



---

Publicly Accessible Penn Dissertations

---

1-1-2012

# Protein Hydrogen Exchange, Dynamics, and Function

John Skinner

*University of Pennsylvania*, [skinnerj@mail.med.upenn.edu](mailto:skinnerj@mail.med.upenn.edu)

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Biochemistry Commons](#), and the [Biophysics Commons](#)

---

## Recommended Citation

Skinner, John, "Protein Hydrogen Exchange, Dynamics, and Function" (2012). *Publicly Accessible Penn Dissertations*. 582.  
<http://repository.upenn.edu/edissertations/582>

This paper is posted at Scholarly Commons. <http://repository.upenn.edu/edissertations/582>  
For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# Protein Hydrogen Exchange, Dynamics, and Function

## **Abstract**

Models derived from X-ray crystallography can give the impression that proteins are rigid structures with little mobility. NMR ensembles may suggest a more dynamic picture, but even these represent a rather narrow range of possibilities close to the lowest energy state. In reality proteins participate in a wide range of dynamics from the subtle and rapid sidechain dynamics that occur in nanoseconds in the PDZ signaling domain to the large and slow rearrangement of secondary structure that takes days in the mitotic checkpoint protein Mad2. Between these extremes are motions on time scales typically associated with protein function, such as those in SNase monitored by hydrogen exchange. The dynamic character of several protein systems, including PDZ domain, Calmodulin, SNase, and Mad2, were explored using a variety of biophysical techniques. This broad investigation demonstrates the dynamic variability between and within proteins. The study of PDZ and Calmodulin illustrates how a computational technique can recapitulate experimental results and provide additional insight into signal transduction. The case of SNase shows that HX NMR data can be exploited to reveal protein dynamics with unprecedented detail. The Mad2 system highlighted some of the pitfalls associated with this technique and some alternative strategies for investigating protein dynamics.

## **Degree Type**

Dissertation

## **Degree Name**

Doctor of Philosophy (PhD)

## **Graduate Group**

Biochemistry & Molecular Biophysics

## **First Advisor**

S Walter Englander

---

**Second Advisor**

Ben E. Black

**Keywords**

CLEANEX-PM, HX, Hydrogen exchange, Mad2, Protein folding, Pump probe molecular dynamics

**Subject Categories**

Biochemistry | Biophysics

# PROTEIN HYDROGEN EXCHANGE, DYNAMICS, AND FUNCTION

**John J. Skinner**

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

Co-Supervisor of Dissertation

---

S. Walter Englander  
Professor  
Biochemistry and Biophysics

---

Ben E. Black,  
Assistant Professor  
Biochemistry and Biophysics

Graduate Group Chairperson

---

Kate Ferguson  
Associate Professor  
Physiology

## Dissertation Committee Members

Kim Sharp, Associate Professor,  
Biochemistry and Biophysics

P. Leslie Dutton, Eldrige Reeves  
Johnson Professor, Biochemistry and  
Biophysics

A. Joshua Wand, Benjamin Rush  
Professor, Biochemistry and Biophysics

Greg Van Duyne, Jacob Gershon-Cohen  
Professor, Biochemistry and Biophysics

Ravi Radhakrishnan, Associate Professor,  
Bioengineering

Doug Barrick, Professor, Biophysics,  
Johns Hopkins University

Dedicated to my grandfather, James  
“Pappy” Skinner, who taught me the  
value of honesty, which is essential in  
the pursuit of Truth.

## ACKNOWLEDGMENTS

I would like everyone who contributed this work. Foremost, I would like to thank my advisers Walter Englander and Ben Black who provided guidance throughout most of my graduate career and Kim Sharp who provided guidance during the early phase of my graduate studies. I would also like to thank Stacey Wood, Leland Mayne, Ninad Prabhu, and Yibing Wu for the training and assistance they provided. In addition to sharing her HXMS data with me, Sandya Ajith proved to be a valuable friend at all times. Samuel Getchell in my tribulations of protein purification and partnered with Brian Wexler to collect some of the Mad2 kinetic data that has been included in this thesis. I would also like to thank J. Nathan Scott for his patience in answering my endless Matlab questions. Many thanks to those who provided materials, data, and assistance including Sergei Vignodagrav who provided Rhodamine B fluorophore, Lee Solomon who helped me incorporate the fluorophore into a synthetic peptide, and Bogumil Zelent who guided me in my fluorescence measurements. Thanks to Michele Vendruscolo and Liu Tong for providing some of the HX predictions used in Chapter 4 and to Tobin Sosnick who not only allowed me to use the NMR facility at University of Chicago, but also allowed me to stay in his home while doing so. I would also like to thank all the members of the Englander and Black labs for helpful discussions throughout my time in those labs. Finally, I want to thank my dog Montana who made lab all the more enjoyable with his presence.

