problem, vice the spectroscopic approach detailed above. Selective labeling will be particularly interesting in the case of amino acids whose sidechains lack methyl groups and whose sidechain dynamics cannot be easily studied by existing methodology.

On a related note, K rates and Q values can provide plentiful grist for MD investigations of NMR dynamics. The analysis of K rates in terms of theoretical K_{rig}

rates and Q values was inspired by earlier MD work aimed at understanding likely values of Q in the absence of experimental information (P 1992). Now that a means of obtaining this information has been developed, the process can be turned around. Work with combinations of MD simulation and K rates will help understand what processes drive changes in K rates, where different regions of high and low K rates should be expected, and how the details of proton-proton correlation functions are assembled into the overall K rate visible at the amide proton.

Work remains to be done on the interaction of amide proton exchange processes with the η_zK experiment. In theory it should be possible to develop a detailed model of the interaction of water magnetization with the η_zK experiment and quantify the contribution of water exchange processes to relaxation. The study of unfolded or partially disordered proteins will be instructive in this regard.

The essential spectroscopy and relaxation theory of the η_zK methodology is applicable to imino protons in nucleic acids and to carbon-bound protons in both proteins and nucleic acids. Although TROSY effects are diminished in such systems, one can in theory still study cross-relaxation between the proton 'up' and proton 'down' LCLO states using the principles of the η_zK NMR experiment and incorporating the same data interpretation and fitting methods. Complications are introduced into the study of carbon-bound protons by the high density of carbon atoms in biomolecules, which introduces problems with carbon/carbon and carbon-proton/carbon-proton relaxation. These systems will provide future challenges for the budding NMR relaxation spectroscopist.

Calmodulin binding to its several calmodulin-binding domains is a structurally and biochemically well-characterized process. Enthalpic and entropic contributions to the free energy of peptide binding have been characterized for numerous peptides. Studying he proton-proton dynamics of calmodulin bound to different peptides provides an interesting opportunity to evaluate changes in proton-proton dynamics as they relate to changes in entropy on protein binding. The temperature response of the η_zK experiment presents another piece of low-hanging fruit for the interested experimentalist.

Finally, experimental work on large proteins awaits. Much work remains to be done on η_zK studies of the Hsc70 NBD. It is hoped that further experimentation, assignment work, and data analysis will provide a clear picture of the dynamics of the Hsc70 NBD and that this work will provide a precedent for an expansion of protein dynamics studies beyond the well-characterized small proteins that currently form the field's stock in trade.

The route to this point has been theoretically and experimentally challenging, but also extraordinarily rewarding. Several outstanding problems in NMR have been resolved. Numerous challenges await on the road ahead. It is my hope that this work will be of use to future spectroscopists and structural biologists, and that it will inspire further discovery in the years to come.