## **ABSTRACT**

# **PROTEIN HYDROGEN EXCHANGE, DYNAMICS, AND FUNCTION**

John J. Skinner

S. Walter Englander

Ben E. Black

Models derived from X-ray crystallography can give the impression that proteins are rigid structures with little mobility. NMR ensembles may suggest a more dynamic picture, but even these represent a rather narrow range of possibilities close to the lowest energy state. In reality proteins participate in a wide range of dynamics from the subtle and rapid sidechain dynamics that occur in nanoseconds in the PDZ signaling domain to the large and slow rearrangement of secondary structure that takes days in the mitotic checkpoint protein Mad2. Between these extremes are motions on time scales typically associated with protein function, such as those in SNase monitored by hydrogen exchange. The dynamic character of several protein systems, including PDZ domain, Calmodulin, SNase, and Mad2, were explored using a variety of biophysical techniques. This broad investigation demonstrates the dynamic variability between and within proteins. The study of PDZ and Calmodulin illustrates how a computational technique can recapitulate experimental results and provide additional insight into signal transduction. The case of SNase shows that HX NMR data can be exploited to reveal protein dynamics with unprecedented detail. The Mad2 system highlighted some of the pitfalls associated with this technique and some alternative strategies for investigating protein dynamics.

# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
1. Protein Dynamics.....	1
1.1 Allostery.....	1
1.2 Relating Protein Dynamics to Hydrogen Exchange Data.....	2
1.3 Conformational Switching.....	3
2. Hydrogen Exchange Theory.....	4
2.1 Hydrogen Exchange Rates.....	4
2.2 Mechanisms of Amide Exposure.....	5
3. Overview of this thesis.....	6
4. Bibliography.....	7
CHAPTER 2: Using pump-probe molecular dynamics to find allosteric pathways.....	12
1. Introduction.....	12
2. Methods.....	18
2.1 PPMD.....	18
2.2 Quantifying the Coupling, or Effectiveness of Transmission of the Pumped Motion.....	20
2.3 Comparison of Coupling Profiles.....	21
2.3.1 Outlier analysis.....	21
2.3.2 Percentile analysis.....	22
2.4 Frequency and Correlation Analysis.....	23
2.5 MD Protocol.....	24

2.6 Additional PPMD Enhancements.....	25
2.7 Proteins Studied.....	26
2.7.1 CaM (PDB entry 1CDL).....	27
2.7.2 PDZ domain protein (PDB entry 1BE9).....	28
3. Results.....	29
4. Discussion.....	46
5. Future Directions.....	55
6. Bibliography.....	57
CHAPTER 3: Enhancing the Accuracy of Hydrogen Exchange Rates Determined by $^1\text{H}$ - $^1\text{H}$ Measurements.....	68
1. Introduction.....	68
2. Theory - Magnetization Transfer.....	69
3. Methods.....	70
3.1 Data Collection.....	70
3.2 Accurate Fitting Algorithm.....	71
3.3 Removal of Background Noise.....	71
4. Results.....	72
4.1 Stepwise Fitting.....	72
4.2 Rate Validation.....	78
4.2.1 pH Dependence.....	78
4.2.2 Comparison to $^2\text{H}$ - $^1\text{H}$ Rates.....	79
4.2.3 Comparison to Theoretical Rates.....	80
5. Conclusions.....	80
6. Bibliography.....	81

CHAPTER 4: Comparing Hydrogen Exchange Models to Measured Experiments.....	84
1. Introduction.....	84
2. Materials and Methods.....	86
2.1 HX Rates.....	86
2.2 Structure-Based Calculations.....	88
3. Results and Discussion.....	89
3.1 Solvent Accessibility.....	89
3.2 Solvent Penetration.....	90
3.3 Electrostatic Effects on HX Rates.....	92
3.3.1 Validating Electrostatic Calculations.....	92
3.3.2 Applying Electrostatic Calculations to SNase.....	95
3.4 Packing Density.....	98
3.5 COREX.....	101
4. Conclusions.....	104
5. Bibliography.....	105
 CHAPTER 5: Interpreting Protein Dynamics from Hydrogen Exchange Data.....	 110
1. Introduction.....	110
2. Materials and Methods.....	111
3. Results and Discussion.....	112
3.1 Hydrogen Bond Acceptor Types.....	112
3.2 Structural Context.....	114
3.2.1 Global unfolding.....	114
3.2.2 Heterogeneous Unfolding in a Beta Sheet.....	117

3.2.3 Varied Exchange in Three $\alpha$ Helices.....	120
3.2.4 Loops and Water.....	123
3.2.5 Resolving Structural Ambiguities.....	127
3.2.6 Local Fluctuations.....	130
3.3 Modeling HX behavior.....	134
3.3.1 Accuracy of $k_{ch}$ .....	134
3.3.2 The HX Competent State.....	136
3.3.3 Multiple Exchange Pathways.....	136
4. Conclusion.....	137
5. Bibliography.....	139
CHAPTER 6: The Mad2 Conformational Transition.....	144
1. Introduction.....	144
1.1 Mad2 Function.....	144
1.2 Mad2 Structural Rearrangement.....	148
1.3 Conformational Switching Models.....	150
1.4 An Unfolding Model for Mad2 Conformational Change.....	153
2. Materials and Methods.....	157
2.1 Mutations.....	157
2.2 Protein Purification.....	158
2.2.1 Cell Growth.....	159
2.2.2 Purification.....	160
2.2.3 Chromatography.....	161
2.3 Peptide Synthesis.....	162
2.4 Spectroscopy.....	163