APPENDIX A: η_zK FITTING CODE (MATLAB)

```
% ---SmallEigXFit_lsqWrap.m---

% SmallEigXFit_lsqWrap.m : Wrapper for implicit-function
implementation of equidistant virtual neighbor proton
approximation fitting to X(t)-type symmetric reconversion
crossover data in the nzK experiment. (Weaver Zuiderweg
2008, Weaver Zuiderweg 2009)  For further information
consult Daniel S. Weaver's PhD dissertation, Proton-Proton
Dynamics In Proteins By NMR Spectroscopy (2010).

% Requires SmallEigXFit_InlineComponent.m

% InputPars are initial guesses for K and DPR, where DPR is
the dot product result between a vector spanning Liouville
population state space which describes the starting state
(normalized to magnitude 1) and a similar vector which
describes the pure undesired state.  In other words, it is
the dot product result between a vector describing the
state you start in and the state you don't want.  Perfect
magnetization preparation results in DPR = 0.

% N is the number of protons involved in the equidistant
virtual proton neighbor approximation.  tspan is the time
points involved in the fit. expX is the values of X(t)
involved in the fit.  OptimStruct contains MATLAB fitting
options for use with lsqcurvefit.

% ETA (the longitudinal cross-correlated relaxation rate)
and R (the longitudinal auto-relaxation rate) don't
actually do anything in X(t) fitting.  Set R to a small
positive number and ETA to a small negative number that is
smaller in absolute magnitude than R.

% 12/2/09 -- Dan Weaver -- dswvr@umich.edu

function
[fitpars,resnorm,residual,exitflag,output]=SmallEigXFit_lsq
Wrap(InputPars,N,ETA,R,tspan,expX,OptimStruct)
```

```
trypars=[InputPars(1) InputPars(2)];

[fitpars,resnorm,residual,exitflag,output]=lsqcurvefit(@(tr
ypars,tspan)
SmallEigXFit_InlineComponent(trypars,tspan,N,ETA,R),trypars
,tspan,expX,[],[],OptimStruct);

end

% ---SmallEigXFit_InlineComponent.m---

 % SmallEigXFit_InlineComponent.m : Guts of implicit-
function implementation of equidistant virtual neighbor
proton approximation fitting to X(t)-type symmetric
reconversion crossover data in the nzK experiment. (Weaver
Zuiderweg 2008, Weaver Zuiderweg 2009)  For further
information consult Daniel S. Weaver's PhD dissertation,
Proton-Proton Dynamics In Proteins By NMR Spectroscopy
(2010).

% Requires SmallEigXFit_lsqWrap.m

% InputPars are initial guesses for K and DPR, where DPR is
the dot product result between a vector spanning Liouville
population state space which describes the starting state
(normalized to magnitude 1) and a similar vector which
describes the pure undesired state.  In other words, it is
the dot product result between a vector describing the
state you start in and the state you don't want.  Perfect
magnetization preparation results in DPR = 0.

% N is the number of protons involved in the equidistant
virtual proton neighbor approximation.  tspan is the time
points involved in the fit. expX is the values of X(t)
involved in the fit.  OptimStruct contains MATLAB fitting
options for use with lsqcurvefit.

% ETA (the longitudinal cross-correlated relaxation rate)
and R (the longitudinal auto-relaxation rate) don't
actually do anything in X(t) fitting.  Set R to a small
positive number and ETA to a small negative number that is
smaller in absolute magnitude than R.

% 12/2/09 -- Dan Weaver -- dswvr@umich.edu


function XRET =
```

```
SmallEigXFit_InlineComponent(trypars,tspan,N,ETA,R)

K1=trypars(1);
DPR=trypars(2);
aAR=(R+ETA);
bAR=(R-ETA);
X=K1/N; % partition your overall K rate
% the "low" nonideal (the one you didn't want)
low_nonideal = DPR / 2^(N/2);
% the "high" nonideal (the one you do want)
high_nonideal = sqrt(2^-N*(1-DPR^2));

a2aSE_all=zeros(1,numel(tspan));
a2bSE_all=zeros(1,numel(tspan));
b2aSE_all=zeros(1,numel(tspan));
b2bSE_all=zeros(1,numel(tspan));

SmallEig_a0nonideal=[high_nonideal;low_nonideal];
% Bscale not needed here because of how XFit works
SmallEig_b0nonideal=[low_nonideal;high_nonideal];

for z=0:N
    % the z=0 and N+1 cases
    if z==0
        SmallEig=zeros(2);
        SmallEig(1,1)=-aAR;
        SmallEig(2,2)=-bAR;
    else  % more interesting cases
        SmallEig=zeros(2);
        SmallEig=-[aAR+(z)*X -(z)*X;-(N+1-z)*X bAR+(N+1-
z)*X];
    end

    % solve this tier via diagonalization
    [v,d]=eig(SmallEig);
    aC=v\SmallEig_a0nonideal;
    bC=v\SmallEig_b0nonideal;

    % aXpl(1,:) : a2a (I-type auto)
    % aXpl(2,:) : a2b (II-type cross)
    % bXpl(1,:) : b2a (III-type auto)
    % bXpl(2,:) : b2b (IV-type cross)
    aXpl=v*((aC*ones(size(tspan))).*(exp(diag(d)*tspan)));
    bXpl=v*((bC*ones(size(tspan))).*(exp(diag(d)*tspan)));

%     if z==0
%         a2acofactor = 1;     % this code accounts for the
relative degeneracy of each state at different z tiers
```

```
%         b2bcofactor = 1;     % it's not needed for the
XFit code because a2a/a2b/b2a/b2b decays cancel out the
relative degeneracy of each state
%     else
%         a2acofactor = z;     % the ratio of the number of
states experiencing each type of relaxation at each z tier
%         b2bcofactor = N-z+1; % the true ratio is
nchoosek(N,z-1) / nchoosek(N,z) = N-z+1 / z
%     end

    % build this tier's contribution to the symmetric
reconversion quartet
    a2aSE=aXpl(1,:);
    a2bSE=aXpl(2,:);
    b2aSE=bXpl(1,:);
    b2bSE=bXpl(2,:);

    % add this tier's contribution to the total quartet —
this is doing the sum in Eqn. (2.32).
    a2aSE_all=a2aSE+a2aSE_all;
    a2bSE_all=a2bSE+a2bSE_all;
    b2aSE_all=b2aSE+b2aSE_all;
    b2bSE_all=b2bSE+b2bSE_all;

end

% construct X(t) from the quartet
XRET=(sqrt(abs(a2bSE_all.*b2aSE_all./a2aSE_all./b2bSE_all))
.*sign(a2bSE_all));

end

% ---SmallEigYFit_lsqWrap.m---

% SmallEigYFit_InlineComponent.m : Guts of implicit-
function implementation of equidistant virtual neighbor
proton approximation fitting to Y(t)-type symmetric
reconversion crossover data in the nzK experiment. (Weaver
Zuiderweg 2008, Weaver Zuiderweg 2009)  For further
information consult Daniel S. Weaver's PhD dissertation,
Proton-Proton Dynamics In Proteins By NMR Spectroscopy
(2010).

% Requires SmallEigXFit_lsqWrap.m
% Requires SmallEigXFit_InlineComponent.m
% Requires SmallEigYFit_InlineComponent.m

% InputPars are initial guesses for nZ and Bscale, where
```

```
Bscale is the initial ratio of IV-type magnetization
intensity to I-type magnetization intensity.  Perfect
equivalency of preparation leads to Bscale = 1.

% N is the number of protons involved in the equidistant
virtual proton neighbor approximation.  K1 and DPR are
values derived from an earlier fit to X(t) data acquired as
part of the nzK experiment.  nZ fitting is dependent on the
value of K1 and DPR because K1 and DPR control cross-
relaxation equilibration between NzHa and NzHb states.  All
Y(t) fits must be conducted after an X(t) fit.  tspan is
the time points involved in the fit. expY is the values of
X(t) involved in the fit.  OptimStruct contains MATLAB
fitting options for use with lsqcurvefit.

% R (the longitudinal auto-relaxation rate) doesn't
actually do anything in Y(t) fitting due to symmetric
reconversion.  Set R to a small positive number that is
larger in absolute magnitude than your expected nZ.

% 12/2/09 -- Dan Weaver -- dswvr@umich.edu


function
[fitpars,resnorm,residual,exitflag,output]=SmallEigYFit_lsq
Wrap(InputPars,N,K1,DPR,R,tspan,expY,OptimStruct)

trypars=[InputPars(1) InputPars(2)];

% precalculate nchoosek to speed up iteration

CoeffColumns=zeros(2*(N+1),1);

CoeffColumns(1)=1;  % number of a-type states for the 0th
energy level tier
CoeffColumns(2)=1;  % number of b-type states for the N+1th
energy level tier (organized this way to be more compact)


for j=3:2:(2*(N+1)-1)
    z=(j-1)/2;
    CoeffColumns(j)=nchoosek(N,z);  % number of a-type
states for this zth tier
    CoeffColumns(j+1)=nchoosek(N,z-1); % number of b-type
states for this zth tier
end

CoeffColumns=(CoeffColumns/2^(N/2))*ones(1,numel(tspan));
```

```
[fitpars,resnorm,residual,exitflag,output]=lsqcurvefit(@(tr
ypars,tspan)
SmallEigYFit_InlineComponent(trypars,tspan,N,K1,DPR,R,Coeff
Columns),trypars,tspan,expY,[],[],OptimStruct);

end

% ---SmallEigYFit_InlineComponent.m---

function YRET =
SmallEigYFit_InlineComponent(trypars,tspan,N,K1,DPR,R,Coeff
Columns)

ETA=trypars(1);
Bscale=trypars(2);
aAR=(R+ETA);
bAR=(R-ETA);
X=K1/N;
% the "low" nonideal (the one you didn't want)
low_nonideal = DPR / 2^(N/2);
% the "high" nonideal (the one you do want)
high_nonideal = sqrt(2^-N*(1-DPR^2));

a2aSE_all=zeros(1,numel(tspan));
b2bSE_all=zeros(1,numel(tspan));
%YRET=zeros(1,numel(tspan));

SmallEig_a0nonideal=[high_nonideal;low_nonideal];
SmallEig_b0nonideal=Bscale*[low_nonideal;high_nonideal];

for z=0:N
    % the z=0 and N+1 cases
    if z==0
        SmallEig=zeros(2);
        SmallEig(1,1)=-aAR;
        SmallEig(2,2)=-bAR;
    else  % more interesting cases
        SmallEig=zeros(2);
        SmallEig=-[aAR+(z)*X -(z)*X;-(N+1-z)*X bAR+(N+1-
z)*X];
    end


    [v,d]=eig(SmallEig);
    aC=v\SmallEig_a0nonideal;
    bC=v\SmallEig_b0nonideal;
```

```
    aXpl=v*((aC*ones(size(tspan))).*(exp(diag(d)*tspan)));
    bXpl=v*((bC*ones(size(tspan))).*(exp(diag(d)*tspan)));

    qqq=(z+1)*2;

        a2acofactor = CoeffColumns(qqq-1,1);      % this
code accounts for the relative degeneracy of each state at
different z tiers
        b2bcofactor = CoeffColumns(qqq,1);        % it's
not needed for the XFit code because a2a/a2b/b2a/b2b decays
cancel out the relative degeneracy of each state
                                                  % the ratio
of the number of states experiencing each type of
relaxation at each z tier
                                                  % the true
ratio is
                                                  %
nchoosek(N,z-1) / nchoosek(N,z) = N-z+1 / z

    a2aSE=a2acofactor*aXpl(1,:);
    b2bSE=b2bcofactor*bXpl(2,:);

    a2aSE_all=a2aSE+a2aSE_all;
    b2bSE_all=b2bSE+b2bSE_all;


end

YRET=b2bSE_all./a2aSE_all;

end


% ---FortranDataMakerHsc70NBDADP092209.m---

% MATLAB constructor program for FORTRAN 90-formatted data
sets.  The output of this program
(NZK_Hsc80ADP_V30C_800_092209.dat) is suitable for feeding
to the (compiled) FORTRAN 90 code detailed in Appendix B.

% The file header looks like this:

%NPEAKS
%NPOINTS
%NBOOT
%SIZE(NLIST)
%SEED
```

```
%IDENT

%NLIST

%TIME SERIES

%X1
%Y1

%X2
%Y2

%...

% The desired X(t) and Y(t) data goes into
XKHsc70ADP30C0921Edit and YKHsc70ADP30C0921Edit here.  t
data goes into tspanKHsc70ADP30C0921Edit, in units of
seconds.

% You'll need to set three variables describing the
parameters of the fit:

% NBOOT: the number of bootstrap trials.  More bootstrap
trials means a bigger bootstrap distribution, but more CPU
time.

% SEED: the seed for your PRNG.

% NLIST: the set of N values (number of neighbor protons)
you want to try in the equidistant virtual neighbor proton
approximation.  More NLIST can produce a more accurate fit
at the cost of CPU time.

% The other parts of the header are as follows:

% NPEAKS: the total number of peaks you're analyzing.
% NPOINTS: the number of time data points you acquired.
% SIZE(NLIST): the number of elements in NLIST
% IDENT: the residue numbers of the peaks you're analyzing.
(Column)
% NLIST: the N values you're using
% TIME SERIES: the time points (in seconds) you're using.

% 12/2/09 -- Dan Weaver -- dswvr@umich.edu

NBOOT=200  % Number of bootstrap trials.
SEED=7777
```

```
NLIST=[1:60]


fid=fopen('NZK_Hsc70ADP_V30C_800_092209.dat','w')

size(XKHsc70ADP30C0921Edit,1)
size(XKHsc70ADP30C0921Edit,2)

fprintf(fid,'%d\n%d\n',size(XKHsc70ADP30C0921Edit,1),size(X
KHsc70ADP30C0921Edit,2))
fprintf(fid,'%d\n',NBOOT);
fprintf(fid,'%d\n',numel(NLIST));
fprintf(fid,'%d\n',SEED);
fprintf(fid,'\n');

for k=1:size(XKHsc70ADP30C0921Edit,1)
    fprintf(fid,'%d\n',k)       % write an identifier
end
fprintf(fid,'\n');

% the time series
for j=1:numel(NLIST)
fprintf(fid,'%d ',NLIST(j));
end
fprintf(fid,'\n');
fprintf(fid,'\n');

% the time series
for j=1:numel(tspanKHsc70ADP30C0921Edit)
fprintf(fid,'%8.6f ',tspanKHsc70ADP30C0921Edit(j));
end
fprintf(fid,'\n');
fprintf(fid,'\n');

% now do the pack of data points

for k=1:size(XKHsc70ADP30C0921Edit,1)
    % do X
    %
    for j=2:size(XKHsc70ADP30C0921Edit,2)
        fprintf(fid,'%8.6f ',XKHsc70ADP30C0921Edit(k,j));
    end
    fprintf(fid,'\n');

    % do X
    for j=2:size(YKHsc70ADP30C0921Edit,2)
        fprintf(fid,'%8.6f ',YKHsc70ADP30C0921Edit(k,j));
    end
```

```
    fprintf(fid,'\n');
    fprintf(fid,'\n');
end

% The output from this program looks like:

% 319
% 17
% 100
% 60
% 7777
%
% 1
% 2
% 3
% 4
% 5
% 6
% 7
% 8
% ... (more IDENT entries)
% 317
% 318
% 319

%1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

%0.010000 0.020000 0.050000 0.100000 0.150000 0.150000
0.200000 0.250000 0.300000 0.350000 0.400000 0.450000
0.450000 0.500000 0.500000 0.600000 0.700000

%0.029385 0.117485 0.180005 0.310843 0.457687 0.474893
0.583361 0.678817 0.717736 0.765287 0.855465 0.896885
0.832786 0.876182 0.888188 0.958358 0.929613
%1.115240 1.015928 0.891429 0.900391 0.863243 0.857982
0.817014 0.862835 0.875419 0.768462 0.816010 0.741965
0.796200 0.787729 0.869371 0.777544 0.822818

%0.023568 0.029613 0.088286 0.148400 0.217747 0.213438
0.277567 0.322326 0.363303 0.404688 0.470356 0.494365
0.515786 0.528648 0.571213 0.594086 0.691085
%1.005916 0.990330 0.890058 0.821870 0.830421 0.820658
0.794776 0.836747 0.873896 0.727304 0.706570 0.685583
0.753597 0.738473 0.819705 0.688665 0.635857
```