2.4.1	Circular Dichroism.....	163
2.4.2	Fluorescence Correlation Spectroscopy.....	163
2.4.3	NMR.....	165
2.5	Hydrogen Exchange Mass Spectroscopy.....	166
2.6	Electrostatic Calculations.....	166
3.	Results and Discussion.....	166
3.1	Separation of Mad2 conformers.....	166
3.2	Mad2 Interconversion.....	169
3.2.1	Macroscopic Interconversion.....	169
3.2.2	Residue Specific Interconversion.....	173
3.2.3	Electrostatic Considerations.....	178
3.2.4	Summary of the Mad2 Intermediate.....	179
3.3	Hydrogen Exchange.....	181
3.3.1	NMR.....	181
3.3.2	Hydrogen Exchange Mass Spectroscopy.....	183
3.4	NMR Peak Assignments.....	185
3.5	Peptide Binding.....	187
4.	Conclusions.....	192
5.	Future Directions.....	194
6.	Bibliography.....	196

## LIST OF TABLES

Table 2.1 Phase Delay of a 10 ps Pump Applied to H76 of PDZ.....	43
Table 2.2 Summary of changes in coupling due to mutations in PDZ domain protein....	46
Table 3.1 pH-independent log $k_0$ rates for SNase.....	77
Table 5.1 SNase protection factors.....	133
Table 6.1 Mad2 mutants used in this study.....	157
Table 6.2 Temperature dependent interconversion rates.....	173
Table 6.3 Rates of decay for O-Mad2 and rise of C-Mad2 peaks.....	175
Table 6.4 Peak decay rates conversion and hydrogen exchange.....	182

## LIST OF FIGURES

Figure 2.1 Fluctuation power spectra of some atomic motions in CaM.....	30
Figure 2.2 Coupling profiles for CaM with the C helix pumped.....	31
Figure 2.3 Coded structural representation of CaM.....	32
Figure 2.4 Coded structural representation of PDZ domain protein.....	32
Figure 2.5 PDZ domain coupling profiles.....	34
Figure 2.6 Plot of coupling strength vs. $C\alpha$ - $C\alpha$ distance for PDZ domain pumped at H76.....	35
Figure 2.7 Percentile analysis of similarity of coupling profiles for CaM.....	37
Figure 2.8 Percentile analysis of similarity of coupling profiles for PDZ.....	38
Figure 2.9 Fluctuation cross-correlation for PDZ domain protein from a control simulation.....	41
Figure 2.10 Time cross-correlation functions for $C\alpha$ atoms in PDZ domain.....	42
Figure 2.11 Time autocorrelation functions for atomic motions in CaM with implicit or explicit solvent.....	45
Figure 3.1 CLEANEX-PM pulse sequence.....	69
Figure 3.2 Uncorrected peak volumes as a function of mixing time for SNase Gly86....	73
Figure 3.3 Peaks volumes after background signal has been subtracted.....	73
Figure 3.4 Linear fits to time courses.....	74
Figure 3.5 Best fit lines using Eq. 3.1 with the approximation $R_{1A} = 0$ .....	75
Figure 3.6 Best fit lines using Eq. 3.1 to fit both $k_{ex}$ and $R_{1A}$ .....	75
Figure 3.7 Fit for all pH's simultaneously.....	76
Figure 3.8 Final pH dependent fitting step.....	77
Figure 3.9 $k_{ex}$ determined from linear slope vs final $k_{ex}$ .....	78
Figure 3.10 $\log k_{ex}$ as a function of pH.....	79

Figure 4.1 Log $k_{\text{ex}}$ vs pH for SNase.....	87
Figure 4.2 Protection factors as a function of distance to the protein surface.....	91
Figure 4.3 $\Delta G$ of deprotonation calculated by Qnifft compared to previously published $\Delta pK_a$ calculations for rubredoxin.....	93
Figure 4.4 $\Delta G$ of deprotonation calculated by Qnifft compared to previously published $\Delta pK_a$ calculations for CI2 and FKBP12.....	93
Figure 4.5 $\Delta G$ of deprotonation calculated by Qnifft compared to previously published $\Delta pK_a$ calculations for ubiquitin.....	94
Figure 4.6 $\Delta pK_a$ from electrostatic calculations vs measured HX rates for SNase.....	95
Figure 4.7 Measured log $P_f$ vs log $P_f$ predicted from local packing density.....	99
Figure 4.8 Log $P_f$ predicted by H-COREX vs measured log $P_f$ .....	102
Figure 5.1 Hydrogen bond inventory for SNase.....	113
Figure 5.2 HX protection pattern indicating global unfolding on one face of the SNase $\beta$ -barrel.....	115
Figure 5.3 Heterogeneous exchange in the SNase $\beta$ -barrel.....	118
Figure 5.4 Sidechains between solvent and Val23 and Leu25.....	120
Figure 5.5 HX behavior in the SNase $\alpha$ -helices.....	121
Figure 5.6 Cooperative motion in the $\alpha 1/\beta 4$ loop.....	124
Figure 5.7 An internal water molecule and several loops.....	125
Figure 5.8 Loops with internal waters.....	127
Figure 5.9 Electron density map surrounding Thr82.....	128
Figure 5.10 The Asn138 protection factor is consistent with the SNase NMR structure.....	129
Figure 5.11 Exchange mechanisms displayed on the structure of SNase.....	131
Figure 6.1 The mitotic checkpoint ensures equal partitioning of chromosomes in anaphase.....	146
Figure 6.2 Mad2 containing complexes.....	151