APPENDIX B: η_zK FITTING CODE (FORTRAN 90)

```
---kfit042409SpaceConserve.f90---

program kft_bigmod

! kfit042409SpaceConserve.f90

! Usage format is kfit <InputFile> <OutputFile>
<MCOutputFile>

! Super-wrapper for equidistant virtual neighbor proton
approximation (EVNPA) fitting to the nzK experiment.
(Weaver Zuiderweg 2008, Weaver Zuiderweg 2009)  For further
information consult Daniel S. Weaver's PhD dissertation,
Proton-Proton Dynamics In Proteins By NMR Spectroscopy
(2010).  Handles writing of fit data to disk and iteration
over different choices of EVNPA proton number N.

! Requires kfitCleanLmdifWrap021809.f90 or something
similar.

! ---FILE FORMATTING INFORMATION---

! <InputFile> formatting is as documented under
FortranDataMakerHsc70NBDADP092209.m in Appendix A.

! ---

! <OutputFile> contains fitting information and is of the
form:

!         1         1    0.100000E+01    0.382229E+01
0.327665E+00   -0.377291E+00    0.102410E+01
0.100000E+01    0.989643E+00    0.100000E+01    0.760935E-
01
!         1         1    0.200000E+01    0.554113E+01
0.228680E+00   -0.443040E+00    0.103332E+01
0.100000E+01    0.592655E+00    0.100000E+01    0.697630E-
01
```

```
!         1         1    0.300000E+01    0.647455E+01
0.162364E+00   -0.484805E+00    0.103951E+01
0.100000E+01    0.380801E+00    0.100000E+01    0.655972E-
01
!         1         1    0.400000E+01    0.701480E+01
0.116151E+00   -0.511919E+00    0.104377E+01
0.100000E+01    0.255653E+00    0.100000E+01    0.627689E-
01
!         1         1    0.500000E+01    0.733075E+01
0.831423E-01   -0.529570E+00    0.104674E+01
0.100000E+01    0.177044E+00    0.100000E+01    0.608216E-
01
!         1         1    0.600000E+01    0.750864E+01
0.592205E-01   -0.540905E+00    0.104881E+01
0.100000E+01    0.125581E+00    0.100000E+01    0.594753E-
01

! ...

!       221         2    0.460000E+02    0.226436E+01
0.357660E-01   -0.304505E+00    0.918647E+00
0.100000E+01    0.511199E-02    0.100000E+01    0.371853E-
01
!       221         2    0.470000E+02    0.226295E+01
0.358358E-01   -0.304461E+00    0.918638E+00
0.100000E+01    0.512782E-02    0.100000E+01    0.371878E-
01
!       221         2    0.480000E+02    0.226160E+01
0.359027E-01   -0.304419E+00    0.918630E+00
0.100000E+01    0.514301E-02    0.100000E+01    0.371902E-
01
!       221         2    0.490000E+02    0.226031E+01
0.359667E-01   -0.304379E+00    0.918622E+00
0.100000E+01    0.515760E-02    0.100000E+01    0.371926E-
01

! which is arranged (double tab separated) as follows:

! MCSETID PEAKNUM N FITK FITDPR FITNZ FITBSCALE XFITOUT
XFITRESID YFITOUT YFITRESID

! MCSETNUM: the unique ID of this MC trial.  You can cross
reference with the data output's .MCout file in order to
find out exactly what Monte Carlo perturbation of the
original data set was the target for this fit.

! PEAKNUM: the number of the peak you are fitting to.
```

```
! N: the choice of N for this fit.

! FITK: the value of K (proton-proton cross-relaxation
rate) resulting from the fit.

! FITDPR: the value of DPR (the dot product result; see
MATLAB code for full explanation) resulting from the fit.

! FITNZ: the value of nZ (longitudinal NH/N cross-
correlated relaxation rate) resulting from the fit.

! FITBSCALE: the value of Bscale (the efficiency of
observing NzHb-type magnetization relative to observing
NzHa-type magnetization) resulting from the fit.

! XFITOUT: the exit condition of lmdif1 (see MINPACK
documentation for lmdif1) for the X(t) segment of the fit.

! XFITRESID: the residual of the fit to X(t).  Note that
this is the sole judge of fit quality as the protocol is
currently implemented.  The best (lowest) XFITRESID value
for a given MCSETNUM is taken as the best fit to that
MCSETNUM.

! YFITOUT: the exit condition of lmdif1 (see MINPACK
documentation for lmdif1) for the Y(t) segment of the fit.

! YFITRESID: the residual of the fit to Y(t) resulting from
the parameters established by the previous fit to X(t).

! ---

! <MCOutputFile> contains Monte Carlo perturbation
information and is of the form:

!         1   3   3   1   5   8  15   8  14  12   7  14   1
16  13  13   5   1
!         2   3   7  12  11   7   8   9   2   4   5  16  10
5  14   3  16   8
!         3   4   2  14   6   1  12   1   8   4   8  12  12
15   8   5   7   3
!         4  17  12   2  13   6  13  10   6   7   8  16  17
3   5  11  11   1
!         5   2   5  13   5  14  16   5   3  17  17  13  17
11   5  17  16   5
!         6   9  17   3  15  10   5   3   3   6  10   4  11
17  12  17  12   1
```

```
!          7   1  14   4  10  14  12   7  12   4   6  17   3
7   4   7  13  12
!

! which is arranged (single tab separated) as follows:

! MCSETID MCSETNUMBERS...

! MCSETID is the unique ID associated with an MC trial.

! MCSETNUMBERS contains NPOINTS integers, each of which
denotes a choice from the orignal data set used to build
the bootstrap ensemble member associated with MCSETID.
Cross-referencing the point indices from MCSETNUMBERS with
the original X(t)/Y(t) data referred to by the PEAKNUM
associated with this MCSETID will allow the interested
individual to reconstruct the exact fitting target of any
MCSETID bootstrap iteration.

! This fitting problem is embarrassingly parallel.  I ran
out of time and couldn't implement parallel computing here.
If you feel an insane urge towards productivity, try
parallelizing this.

! 12/2/09 -- Dan Weaver -- dswvr@umich.edu


implicit none
integer, parameter :: fid = 51, fidOutputList = 61,
fidResid = 62
integer :: qqq, iii, jjj, kkk, info, lwa, m, n, NNNSize,
NResid, NMC, rngSeed
integer, allocatable :: iwa(:), NNNList(:), residList(:)
double precision, allocatable :: X(:,:), Y(:,:)
real :: dummy
integer, parameter :: lstate=16
integer :: j, genid, subid, seed, lseed, state(lstate),
randinit_info, rand_info, MCcounter
double precision :: minval, maxval
double precision, allocatable :: thisX(:), thisY(:),
tspan(:), thisTspan(:), MonteCarloRands(:)
integer, allocatable :: MonteCarloRandsCeil(:)
! output arrays for kfit_wrapper:
double precision, allocatable :: OutputList(:,:),
OutputListXResid(:,:), &
OutputListYResid(:,:)

real :: tarray(2), result
```

```
external kfit_wrapper, dtime, dranduniform, drandinitialize

character :: InputFileName*200
character :: OutputFileName*200
character :: ResidFileName*200
integer :: i, iargc, nargs
nargs = iargc()

if (nargs .eq. 3) then

call getarg( 1, InputFileName)
call getarg( 2, OutputFileName)