Figure 6.3 Mad2 unfolding and refolding considerations.....	155
Figure 6.4 Mad2 purification gel.....	161
Figure 6.5 Anion exchange and size exclusion chromatograms for Mad2 <sup>RQ</sup> .....	168
Figure 6.6 CD spectra for Mad2 <sup>RF</sup> .....	169
Figure 6.7 Mad2 interconversion.....	170
Figure 6.8 Structure highlighting interactions involving the residues that have been mutated in Mad2 <sup>RQ</sup> and Mad2 <sup>RF</sup> .....	172
Figure 6.9 <sup>1</sup> H- <sup>15</sup> N HSQC spectra for Mad2.....	174
Figure 6.10 <sup>1</sup> H- <sup>15</sup> N peaks that shift from open to closed conformations mapped onto O-Mad2 and C-Mad2 structures.....	177
Figure 6.11 Electrostatic analysis of differences in binding affinity for O-Mad2 and C-Mad2.....	179
Figure 6.12 HX mass spectroscopy results for Mad2 <sup>RF</sup> .....	184
Figure 6.13 2D projections of HNCA and HN(CO)CA spectra.....	186
Figure 6.14 FCS results for RhB-MBP1 binding to Mad2.....	188
Figure 6.15 RhB-MBP1 binding monitored by fluorescence intensity monitored as a function of time.....	189

## ABBREVIATIONS

- APC – Anaphase promoting complex/cyclosome
- BSA – Bovine serum albumin
- CaM - Calmodulin
- CD – Circular dichroism
- Cdc20 – Cell division cycle 20. A protein regulated by Mad2
- C-Mad2 – Closed conformation of Mad2
- FCS – Fluorescence correlation spectroscopy
- FPLC – Fast pressure liquid chromatography
- GdmCl – Guanidinium chloride
- HSQC – Heteronuclear single quantum coherence
- HX – Hydrogen exchange
- IPTG – Isopropyl  $\beta$ -D-1-thiogalactopyranoside
- LB – Lysogeny broth
- Mad1 – Mitotic arrest deficient protein 1
- Mad2 – Mitotic arrest deficient protein 2
- Mad2 <sup>$\Delta$ C</sup> – Mad2 with the last 10 residues truncated
- Mad2 <sup>$\Delta$ N</sup> – Mad2 with the first 10 residues truncated
- Mad2<sup>RF</sup> – Mad2 R133A/F141A mutant
- Mad2<sup>RQ</sup> – Mad2 R133E/Q134A mutant
- MBP1 – Mad2 binding peptide 1
- MCC – Mitotic checkpoint complex
- MD – Molecular dynamics
- NATA – N-acetyl-L-tryptophanamide

NMR – Nuclear Magnetic Resonance

O-Mac2 – Open conformation of Mad2

PDZ – PSD-95, Discs Large, Zo-1. A common protein domain.

P<sub>f</sub> – Protection factor

PMSF - Phenylmethanesulfonylfluoride

PPMD – Pump-probe molecular dynamics

RhB-MBP1 – MBP1 covalently linked to a Rhodamine B derivative

SEC – Size exclusion chromatography

SNase – Staphylococcal nuclease

WT Mad2 – Post PreScission™ cleavage form of Mad2. Pro has been inserted at the +2 position

# CHAPTER 1

## INTRODUCTION

### 1. Protein Dynamics

Models derived from X-ray crystallography can give the impression that proteins are rigid structures with little mobility. NMR ensembles may suggest a more dynamic picture, but even these represent a rather narrow range of possibilities close to the lowest energy state. In reality proteins participate in a wide range of dynamics from the subtle and rapid sidechain dynamics that occur in nanoseconds in the PDZ signaling domain (1) to the large and slow rearrangement of secondary structure that takes days in the mitotic checkpoint protein Mad2 (2). Between these extremes are motions on time scales typically associated with protein function, such as those in SNase monitored by hydrogen exchange.

#### 1.1 Allostery

Allosteric effects require interactions between spatially separated sites without obvious structural reconfiguration. The mechanism underlying this process is not completely understood. Allostery can be thought of as occurring by modulation of dynamic pathways (3) or by perturbing the overall protein ensemble (4). These models are not necessarily mutually exclusive as an ensemble view of allostery still requires

distal domains to somehow communicate with one another and a pathway view is compatible with multiple routes.

One popular method for investigating protein dynamics is to apply molecular dynamics (MD) simulations. Unfortunately, it is difficult to parse the dynamics relating to allosteric regulation from the other dynamics occurring throughout a protein. Chapter 2 describes a novel technique called pumped-probe MD (ppmd; (1) that allows for the detection of allosteric pathways by applying a low frequency pumping force to an MD simulation and monitoring how the force transfers through the protein.