call getarg( 3, ResidFileName)
write(*,*) InputFileName
write(*,*) OutputFileName
write(*,*) ResidFileName

else

write(*,*) "Format is kfit <InputFile> <OutputFile>
<MCOutputFile>"
stop

end if

!!--------------
!! START READING HERE
!!--------------

open(unit=fid, file=InputFileName, status='old')
read(fid,"I4") NResid
read(fid,"I4") m
read(fid,"I4") NMC
read(fid,"I4") NNNSize
read(fid,"I8") rngSeed
read(fid,*)

allocate(residList(1:NResid))

do jjj=1,NResid
read(fid,"I4") residList(jjj)
end do

allocate(NNNlist(1:NNNsize))

! column major order makes it the right way to do X(m,
Nresid)
```

```
allocate(tspan(1:m),thisTspan(1:m),X(1:m,1:NResid),Y(1:m,1:
NResid),&
thisX(1:m),thisY(1:m))
allocate(OutputList(9,NNNsize),OutputListXResid(m,NNNSize),
&
OutputListYResid(m,NNNSize))

read(fid,*)
read(fid,*) NNNlist
read(fid,*)
read(fid,*) tspan
read(fid,*)

do jjj=1,Nresid
read(fid,*) X(:,jjj)
read(fid,*) Y(:,jjj)
read(fid,*)
end do

close(fid)

genid = 1
subid = 1
seed = rngSeed
lseed = 1
minval=0.d0
maxval=DBLE(m)
allocate(MonteCarloRands(1:m),MonteCarloRandsCeil(1:m))
call
drandinitialize(genid,subid,seed,lseed,state,lstate,randini
t_info)

!! --------------
!! END READING BIT
!! --------------

open(unit=fidOutputList, file=OutputFileName)
open(unit=fidResid, file=ResidFileName)

MCcounter=1   ! use this as an index

! -- start running through residues

do iii=1,NResid

do qqq=1,NMC
```

```
call
dranduniform(m,minval,maxval,state,MonteCarloRands,rand_inf
o)
MonteCarloRandsCeil=int(ceiling(MonteCarloRands))

thisX=X(MonteCarloRandsCeil,iii)
thisY=Y(MonteCarloRandsCeil,iii)
thisTspan=tspan(MonteCarloRandsCeil)

! -- start running through NNNlist

call
kfit_wrapper(m,thisTspan,thisX,thisY,NNNsize,NNNlist,Output
List,&
OutputListXResid,OutputListYResid)

do jjj=1,NNNsize
write(fidOutputList,FMT="I10.1",ADVANCE="NO") MCcounter
write(fidOutputList,FMT="I10.1",ADVANCE="NO") iii
do kkk=1,9
write(fidOutputList,FMT="E16.6",ADVANCE="NO")
OutputList(kkk,jjj)
end do
! now advance
write(fidOutputList,*,ADVANCE="YES")
end do

! use fidResid as a repository for your MC point
information
write(fidResid,FMT="I10.1",ADVANCE="NO") MCcounter
do kkk=1,m
write(fidResid,FMT="I4.1",ADVANCE="NO")
MonteCarloRandsCeil(kkk)
end do
! now advance
write(fidResid,*,ADVANCE="YES")

MCcounter=MCcounter+1
end do
end do

deallocate(residList,NNNlist,tspan,thisTspan,X,Y,thisX,this
Y,OutputList,OutputListXResid,OutputListYResid,MonteCarloRa
nds,MonteCarloRandsCeil)

close(fidOutputList)
close(fidResid)
```

```
end program kft_bigmod

! ----------------------------------------
```

```
---kfitCleanLmdifWrap021809.f90---

module kfit_data_module

implicit none

integer :: LWORK,NNN,NTpoints,FullSysSize
double precision :: workXK, workDPR
double precision, allocatable :: tspan(:), aXpl(:,:),
bXpl(:,:), EndaXpl(:,:), EndbXpl(:,:), &
a2a(:), a2b(:), b2a(:), b2b(:), WORK(:)
double precision, allocatable :: commonX(:), commonY(:),
CoeffColumns(:)

end module

!-------

subroutine
kfit_wrapper(m,timepoints,thisX,thisY,NNNsize,NNNlist, &
OutputList,OutputListXResid,OutputListYResid)

use kfit_data_module

implicit none

! kfitCleanLmdifWrap021809.f90

! Wrapper for equidistant virtual neighbor proton
approximation (EVNPA) fitting to the nzK experiment.
(Weaver Zuiderweg 2008, Weaver Zuiderweg 2009)  For further
information consult Daniel S. Weaver's PhD dissertation,
Proton-Proton Dynamics In Proteins By NMR Spectroscopy
(2010).  Handles X(t) and Y(t) calls to lmdif1 function
based on the t, X(t), and Y(t) data it's been handed by
kfit0424SpaceConserve.  Very similar in concept to
SmallEigXFit_lsqWrap.m and SmallEigYFit_lsqWrap.m.

! Requires kfit0424SpaceConserve or something similar.
! Requires BLAS Basic Linear Algebra Subprograms.
! Requires LAPACK.
! Requires MINPACK.


integer :: jjj, kkk, zzz, Kinfo, nzinfo, lwa, m, n,
NNNsize, NNNlist(NNNsize)
double precision :: Kpars(2), nzpars(2), tol, dummyXsum,
dummyYsum
```

```
double precision ::
timepoints(m),thisX(m),thisY(m),OutputList(9,NNN),OutputLis
tXResid(m,NNN),OutputListYResid(m,NNN)
integer, allocatable :: iwa(:)
double precision, allocatable :: Xout(:), Yout(:),wa(:)

real :: tarray(2), result

!integer :: factorial
double precision :: enorm, dpmpar, dsum, factorial
external :: enorm, kftfcn, nzftfcn, lmdif1, dpmpar, dsum,
factorial

tol=dsqrt(dpmpar(1))

n=2
NTpoints=m
LWA=(NTpoints*2+5*2+NTpoints)
LWORK=400 ! provisional

allocate(iwa(1:2),wa(1:LWA),WORK(1:LWORK))
allocate(tspan(1:m), commonX(1:m), commonY(1:m), Xout(1:m),
Yout(1:m),  aXpl(1:2,1:m), bXpl(1:2,1:m), &
EndaXpl(1:2,1:m), EndbXpl(1:2,1:m), a2a(1:m), a2b(1:m),
b2a(1:m), b2b(1:m))

!R=2.d0
!ETA=-1.d0

! set up residue/time points information

commonX=thisX
commonY=thisY
tspan=timepoints

! -- start running through NNNlist

do jjj=1,NNNsize

NNN=NNNlist(jjj)
FullSysSize=2*(NNN+1) ! number of energy levels in an NNN-
proton system

allocate(CoeffColumns(1:FullSysSize))
CoeffColumns(1)=1.d0/(factorial(1.d0*NNN))
CoeffColumns(2)=1.d0/(factorial(1.d0*NNN))
```

```
do kkk=3,FullSysSize-1,2
zzz=(kkk-1)/2;

CoeffColumns(kkk)=1.d0/(factorial(dble(NNN-
zzz))*factorial(dble(zzz)))
CoeffColumns(kkk+1)=1.d0/(factorial(dble(NNN-(zzz-
1)))*factorial(dble(zzz-1)))
end do

Kpars=(/1.d0, 0.d0/)
wa=0.d0
iwa=0

call
lmdif1(kftfcn,NTpoints,2,Kpars,Xout,tol,Kinfo,iwa,wa,lwa)
nzpars=(/-1.d0, 1.d0/)
wa=0.d0
iwa=0
workXK=Kpars(1)
workDPR=Kpars(2)

call
lmdif1(nzftfcn,NTpoints,2,nzpars,Yout,tol,nzinfo,iwa,wa,lwa
)

dummyXsum=dsum(NTpoints,Xout*Xout,1)
dummyYsum=dsum(NTpoints,Yout*Yout,1)
OutputList(:,jjj)=(/NNN*1.d0,Kpars(1)*NNN,Kpars(2),nzpars(1
),nzpars(2),Kinfo,dummyXsum,nzinfo,dummyYsum/)
OutputListXResid(:,jjj)=Xout
OutputListYResid(:,jjj)=Yout
deallocate(CoeffColumns)
end do

deallocate(iwa,wa,WORK,tspan,commonX,commonY,Xout,Yout,aXpl
,bXpl,EndaXpl,EndbXpl,a2a,a2b,b2a,b2b)

end subroutine kfit_wrapper

! ----------------------------------------

subroutine kftfcn(m,n,x,fvec,iflag)

use kfit_data_module
implicit none

! Guts for equidistant virtual neighbor proton
approximation (EVNPA) fitting to the nzK experiment.
```

```
(Weaver Zuiderweg 2008, Weaver Zuiderweg 2009)  For further
information consult Daniel S. Weaver's PhD dissertation,
Proton-Proton Dynamics In Proteins By NMR Spectroscopy
(2010).  Handles X(t) calls from lmdif1 function.  Very
similar in concept to SmallEigXFit_InlineComponent.m.

! Requires kfit0424SpaceConserve or something similar.
! Requires BLAS Basic Linear Algebra Subprograms.
! Requires LAPACK.
! Requires MINPACK.

logical :: SELECT

integer :: m,n,iflag, jjj,kkk,ILO,IHI,INFODGEBAL,INFODGEHRD
,&
INFO,DUMMYINT,INFODORGHR,INFODHSEQR,INFODTREVC

double precision ::
aAR,bAR,XK,DPR,VL,DUMMYDP,fakeR,fakeETA,ALPHA_SCALE, &
loni, hini, MURFIN, zzz, NNNDP, x(n), fvec(m)

double precision :: SmallEig(2,2), TAU, QSmallEig(2,2),
ScaleQSmallEig(2,2), WR(2), &
WI(2), WORKDTREVC(6), a0ni(2), b0ni(2), aC(2), bC(2),
aD(2), aDU, aDL, bD(2), bDU, bDL

double precision :: DDOT, DNRM2
external :: DGEHRD, DORGHR, DHSEQR, DTREVC, DGTSV, DDOT,
DNRM2
intrinsic :: MIN, MAX, DSQRT, DEXP, DSIGN, DABS

!     set up autorelaxation

ALPHA_SCALE=1.0d0
MURFIN=0.0d0

! it doesn't matter what these are when you're doing the K
fitting...
fakeR=2.d0
fakeETA=-1.d0

aAR=fakeR+fakeETA
bAR=fakeR-fakeETA

XK=x(1)       ! K1 / N
DPR=x(2)      ! dot product result

!initialize a2a etc
```

```
a2a=0.d0
a2b=0.d0
b2a=0.d0
b2b=0.d0

!write(*,*) FullSysSize

do kkk=1,FullSysSize-1,2

if (kkk .eq. 1) then

! set up the all-a and all-b states
SmallEig=0.d0
SmallEig(1,1)=-aAR
SmallEig(2,2)=-bAR

else
! do a more interesting state

zzz=(kkk-1.d0)/2.d0
SmallEig(1,1) = -(aAR+zzz*XK)
SmallEig(1,2) = zzz*XK
SmallEig(2,1) = (NNN+1.d0-zzz)*XK
SmallEig(2,2) = -(bAR+(NNN+1.d0-zzz)*XK)

endif

ILO=1
IHI=2

call DGEHRD( 2, ILO, IHI, SmallEig, 2, TAU, WORK, LWORK,
INFODGEHRD)

QSmallEig=SmallEig

call DORGHR( 2, ILO, IHI, QSmallEig, 2, TAU, WORK, LWORK,
INFODORGHR)

!write(*,*) "INFODORGHR: ", INFODORGHR

call DHSEQR( 'S', 'V', 2, ILO, IHI, SmallEig, 2, WR, WI,
QSmallEig, 2, WORK, LWORK, INFODHSEQR )

!write(*,*) "INFODHSEQR: ", INFODHSEQR

call DTREVC( 'R', 'B', SELECT, 2, SmallEig, 2, VL, 1,
QSmallEig, 2, 2, 2, WORKDTREVC, INFODTREVC )
```

```
!write(*,*) "INFODTREVC: ", INFODTREVC

ScaleQSmallEig(:,1)=QSmallEig(:,1)/dnrm2(2,QSmallEig(:,1),1
)
ScaleQSmallEig(:,2)=QSmallEig(:,2)/dnrm2(2,QSmallEig(:,2),1
)

loni=DPR / (2.d0**(NNN/2.d0))
hini=dsqrt(dabs((2.d0**-NNN)*(1.d0-DPR**2.d0)))

a0ni=(/hini, loni/)
b0ni=(/loni, hini/)

aC=a0ni
bC=b0ni

aDL=ScaleQSmallEig(2,1)
aD(1)=ScaleQSmallEig(1,1)
aD(2)=ScaleQSmallEig(2,2)
aDU=ScaleQSmallEig(1,2)

! Dx is destroyed
bDL=aDL
bD=aD
bDU=aDU

call DGTSV(2,1,aDL,aD,aDU,aC,2,INFO)
call DGTSV(2,1,bDL,bD,bDU,bC,2,INFO)

do jjj=1,NTpoints
aXpl(:,jjj)=aC*dexp(tspan(jjj)*WR)
bXpl(:,jjj)=bC*dexp(tspan(jjj)*WR)
end do

call DGEMM('N','N', 2, NTpoints, 2, ALPHA_SCALE,
ScaleQSmallEig, 2, aXpl, 2, MURFIN, EndaXpl, 2)
call DGEMM('N','N', 2, NTpoints, 2, ALPHA_SCALE,
ScaleQSmallEig, 2, bXpl, 2, MURFIN, EndbXpl, 2)

a2a=a2a+EndaXpl(1,:)
a2b=a2b+EndaXpl(2,:)
b2a=b2a+EndbXpl(1,:)
b2b=b2b+EndbXpl(2,:)

end do
```

```
fvec=dsqrt(dabs(b2a*a2b/a2a/b2b))*dsign(ALPHA_SCALE,b2a)-
commonX


end subroutine kftfcn


! ------------------------------

! ----------------------------------------

subroutine nzftfcn(m,n,x,fvec,iflag)

use kfit_data_module
implicit none

! Guts for equidistant virtual neighbor proton
approximation (EVNPA) fitting to the nzK experiment.
(Weaver Zuiderweg 2008, Weaver Zuiderweg 2009)  For further
information consult Daniel S. Weaver's PhD dissertation,
Proton-Proton Dynamics In Proteins By NMR Spectroscopy
(2010).  Handles Y(t) calls from lmdif1 function.  Very
similar in concept to SmallEigYFit_InlineComponent.m.

! Requires kfit0424SpaceConserve or something similar.
! Requires BLAS Basic Linear Algebra Subprograms.
! Requires LAPACK.
! Requires MINPACK.

logical :: SELECT

integer :: m,n,iflag,jjj,kkk,ILO,IHI,INFODGEBAL,INFODGEHRD
,&
INFO,DUMMYINT,INFODORGHR,INFODHSEQR,INFODTREVC

double precision ::
aAR,bAR,XK,DPR,VL,DUMMYDP,ALPHA_SCALE,R,Bscale,ETA, &
loni, hini, MURFIN, zzz, NNNDP, x(n), fvec(m)

double precision :: SmallEig(2,2), TAU, QSmallEig(2,2),
ScaleQSmallEig(2,2), WR(2), &
WI(2), WORKDTREVC(6), a0ni(2), b0ni(2), aC(2), bC(2),
aD(2), aDU, aDL, bD(2), bDU, bDL

double precision :: DDOT, DNRM2
external :: DGEHRD, DORGHR, DHSEQR, DTREVC, DGTSV, DDOT,
DNRM2
intrinsic :: MIN, MAX, DSQRT, DEXP, DSIGN, DABS
```

```
!     set up autorelaxation

ALPHA_SCALE=1.0d0
MURFIN=0.0d0

ETA=x(1)
Bscale=x(2)
R=2.d0!*ETA

aAR=R+ETA
bAR=R-ETA

XK=workXK       ! K1 / N
DPR=workDPR      ! dot product result

!initialize a2a etc

a2a=0.d0
a2b=0.d0
b2a=0.d0
b2b=0.d0

do kkk=1,FullSysSize-1,2

if (kkk .eq. 1) then

! set up the all-a and all-b states
SmallEig=0.d0
SmallEig(1,1)=-aAR
SmallEig(2,2)=-bAR

else
! do a more interesting state

zzz=(kkk-1.d0)/2.d0
SmallEig(1,1) = -(aAR+zzz*XK)
SmallEig(1,2) = zzz*XK
SmallEig(2,1) = (NNN+1.d0-zzz)*XK
SmallEig(2,2) = -(bAR+(NNN+1.d0-zzz)*XK)

endif

ILO=1
IHI=2

call DGEHRD( 2, ILO, IHI, SmallEig, 2, TAU, WORK, LWORK,
INFODGEHRD)
```

```
QSmallEig=SmallEig

call DORGHR( 2, ILO, IHI, QSmallEig, 2, TAU, WORK, LWORK,
INFODORGHR)

!write(*,*) "INFODORGHR: ", INFODORGHR

call DHSEQR( 'S', 'V', 2, ILO, IHI, SmallEig, 2, WR, WI,
QSmallEig, 2, WORK, LWORK, INFODHSEQR )

!write(*,*) "INFODHSEQR: ", INFODHSEQR

call DTREVC( 'R', 'B', SELECT, 2, SmallEig, 2, VL, 1,
QSmallEig, 2, 2, 2, WORKDTREVC, INFODTREVC )

!write(*,*) "INFODTREVC: ", INFODTREVC

ScaleQSmallEig(:,1)=QSmallEig(:,1)/dnrm2(2,QSmallEig(:,1),1
)
ScaleQSmallEig(:,2)=QSmallEig(:,2)/dnrm2(2,QSmallEig(:,2),1
)

loni=DPR / (2.d0**(NNN/2.d0))
hini=dsqrt(dabs((2.d0**-NNN)*(1.d0-DPR**2.d0)))

a0ni=(/hini, loni/)
b0ni=(/Bscale*loni, Bscale*hini/)

aC=a0ni
bC=b0ni

aDL=ScaleQSmallEig(2,1)
aD(1)=ScaleQSmallEig(1,1)
aD(2)=ScaleQSmallEig(2,2)
aDU=ScaleQSmallEig(1,2)

! Dx is destroyed
bDL=aDL
bD=aD
bDU=aDU

call DGTSV(2,1,aDL,aD,aDU,aC,2,INFO)
call DGTSV(2,1,bDL,bD,bDU,bC,2,INFO)

do jjj=1,NTpoints
aXpl(:,jjj)=aC*dexp(tspan(jjj)*WR)
bXpl(:,jjj)=bC*dexp(tspan(jjj)*WR)
```

```
end do

call DGEMM('N','N', 2, NTpoints, 2, ALPHA_SCALE,
ScaleQSmallEig, 2, aXpl, 2, MURFIN, EndaXpl, 2)
call DGEMM('N','N', 2, NTpoints, 2, ALPHA_SCALE,
ScaleQSmallEig, 2, bXpl, 2, MURFIN, EndbXpl, 2)

a2a=a2a+CoeffColumns(kkk)*EndaXpl(1,:)
!a2b=a2b+EndaXpl(2,:)
!b2a=b2a+EndbXpl(1,:)
b2b=b2b+CoeffColumns(kkk+1)*EndbXpl(2,:)

end do


fvec=(b2b/a2a)-commonY

end subroutine nzftfcn


!-----Factorial-----------------------------------------------
----------
!
!  Function to calculate factorials recursively
!
!-----------------------------------------------------------
-----------
RECURSIVE FUNCTION Factorial(n)  RESULT(Fact)

IMPLICIT NONE
DOUBLE PRECISION :: Fact
DOUBLE PRECISION, INTENT(IN) :: n

IF (n == 0.d0) THEN
   Fact = 1
ELSE
   Fact = n * Factorial(n-1.d0)
END IF

END FUNCTION Factorial
```