## **1.2 Relating Protein Dynamics to Hydrogen Exchange Data**

Hydrogen exchange (HX) has proved to be a valuable technique for studying protein dynamics. What makes this method so powerful is that it is only sensitive to the populations in which a backbone amide is exposed to solvent. For most residues this corresponds to higher energy states that become populated anywhere from  $10^{-1}$  to  $10^{-9}$  of the time. Because HX rates are known for unfolded peptides (5, 6), HX rates measured for folded proteins can be translated into stability constants. Combining this technique with 2D NMR (7) allows for the determination of residue resolved stabilities.

HX rates have often been associated with protein folding (8). However, faster rates often result from local motions that occur on timescales relevant to many protein functions. Obtaining HX rates for marginally protected and unprotected residues has become easier with advances in NMR magnetization transfer experiments (9–11). Yet the determination of rates from these experiments has been hindered by approximations

made to fitting algorithms. Chapter 3 describes a novel algorithm for accurately determining HX rates from magnetization transfer experiments.

Several groups have applied models to relate HX data to protein structure (12–14). Chapter 4 evaluates the effectiveness of these models at predicting HX rates for Staphylococcal nuclease (SNase). The failure of each of these methods to accurately predict HX rates inspired the detailed examination of the SNase structure in relation to measured HX rates, discussed in Chapter 5. The structural analysis leads to several general conclusions including that residues with multiple layers of secondary structure between themselves and the edges of those structures tend to exchange as cooperative units regardless of the proximity to solvent and that an amide can hydrogen bond to a water molecule without becoming HX competent. In several cases involving local structural fluctuations detailed motions could be inferred from the context of nearby residues based on HX rates and local structure.

### **1.3 Conformational Switching**

Mad2 is one of a handful of proteins, known as metamorphic proteins (15), that undergo a drastic structural rearrangement as part of their function. Other such proteins include RNA polymerase (16, 17), viral glycoprotein (18), chloride ion channel (19), lysozyme (20), and chemokine lymphotactin (21, 22). Of this group, Mad2 is the only protein known to exist in both of its conformations in the absence of ligand or environmental perturbation. In the case of Mad2 this conformation change includes the rearrangement of two  $\beta$  strands and the transition of another  $\beta$  strand into an  $\alpha$  helix (2).

Conformational switching is required for Mad2 to bind either its upstream effector or the protein Mad2 regulates (23). While this switching is necessary for Mad2's function as a mitotic regulatory protein, the structural details of this transition have not been described (24). Chapter 6 of this thesis describes an investigation of the intermediate I-Mad2 conformation by various biophysical techniques.

## 2. Hydrogen Exchange Theory

### 2.1 Hydrogen Exchange Rates

Backbone amides have a  $pK_a > 18$ , which means they are in constant exchange with solvent even though the population is never significantly deprotonated. Above  $pH \sim 3$  HX is catalyzed by  $OH^-$ , thus the rate of HX increases by 10-fold for each pH unit.  $N^-$  is a stronger base than  $OH^-$  by  $>100$ -fold so amide NH to  $OH^-$  ion collision occurs  $>100$  times before the proton is actually carried away. For a polypeptide chain in a random coil HX rates at  $pH 7$  at  $0^\circ C$  are about  $1 s^{-1}$ . These rates are dependent on sequence, temperature, pH and isotope, which have been accurately calibrated (5, 6).

When an amide is hydrogen bonded it is isolated from solvent and therefore completely protected from attack by  $OH^-$  catalyst. HX occurs only when the amide samples a state in which it is exposed to solvent. The resulting HX rate can be described as follows:

$$k_{obs} = k_{op}k_{ch}/(k_{op} + k_{cl} + k_{ch}) \quad \text{Eq. 1.1}$$

where  $k_{ch}$  is the known HX rate for the amide if it were in a random coil,  $k_{op}$  is the rate for opening the amide to solvent, and  $k_{cl}$  is the reclosing rate. In most cases  $k_{cl} \gg k_{ch}$  and  $k_{cl} \gg k_{op}$ , which reduces Eq. 1 to

$$k_{obs} = k_{op}k_{ch}/k_{cl} \quad \text{Eq. 1.2}$$

This is known as the EX2 case. Since  $k_{ch}$  is known, one can determine the ratio  $k_{op}/k_{cl}$  which is trivially converted into stability against the opening reaction. It should be noted that this analysis assumes that the exposed state exchanges at the same rate as a random coil, which may not be true in all cases (discussed in Chapter 5).

When pH is high and  $k_{cl}$  is slow ( $k_{cl} \ll k_{ch}$ ), we enter the EX1 regime. Here Eq. 1.1 reduces to

$$k_{obs} = k_{op}k_{ch}/k_{ch} = k_{op} \quad \text{Eq. 1.3}$$

EX1 behavior has only been observed in the most stable regions of very stable proteins.

## 2.2 Mechanisms of Amide Exposure

The stability and (un)folding rates of proteins are typically determined by methods that analyze the protein as a whole or possibly one specific region of the protein. NMR HX allows the determination of these properties on a residue-by-residue basis. This has led to the observation that hydrogen bonded amides exchange by three different exposure mechanisms: global unfolding, sub-global unfolding, and local fluctuations (8).