REFERENCES

1. Gilson, M. K. & Zhou, H.-X. (2007). Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 21-42.
2. Tsai, C.-J., del Sol, A. & Nussinov, R. (2008). Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play *J. Mol. Biol.* **378**, 1-11.
3. Meirovitch, H. (2007). Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr. Op. Struct. Biol.* **17**, 181-186.
4. Smock, R. G. & Gierasch, L. M. (2009). Sending Signals Dynamically. *Science* **324**, 198-203.
5. Sturtevant, J. M. (1977). Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci. USA* **74**, 2236-2240.
6. Karplus, M., Ichiye, T. & Pettitt, B. M. (1987). Configurational entropy of native proteins. *Biophys. J.* **52**, 1083-1085.
7. Cooper, A. & Dryden, D. T. F. (1984). Allostery without conformational change: a plausible model. *Eur. Biophys. J.* **11**, 103-109.
8. Brooks, B. R., Brooks III, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., V., O., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. & Karplus, M. (2009). CHARMM: The Biomolecular Simulation Program. *J. Comp. Chem.* **30**, 1545-1615.
9. Case, D. A., Darden, T. A., Cheatham, I., T.E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V. & Kollman, P. A. (2008). AMBER 10 10 edit. University of California, San Francisco, San Francisco.
10. Philippe, W. S., Scott, W. R. P., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T. & Kruger, P. (1999). The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **103**, 3596-3607.
11. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559.
12. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-9.
13. Akke, M., Brüschweiler, R. & Palmer, A. G. I. (1993). NMR Order Parameters and Free Energy: An Analytical Approach and Its Application to Cooperative Ca2+ Binding by Calbindin D9k. *J. Am. Chem. Soc.* **115**, 9832-9833.

14. Li, Z., Raychaudhuri, S. & Wand, A. J. (1996). Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Science* **5**, 2647-2650.
15. Yang, D. & Kay, L. E. (1996). Contributions to Conformational Entropy Arising from Bond Vector Fluctuations Measured from NMR-Derived Order Parameters: Application to Protein Folding. *J. Mol. Biol.* **263**, 369-382.
16. Lee, A. L., Kinnear, S. A. & Wand, A. J. (2000). Redistribution and loss of side chain entropy upon formation of a calmodulin−peptide complex. *Nature Structural Biology* **7**, 72-77.
17. Lee, A. L., Sharp, K. A., Kranz, J. K., Song, X.-J. & Wand, A. J. (2002). Temperature Dependence of the Internal Dynamics of a Calmodulin-Peptide Complex. *Biochemistry* **41**, 13814-13825.
18. Best, R. B., Clarke, J. & Karplus, M. (2005). What Contributions to Protein Side-chain Dynamics are Probed by NMR Experiments? A Molecular Dynamics Simulation Analysis *J. Mol. Biol.* **349**, 185-203.
19. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. (2007). Conformational entropy in molecular recognition by proteins. *Nature* **448**, 325-329.
20. Wang, T., Cai, S. & Zuiderweg, E. R. P. (2003). Temperature dependence of anisotropic protein backbone dynamics. *J. Am. Chem. Soc.* **125**, 8639-8643.
21. Jarymowycz, V. A. & Stone, M. J. (2006). Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev.* **106**, 1624-71.
22. Mandel, A. M., Akke, M. & Palmer, A. G. I. (1995). Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144-63.
23. Palmer, A. G. I., Rance, M. & Wright, P. E. (1991). Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371-4380.
24. Bodenhausen, G. & Ruben, D. J. (1980). Heteronuclear 2D correlation spectra with double in-phase transfer steps. *Chem. Phys. Letters* **69**, 185-188.
25. Fischer, M. W. F., Majumdar, A. & Zuiderweg, E. R. P. (1998). Protein NMR relaxation: theory, applications and outlook. *Prog. Nucl. Magn. Reson.* **33**, 207-272.
26. Palmer, A. G. I. (2001). NMR PROBES OF MOLECULAR DYNAMICS: Overview and Comparison with Other Techniques. *Annual Review of Biophysics and Biomolecular Structure* **30**, 129-155.
27. Pervushin, K. V., Riek, R., Wider, G. & Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366-71.
28. Zhu, G., Xia, Y., Nicholson, L. K. & Sze, K. H. (2000). Protein Dynamics Measurements by TROSY-Based NMR Experiments *J. Magn. Reson.* **143**, 423-426.
29. Zhu, G. & Yao, X. (2007). TROSY-based NMR experiments for NMR studies of large biomolecules *Prog. Nucl. Magn. Res. Sp.* **52**, 49-68.

30. Ferrage, F., Piserchio, A., Cowburn, D. & Ghose, R. (2008). On the measurement of 15N–{1H} nuclear Overhauser effects *J. Magn. Reson.* **192**, 302-313.
31. Igumenova, T. I., Frederick, K. K. & Wand, A. J. (2006). Characterization of the Fast Dynamics of Protein Amino Acid Side Chains Using NMR Relaxation in Solution. *Chem. Rev.* **106**, 1672-1699.
32. Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A. & Torchia, D. A. (1992). Dynamics of methyl groups in proteins as studied by proton-detected carbon-13 NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **31**, 5253-5263.
33. Muhandiram, D. R., Yamazaki, T., Sykes, B. D. & Kay, L. E. (1995). Measurement of 2H T1 and T1.rho. Relaxation Times in Uniformly 13C-Labeled and Fractionally 2H-Labeled Proteins in Solution. *J. Am. Chem. Soc.* **117**, 11536-11544.
34. Millet, O., Muhandiram, D. R., Skrynnikov, N. R. & Kay, L. E. (2002). Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in 13C-Labeled and Fractionally 2H-Enriched Proteins in Solution. *J. Am. Chem. Soc.* **124**, 6439-6448.
35. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. (2003). Cross-Correlated Relaxation Enhanced 1H−13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.* **125**, 10420-10428.
36. Ollerenshaw, J. E., Tugarinov, V. & Kay, L. E. (2005). Methyl TROSY: explanation and experimental verification. *Magn. Reson. Chem.* **41**, 843-852.
37. Peng, J. W. & Wagner, G. (1992). Mapping of spectral density functions using heteronuclear NMR relaxation measurements *J. Magn. Reson.* **98**, 308-322.
38. Skrynnikov, N. R., Millet, O. & Kay, L. E. (2002). Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions. *J. Am. Chem. Soc.* **124**, 6449-6460.
39. Ferrage, F., Pelupessy, P., Cowburn, D. & Bodenhausen, G. (2006). Protein Backbone Dynamics through 13C‘−13Cα Cross-Relaxation in NMR Spectroscopy. *J. Am. Chem. Soc.* **128**, 11072-11078.
40. Zeng, L., Fischer, M. W. F. & Zuiderweg, E. R. P. (1996). Study of Protein Dynamics in Solution by Measurement of 13Ca-13CO NOE and 13CO Longitudinal Relaxation. *J. Biomol. NMR* **7**, 157-162.
41. Fischer, M. W. F., Zeng, L., Majumdar, A. & Zuiderweg, E. R. P. (1998). Characterizing Semi-Local Motions in Proteins by NMR Relaxation Studies. *Proc. Natl. Acad. Sci. USA* **95**, 8016-8019.
42. Wang, T., Weaver, D. S., Cai, S. & Zuiderweg, E. R. P. (2006). Quantifying Lipari-Szabo modelfree parameters from 13CO NMR relaxation experiments. *J. Biomol. NMR* **36**, 79-102.
43. Neuhaus, D. & Williamson, M. P. (2000). *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. 2nd edit. Methods in Stereochemical Analysis (Marchand, A. P., Ed.), Wiley-VCH, New York.
44. Tropp, J. (1980). Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. *J. Chem. Phys.* **72**.