For amides that exchange by global unfolding, HX exhibits the same stability, denaturant dependence, and unfolding rate ( $k_{op}$ ) as the protein as a whole as measured by other methods (25, 26). Collectively, these residues appear to represent the first fragment or “foldon” of the protein to fold and the last to unfold (8). This first foldon may be stable independently of the rest of the protein and can provide the nucleus or template upon which the rest of the structure can form (27).

Amides that exchange by a sub-global unfolding mechanism have a lower stability and shallower denaturant dependence than their global counterparts (27). The shallower denaturant dependence is due to the partially unfolded conformers exposing less surface area than the globally unfolded state. When these amides are clustered into foldons by their equilibrium and kinetic properties and mapped onto the protein, they are typically found to be clustered spatially as well. These units are thought to build upon the global scaffold and one another in a sequential manner (28).

Exchange by local fluctuations is the least understood of the three mechanisms. These residues exhibit no denaturant dependence, which presumably means little or no additional surface area is exposed in the exchange competent state. From the perspective of one studying protein folding, local fluctuations are an annoyance which one tries to circumvent by adding denaturant until the stability becomes such that the amide exchanges by way of a more interesting global or sub-global unfolding. However, these fluctuations may play an important role in other protein functions.

### **3. Overview of This Thesis**

The dynamic character of several protein systems, including PDZ domain, Calmodulin, SNase and Mad2, were explored using a variety of biophysical techniques. This broad investigation demonstrates the dynamic variability between and within proteins. The study of PDZ and Calmodulin illustrates how a computational technique can recapitulate experimental results and provide additional insight into signal transduction. The case of SNase shows that HX NMR data can be exploited to reveal protein dynamics with unprecedented detail. The Mad2 system highlighted some of the pitfalls associated with this technique and some alternative strategies for investigating protein dynamics.

#### **4. Bibliography**

1. Sharp, K., and J.J. Skinner. 2006. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins*. 65: 347-361.
2. Luo, X., Z. Tang, G. Xia, K. Wassmann, T. Matsumoto, J. Rizo, and H. Yu. 2004. The Mad2 spindle checkpoint protein has two distinct natively folded states. *Nat Struct Mol Biol*. 11: 338-345.
3. del Sol, A., C.-J. Tsai, B. Ma, and R. Nussinov. 2009. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure*. 17: 1042-1050.
4. Wrabl, J.O., J. Gu, T. Liu, T.P. Schrank, S.T. Whitten, and V.J. Hilser. 2011. The role of protein conformational fluctuations in allostery, function, and evolution. *Biophys. Chem*. 159: 129-141.

5. Bai, Y., J.S. Milne, L. Mayne, and S.W. Englander. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins*. 17: 75-86.
6. Connelly, G.P., Y. Bai, M.F. Jeng, and S.W. Englander. 1993. Isotope effects in peptide group hydrogen exchange. *Proteins*. 17: 87-92.
7. Wagner, G., and K. Wüthrich. 1982. Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution: Studies with two-dimensional nuclear magnetic resonance. *Journal of Molecular Biology*. 160: 343-361.
8. Englander, S.W., L. Mayne, and M.M.G. Krishna. 2007. Protein Folding and Misfolding: Mechanism and Principles. *Quarterly Reviews of Biophysics*. 40: 287-326.
9. Gemmecker, G., W. Jahnke, and H. Kessler. 1993. Measurement of fast proton exchange rates in isotopically labeled compounds. *J. Am. Chem. Soc.* 115: 11620-11621.
10. Hwang, T.-L., S. Mori, A.J. Shaka, and P.C.M. van Zijl. 1997. Application of Phase-Modulated CLEAN Chemical EXchange Spectroscopy (CLEANEX-PM) to Detect Water-Protein Proton Exchange and Intermolecular NOEs. *J. Am. Chem. Soc.* 119: 6203-6204.
11. Hwang, T.-L., P.C.M. van Zijl, and S. Mori. 1998. Accurate Quantitation of Water-amide Proton Exchange Rates Using the Phase-Modulated CLEAN Chemical

- EXchange (CLEANEX-PM) Approach with a Fast-HSQC (FHSQC) Detection Scheme. *Journal of Biomolecular NMR*. 11: 221-226.
12. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2008. A Billion-fold Range in Acidity for the Solvent-Exposed Amides of *Pyrococcus furiosus* Rubredoxin. *Biochemistry*. 47: 6178-6188.
  13. Best, R.B., and M. Vendruscolo. 2006. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure*. 14: 97-106.
  14. Hilser, V.J., and E. Freire. 1996. Structure-based Calculation of the Equilibrium Folding Pathway of Proteins. Correlation with Hydrogen Exchange Protection Factors. *Journal of Molecular Biology*. 262: 756-772.
  15. Murzin, A.G. 2008. Biochemistry. Metamorphic proteins. *Science*. 320: 1725-1726.
  16. Yin, Y.W., and T.A. Steitz. 2002. Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase. *Science*. 298: 1387 -1395.
  17. Tahirov, T.H., D. Temiakov, M. Anikin, V. Patlan, W.T. McAllister, D.G. Vassylyev, and S. Yokoyama. 2002. Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature*. 420: 43-50.
  18. Roche, S., F.A. Rey, Y. Gaudin, and S. Bressanelli. 2007. Structure of the Prefusion Form of the Vesicular Stomatitis Virus Glycoprotein G. *Science*. 315: 843 -848.

19. Littler, D.R., S.J. Harrop, W.D. Fairlie, L.J. Brown, G.J. Pankhurst, S. Pankhurst, M.Z. DeMaere, T.J. Campbell, A.R. Bauskin, R. Tonini, M. Mazzanti, S.N. Breit, and P.M.G. Curmi. 2004. The Intracellular Chloride Ion Channel Protein CLIC1 Undergoes a Redox-controlled Structural Transition. *Journal of Biological Chemistry*. 279: 9298 -9305.
20. Xu, M., A. Arulandu, D.K. Struck, S. Swanson, J.C. Sacchettini, and R. Young. 2005. Disulfide Isomerization After Membrane Release of Its SAR Domain Activates P1 Lysozyme. *Science*. 307: 113 -117.
21. Tuinstra, R.L., F.C. Peterson, S. Kutlesa, E.S. Elgin, M.A. Kron, and B.F. Volkman. 2008. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proceedings of the National Academy of Sciences*. 105: 5057 -5062.
22. Alexander-Brett, J.M., and D.H. Fremont. 2007. Dual GPCR and GAG mimicry by the M3 chemokine decoy receptor. *The Journal of Experimental Medicine*. 204: 3157 -3172.
23. Luo, X., Z. Tang, J. Rizo, and H. Yu. 2002. The Mad2 Spindle Checkpoint Protein Undergoes Similar Major Conformational Changes Upon Binding to Either Mad1 or Cdc20. *Molecular Cell*. 9: 59-71.
24. Skinner, J.J., S. Wood, J. Shorter, S.W. Englander, and B.E. Black. 2008. The Mad2 partial unfolding model: regulating mitosis through Mad2 conformational switching. *The Journal of Cell Biology*. 183: 761 -768.

25. Bai, Y., J.J. Englander, L. Mayne, J.S. Milne, and S.W. Englander. 1995.  
Thermodynamic parameters from hydrogen exchange measurements. In: *Energetics of Biological Macromolecules*. Academic Press. pp. 344-356.
26. Huyghues-Despointes, B.M.P., C.N. Pace, S.W. Englander, and J.M. Scholtz.  
Measuring the Conformational Stability of a Protein by Hydrogen Exchange. In:  
*Protein Structure, Stability, and Folding*. New Jersey: Humana Press. pp. 069-092.
27. Bai, Y., T. Sosnick, L. Mayne, and S. Englander. 1995. Protein folding  
intermediates: native-state hydrogen exchange. *Science*. 269: 192 -197.
28. Krishna, M.M.G., H. Maity, J.N. Rumbley, Y. Lin, and S.W. Englander. 2006.  
Order of Steps in the Cytochrome c Folding Pathway: Evidence for a Sequential  
Stabilization Mechanism. *Journal of Molecular Biology*. 359: 1410-1419.

## CHAPTER 2

### Using Pump-Probe Molecular Dynamics to Find Allosteric Pathways

#### 1. Introduction

Pump-probe molecular dynamics (PPMD) was developed by Dr. Kim Sharp during my brief time in his laboratory prior to my Ph.D. candidacy. While that work was largely driven by Dr. Sharp, I was able to contribute enough to be included as an author on the manuscript that introduced this technique (1). That work has been included here for the sake of completeness.

PPMD is a novel method for analyzing the dynamics of proteins. A set of oscillating forces are applied to a set of atoms or residues as part of a molecular dynamics simulation. How these forces propagate to other parts of the protein is probed using a Fourier transform of atomic motions. From this analysis, a coupling profile is determined which quantifies the degree of interaction between pump and probe residues. Various physical properties of the method such as reciprocity and speed of transmission are examined to establish the soundness of the method. The coupling strength can be used to address questions such as the degree of interaction between different residues at the level of dynamics, and identify propagation of influence of one part of the protein on another via “pathways” through the protein. The method is illustrated by analysis of coupling between different secondary structure elements in the allosteric protein calmodulin, and by analysis of pathways of residue–residue interaction in the PDZ domain protein previously elucidated by genomics and mutational studies.

Proteins are dynamic objects, and motions of a protein play an important part in their function. Functionally important conformational changes in proteins are typically driven by energies of only a few  $kT$ , provided, for example, by the binding of ligands or other proteins. Techniques such as NMR and hydrogen exchange (HX) provide detailed site resolved dynamic information on proteins, revealing the stability, extent, and time scale of motion of individual groups through HX protection factors (2), the generalized order parameter ( $S^2$ ), relaxation rates ( $\tau$ ), chemical shift averaging, and other quantities (3). Molecular dynamics (MD) simulations also provide a detailed description of protein motion and play an important role in the interpretation of experimental probes of protein dynamics. With the routine ability to do all atom simulations on multianosecond time scales and longer, the amount of information provided by these simulations is enormous. Analyzing the fluctuations in a useful way and relating them to specific experiments is nontrivial.