45. Olejniczak, E. T., Dobson, C. M., Karplus, M. & Levy, R. M. (1984). Motional averaging of proton nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme. *J. Am. Chem. Soc.* **106**, 1923-1930.
46. LeMaster, D. M., Kay, L. E., Brünger, A. T. & Prestegard, J. H. (1988). Protein dynamics and distance determination by NOE measurements. *FEBS Lett.* **236**, 71-76.
47. Palmer, A. G. I. & Case, D. A. (1992). Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide. *J. Am. Chem. Soc.* **114**, 9059-9067.
48. Post, C. B. (1992). Internal motional averaging and three-dimensional structure determination by nuclear magnetic resonance. *J. Mol. Biol.* **224**, 1087-1101.
49. Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M. & Ernst, R. R. (1992). Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. *J. Am. Chem. Soc.* **114**, 2289-2302.
50. Abseher, R., Lüdemann, S., Schreiber, H. & Steinhauser, O. (1995). NMR Cross-relaxation Investigated by Molecular Dynamics Simulation: A Case Study of Ubiquitin in Solution. *J. Mol. Biol.* **249**, 604-624.
51. Schneider, T. R., Brünger, A. T. & Nilges. (1999). Influence of internal dynamics on accuracy of protein NMR structures: derivation of realistic model distance data from a long molecular dynamics trajectory. *J. Mol. Biol.* **285**, 727-740.
52. Daura, X., Antes, I., van Gunsteren, W. F., Thiel, W. & Mark, A. E. (1999). The Effect of Motional Averaging on the Calculation of NMR-Derived Structural Properties. *Proteins: Structure, Function, and Genetics* **36**, 542-555.
53. Zagrovic, B. & van Gunsteren, W. F. (2006). Comparing Atomistic Simulation Data With the NMR Experiment: How Much Can NOEs Actually Tell Us? *Proteins: Structure, Function, and Bioinformations* **63**, 210-218.
54. Olejniczak, E. T., Poulsen, F. M. & Dobson, C. M. (1981). Proton Nuclear Overhauser Effects and Protein Dynamics. *J. Am. Chem. Soc.* **103**, 6574-6580.
55. Schleucher, J. & Wijmenga, S. S. (2002). How to Detect Internal Motion by Homonuclear NMR. *J. Am. Chem. Soc.* **124**, 5881-5889.
56. Pelupessy, P., Espallargas, G. M. & Bodenhausen, G. (2003). Symmetrical reconversion: measuring cross-correlation rates with enhanced accuracy. *J. Magn. Reson.* **161**, 258-264.
57. Lukas, T. J., Burgess, W. H., Prendergast, F. G., Lau, W. & Watterson, D. M. (1986). Calmodulin binding domains: characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. *Biochemistry* **25**, 1458-1464.
58. Crivici, A. & Ikura, M. (1995). Molecular and Structural Basis of Target Recognition by Calmodulin. *Annual Review of Biophysics and Biomolecular Structure* **24**, 85-116.
59. Yamniuk, A. P. & Vogel, H. J. (2007). Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Molecular Biotechnology* **27**, 33-57.
60. Swulius, M. T. & Waxham, M. N. (2008). Ca2+/Calmodulin-dependent Protein Kinases. *Cell. Mol. Life Sci.* **65**, 2637-2657.

61. Wintrode, P. L. & Privalov, P. L. (1997). Energetics of Target Peptide Recognition by Calmodulin: A Calorimetric Study. *J. Mol. Biol.* **266**, 1050-1062.
62. Brokx, R. D., Lopez, M. M., Vogel, H. J. & Makhatadze, G. I. (2001). Energetics of Target Peptide Binding by Camodulin Reveals Different Modes of Binding. *J. Biol. Chem.* **276**, 14083-14091.
63. Yang, C., Jas, G. S. & Kuczera, K. (2004). Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *BBA - Proteins & Proteomics* **1697**, 289-300.
64. Wilson, M. A. & Brünger, A. T. (2000). The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J. Mol. Biol.* **301**.
65. Valentine, K. G., Ng, H. L., Schneeweis, J. K., Kranz, J. K., Frederick, K. K., Alber, T. & Wand, A. J. (2007). Ultrahigh resolution crystal structure of calmodulin-smooth muscle light kinase peptide complex.
66. Woelk, T., Sigismund, S., Penengo, L. & Polo, S. (2007). The ubiquitination code: a signalling problem. *Cell Division* **2**.
67. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.* **120**, 6836-6837.
68. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128-132.
69. Richter, B., Gsponer, J., Varnai, P., Salvatella, X. & Vendruscolo, M. (2007). The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins *J. Biomol. NMR* **37**, 117-135.
70. Lange, O. F., Lakomek, N.-A., Fares, C., Schröder, G. F., Walter, K. F. A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C. & de Groot, B. L. (2008). Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **320**, 1471-1475.
71. Madl, T., Bermel, W. & Zangger, K. (2009). Use of Relaxation Enhancements in a Paramagnetic Environment for the Structure Determination of Proteins Using NMR Spectroscopy. *Angew. Chem. Int. Ed.* **48**, 8259-8262.
72. Mayer, M. P. & Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol. Life Sci.* **62**, 670-84.
73. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**, 623-628.
74. Wilbanks, S. M., DeLuca-Flaherty, C. & McKay, D. B. (1994). Structural basis of the 70-kilodalton heat shock cognate protein ATP hydrolytic activity. I. Kinetic analyses of active site mutants. *J. Biol. Chem.* **269**, 12893-12898.
75. Jiang, J., Maes, E. G., Taylor, A. B., Wang, L., Hinck, A. P., Lafer, E. M. & Sousa, R. (2007). Structural basis of J cochaperone binding and regulation of Hsp70. *Mol. Cell* **28**, 422-433.
76. Weaver, D. S. & Zuiderweg, E. R. P. (2008). Eta(z)/kappa: A transverse relaxation optimized spectroscopy NMR experiment measuring longitudinal relaxation interference. *J. Chem. Phys.* **128**.

77. Weaver, D. S. & Zuiderweg, E. R. P. (2009). Protein proton–proton dynamics from amide proton spin flip rates *J. Biomol. NMR* **45**, 99-119.
78. Werbelow, L. G. & Grant, D. (1977). Intramolecular dipolar relaxation in multispin systems. *Adv. Magn. Reson.* **9**, 189-299.
79. Pervushin, K. V., Wider, G. & Wüthrich, K. (1998). Single Transition-to-Single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J. Biomol. NMR* **12**, 345-348.
80. Rance, M., Loria, J. P. & Palmer, A. G. I. (1999). Sensitivity Improvement of Transverse Relaxation-Optimized Spectroscopy. *J. Magn. Reson.* **141**, 180-184.
81. Meissner, A., Schulte-Herbrüggen, T., Briand, J. & Sørensen, O. W. (1998). Double spin-state-selective coherence transfer. Application for twodimensional selection of multiplet components with long transverse relaxation times *Molecular Physics* **95**, 1137-1142.
82. Andersson, P., Annila, A. & Otting, G. (1998). An α/β-HSQC-α/β Experiment for Spin-State Selective Editing of IS Cross Peaks. *J. Magn. Reson.* **133**, 364-367.
83. Meissner, A., Duus, J. Ø. & Sørensen, O. W. (1997). Integration of spin-state-selective excitation into 2D NMR correlation experiments with heteronuclear ZQ/2Q π rotations for 1JXH. *J. Biomol. NMR* **10**, 89-94.
84. Schulte-Herbrüggen, T. & Sørensen, O. W. (2000). Clean TROSY: Compensation for Relaxation-Induced Artifacts. *J. Magn. Reson.* **144**, 123-128.
85. Roehrl, M. H. A., Heffron, G. J. & Wagner, G. (2005). Correspondence between spin-dynamic phases and pulse program phases of NMR spectrometers *J. Magn. Reson.* **174**, 325-330.
86. Mandal, P. K. & Majumdar, A. (2004). A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts in Magn. Reson. A* **20A**, 1-23.
87. Bouguet-Bonnet, S., Mutzenhardt, P. & Canet, D. (2004). Measurement of 15N csa/dipolar cross-correlation rates by means of Spin State Selective experiments. *J. Biomol. NMR* **30**, 133-142.
88. Pelupessy, P., Ferrage, F. & Bodenhausen, G. (2007). Accurate measurement of longitudinal cross-relaxation rates in nuclear magnetic resonance. *J. Chem. Phys* **126**.
89. Morris, G. A. & Freeman, R. J. (1979). Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **101**, 760-762.
90. Morris, G. A. & Freeman, R. J. (1980). Sensitivity Enhancement in 15N NMR: Polarization Transfer Using The INEPT Pulse Sequence. *J. Am. Chem. Soc.* **102**, 428-429.
91. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions *J. Biomol. NMR* **2**, 661-665.
92. Cavanagh, J. & Rance, M. (1993). Sensitivity-Enhanced NMR Techniques for the Study of Biomolecules. In *Annual reports on NMR spectroscopy* (Webb, G. A., ed.), Vol. 27, pp. 1-58. Academic Press, San Diego, CA.
93. Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665.

94. Muhandiram, D. R., Xu, G. Y. & Kay, L. E. (1993). An enhanced-sensitivity pure absorption gradient 4D 15N, 13C-edited NOESY experiment. *J. Biomol. NMR* **3**, 463-470.
95. Schleucher, J., Sattler, M. & Griesinger, C. (1993). Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNCO Experiment. *Angewandte Chemie* **32**, 1489-1491.
96. Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sørensen, O. W. & Griesinger, C. (1994). A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients *J. Biomol. NMR* **4**, 301-306.
97. Moré, J. J. (1978). The Levenberg-Marquart algorithm: Implementation and theory. In *Numerical Analysis* (Watson, G. A., ed.), Vol. 630, pp. 105-116. Springer, Berlin / Heidelberg.
98. Efron, B. & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability, CRC Press, Boca Raton, FL.
99. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2002). *Numerical Recipes in C++*. 2nd edit, Cambridge University Press, New York.
100. Argonne National Laboratory (1980). User Guide for MINPACK-1. Moré, J. J., Garbow, B. S. & Hillstrom, K. E.
101. Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. & Sorensen, D. (1999). *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
102. Blackford, L. S., Demmel, J., Dongarra, J., Duff, I., Hammarling, S., Henry, G., Heroux, M., Kaufman, L., Lumsdaine, A., Petitet, A., Pozo, R., Remington, K. & Whaley, R. C. (2002). An Updated Set of Basic Linear Algebra Subprograms (BLAS). *ACM Trans. Math. Soft.* **28**, 135-151.
103. Dongarra, J. (2002). Basic Linear Algebra Subprograms Technical Forum Standard. *International Journal of High Performance Applications and Supercomputing* **16**, 1-199.
104. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, B. I., Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**, W375-W383.
105. Yip, P. F. & Case, D. A. (1991). In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy* (Hoch, J. C., Poulsen, F. M. & Redfield, C., eds.), pp. 317-330. Plenum Press, New York.
106. Woessner, D. E. (1962). Spin Relaxation Processes in a Two-Proton System Undergoing Anisotropic Reorientation. *J. Chem. Phys* **36**.
107. Brüschweiler, R., Liao, X. & Wright, P. E. (1995). Long-Range Motional Restrictions in a Multidomain Zinc-Finger Protein from Anisotropic Tumbling. *Science* **268**, 886-889.
108. Lee, L. K., Rance, M., Chazin, W. & Palmer, A. G. I. (1997). Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13Cα nuclear spin relaxation *J. Biomol. NMR* **9**, 287-298.

109. Kroenke, C. D., Loria, J. P., Lee, L. K., Rance, M. & Palmer, A. G. I. (1998). Longitudinal and Transverse 1H−15N Dipolar/15N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules. *J. Am. Chem. Soc.* **120**, 7905-7915.
110. Wang, C., Rance, M. & Palmer, A. G. I. (2003). Mapping Chemical Exchange in Proteins with MW > 50 kD. *J. Am. Chem. Soc.* **125**, 8968-8969.
111. Cruickshank, D. W. J. (1999). Remarks about protein structure precision. *Acta Cryst.* **D55**, 583-601.
112. Chin, D. & Means, A. R. (2000). Calmodulin: a prototypical calcium sensor *Trends in Cell Biology* **10**, 322-328.
113. Olwin, B. B. & Storm, D. R. (1985). Calcium binding to complexes of calmodulin and calmodulin binding proteins. *Biochemistry* **24**, 8081-8086.
114. Spera, S., Ikura, M. & Bax, A. (1991). Measurement of the exchange rates of rapidly exchanging amide protons: Application to the study of calmodulin and its complex with a myosin light chain kinase fragment *J. Biomol. NMR* **1**, 155-165.
115. Tugarinov, V., Sprangers, R. & Kay, L. E. (2007). Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 1743-1750.
116. Peti, W., Meiler, J., Brüschweiler, R. & Griesinger, C. (2002). Model-Free Analysis of Protein Backbone Motion from Residual Dipolar Couplings. *J. Am. Chem. Soc.* **124**, 5822-5833.
117. Vögeli, B., Yao, L. & Bax, A. (2008). Protein backbone motions viewed by intraresidue and sequential HN–Hα residual dipolar couplings. *J. Biomol. NMR* **41**, 17-28.
118. Pang, Y., Buck, M. & Zuiderweg, E. R. P. (2002). Backbone Dynamics of the Ribonuclease Binase Active Site Area Using Multinuclear (15N and 13CO) NMR Relaxation and Computational Molecular Dynamics. *Biochemistry* **41**, 2655-2666.
119. Pellecchia, M., Pang, Y., Wang, L., Kurochkin, A. V., Kumar, A. & Zuiderweg, E. R. P. (1999). Quantitative Measurement of Cross-Correlations between 15N and 13CO Chemical Shift Anisotropy Relaxation Mechanisms by Multiple Quantum NMR. *J. Am. Chem. Soc.* **121**, 9165-9170.
120. Zhang, F. & Brüschweiler, R. (2002). Contact Model for the Prediction of NMR N−H Order Parameters in Globular Proteins. *J. Am. Chem. Soc.* **124**, 12654-12655.
121. Ming, D. & Brüschweiler, R. (2004). Prediction of methyl-side Chain Dynamics in Proteins *J. Biomol. NMR* **29**, 363-368.
122. Hicke, L. (2001). Protein regulation by monoubiquitin. *Nature Rev. Mol. Cell. Bio.* **2**, 195-201.
123. Di Fiore, P. P., Polo, S. & Hofmann, K. (2003). When ubiquitin meets ubiquitin receptors: a signalling connection. *Nature Rev. Mol. Cell. Bio.* **4**, 491-497.
124. Weber, P. L., Brown, S. C. & Mueller, L. (1987). Sequential 1H NMR assignments and secondary structure identification of human ubiquitin. *Biochemistry* **26**, 7282-7290.

125. Di Stefano, D. L. & Wand, A. J. (1987). Two-dimensional 1H NMR study of human ubiquitin: a main chain directed assignment and structure analysis. *Biochemistry* **26**, 7272-7281.
126. Schneider, D. M., Dellwo, M. J. & Wand, A. J. (1992). Fast internal main-chain dynamics of human ubiquitin. *Biochemistry* **31**, 3645-3652.
127. Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). Rotational Diffusion Anisotropy of Human Ubiquitin from 15N NMR Relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566.
128. Fushman, D., Tjandra, N. & Cowburn, D. (1998). Direct Measurement of 15N Chemical Shift Anisotropy in Solution. *J. Am. Chem. Soc.* **120**, 10947-10952.
129. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **194**, 531-544.
130. Paci, E. & Karplus, M. (1999). Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288**, 441-459.
131. Best, R. B. & Vendruscolo, M. (2004). Determination of Protein Structures Consistent with NMR Order Parameters. *J. Am. Chem. Soc.* **126**, 8090-8091.
132. Schlossman, D. M., Schmid, S. L., Braell, W. A. & Rothman, J. E. (1984). An enzyme that removes clathrin coats: purification of an uncoating ATPase. *J. Cell Biol.* **99**, 723-733.
133. Pelham, H. R. B. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins *Cell* **46**, 959-961.
134. Chappell, T. G., Konforti, B. B., Schmid, S. L. & Rothman, J. E. (1987). The ATPase core of a clathrin uncoating protein. *J. Biol. Chem.* **262**, 746-751.
135. Deluca-Flaherty, C., Flaherty, K. M., McIntosh, L. J., Bahrami, B. & McKay, D. B. (1988). Crystals of an ATPase fragment of bovine clathrin uncoating ATPase. *J. Mol. Biol.* **200**, 749-750.
136. Holmes, K. C., Sander, C. & Valencia, A. (1993). A new ATP-binding fold in actin, hexokinase and Hsc70. *Trends Cell Biol.* **3**, 53-59.
137. O'Brien, M. C. & McKay, D. B. (1995). How Potassium Affects the Activity of the Molecular Chaperone Hsc70 I. POTASSIUM IS REQUIRED FOR OPTIMAL ATPase ACTIVITY. *J. Biol. Chem.* **270**, 2247-2250.
138. O'Brien, M. C., Flaherty, K. M. & McKay, D. B. (1996). Lysine 71 of the Chaperone Protein Hsc70 Is Essential for ATP Hydrolysis. *J. Biol. Chem.* **271**, 15874-15878.
139. Johnson, E. R. & McKay, D. B. (1999). Mapping the Role of Active Site Residues for Transducing an ATP-Induced Conformational Change in the Bovine 70-kDa Heat Shock Cognate Protein. *Biochemistry* **38**, 10823-10830.
140. Barthel, T. K., Zhang, J. & Walker, G. C. (2001). ATPase-Defective Derivatives of Escherichia coli DnaK That Behave Differently with Respect to ATP-Induced Conformational Change and Peptide Release *J. Bacteriology* **183**, 5482-5490.
141. Radkiewicz, J. L. & Brooks, C. L. I. (2000). Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase. *J. Am. Chem. Soc.* **122**, 225-231.
142. Antoniou, D., Caratzoulas, S., Kalyanaraman, C., Mincer, J. S. & Schwartz, S. D. (2002). Barrier passage and protein dynamics in enzymatically catalyzed reactions. *Eur. J. Biochem.* **269**, 3103-3112.

143. Kern, D. & Zuiderweg, E. R. P. (2003). The role of dynamics in allosteric regulation. *Curr. Op. Struct. Biol.* **13**, 748-757.
144. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2006). The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **313**, 1636-1642.
145. Goodey, N. M. & Benkovic, S. (2008). Allosteric regulation and catalysis emerge via a common route. *Nature Chem. Biol.* **4**, 474-481.
146. Wand, A. J. (2001). Dynamic activation of protein function: A view emerging from NMR spectroscopy. *Nature Struct. Biol.* **8**, 926-931.
147. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. & Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-121.
148. Henzer-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-916.
149. Henzler-Wildman, K. A. & Kern, D. (2007). Dynamic personalities of proteins. *Nature* **450** , 964-972.
150. Watt, E. D., Shimada, H., Kovrigin, E. L. & Loria, J. P. (2007). The mechanism of rate-limiting motions in enzyme function. *Proc. Natl. Acad. Sci. USA* **104**, 11981-11986.
151. Lee, G. M. & Craik, C. S. (2009). Trapping Moving Targets With Small Molecules. *Science* **324**, 213-215.
152. Gong, Q. & Ishima, R. (2007). 15N–{1H} NOE experiment at high magnetic field strengths *J. Biomol. NMR* **37**, 147-157.