An important class of methods for studying protein motion is based on frequency analysis. An early example is the now classic method of normal mode (harmonic) analysis (4), which decomposes the possible motions of a protein around a minimum conformation into harmonic, orthogonal modes. An important insight from this analysis is that the lowest frequency modes represent the softest, most thermodynamically accessible ways a protein could change its conformation (the stiffness being proportional to the square of the frequency). These modes involve long range, concerted motions because they are low frequency, i.e. they have a large effective mass, and hence involve many atoms (5). A related method is the quasi-harmonic method that uses coordinate fluctuation covariance matrices obtained from MD to model modes of conformational

change (6–9). This allows for a limited amount of anharmonicity. The coordinate covariance matrices may then be analyzed in terms of eigen-vectors and subjected to the same frequency analysis as with normal modes. Principle component and essential dynamics analysis also use effective modes obtained from coordinate fluctuation covariances(10–12). However, extracting, interpreting, and using the modes obtained by this kind of frequency analysis is not easy (13–16). Alternatively, coupling between different atoms, residues, or segments may be analyzed directly using the covariance terms (17). Another approach is to use simplified harmonic models (Elastic network or Tirion type potentials; (18, 19) that can be combined with other treatments of large anharmonic motion (20) and sequence/mutation data (21). Fourier transformation (FT) and filtering of frequencies can be used to simplify and analyze MD trajectories (22, 23). Removal of high frequency motions allows clearer analysis of the putatively more interesting, or at least large scale, low frequency motions. Other methods take this a step further by actively manipulating selected frequency components of the velocity during MD simulations to probe, or drive conformational changes (24–27). Dynamics quantities such as amide and methyl NMR order parameters and relaxation rates can be obtained directly from MD simulations, and are most effectively obtained through the frequency domain via fast Fourier transform (FFT; (28–31).

The view of native protein motion as a superposition of oscillatory motions (harmonic or anharmonic) of different frequencies around a minimum energy conformation has provided an attractive model for allosteric (literally “other shape”) interactions (8, 20, 32, 33), which gives further impetus to frequency-based methods of analysis. The logic is that allosteric effects require interaction between spatially separated

sites. One such mechanism is via collective motion of a large spatial array of atoms, which in turn is characteristic of “low frequency modes” of protein motion. This model (dubbed here the low frequency mode model) naturally invokes analytical methods such as normal mode analysis, essential dynamics, and quasi-harmonic analysis. A different but not necessarily mutually exclusive view of allostery comes from many experiments, the specific residue–residue interaction model. This model does not derive from a low frequency mode view of protein motions, and different ways to analyze MD simulations are required if they are to help interpret these types of experiments.

The specific residue–residue interaction model for allostery emerges from many studies on different proteins with a variety of methods, of which we mention a few pertinent examples. Some classic examples involving oxygen carrying proteins and allosteric enzymes such as glycogen phosphorylase and phosphofructokinase have been reviewed in detail by Perutz (34). For example, in hemoglobin, oxygen binding to heme iron causes a flattening of the heme-porphyrin plane, which is transmitted to the distal histidine, then via a leucine and isoleucine on the F-helix, to movement of the end of the F-helix, and the CD loop in the cooperative  $\alpha 1\beta 2$  and  $\alpha 2\beta 1$  interfaces (35).

Using a statistical mechanical ensemble model of protein fluctuations Friere (36) traced the effect of substrate binding to Lysozyme at helix F through a specific pathway involving residues 24–37 on a neighboring helix, through residues 8–15 on the next helix on to residues in a sheet region on the opposite side of the protein.

In Calmodulin (CaM) residue Y138 has been shown to interact with residues E82, E78, Q79, and D80 of the linker region between helices D and E, which in turn interact

with helix A. These specific interactions are required for cooperativity and linkage of  $\text{Ca}^{2+}$  binding and peptide substrate binding (37–40).

In the PDZ class of proteins (a peptide binding domain found in signaling proteins) a combination of genomics analysis of sequences, mutations, and binding assays has traced specific residue–residue interactions necessary for allostery, for example from residue H76 through F29 and E57 to A51 on the other side of the protein (41). This coupling pathway has also recently been detected in dynamic behavior on the ps to ns timescale from changes in NMR-derived backbone and side chain order parameters (42). A similar sequence/mutation analysis of coupling has been done on the large class of G-coupled protein receptors (GPCR). Significantly, these networks of interactions are quite sparse, i.e. relatively few of the residues mediate allostery, and not all close residues interact in way relevant for allostery (43). More generally, recent NMR experiments show that mutations cause changes in dynamics that propagate along nonhomogenous, long range, and apparently specific paths (42, 44, 45).

Significantly, the specific residue–residue interactions that are mapped out by the experimental and genomic analyses described above are far from obvious by retrospective analysis of these systems in purely structural terms, i.e. in terms of distances between residues. It is hard to explain in purely structural terms why particular residue interactions are important, while others of equal or lesser distance are not. An extra dimension to the interactions must arise from the motions these groups undergo, and the coupling between them, i.e. from protein dynamics. However, in terms of the specific residue–residue interaction model for allostery it is difficult with existing MD techniques to frame and test simple hypotheses about coupling such as the following: Does residue X

influence Y, and by how much? Are they more strongly coupled than an arbitrary pair X-Z. How does influence propagate as a function of direction and distance? Can one detect pathways of allosteric action analogous to those found experimentally? Even experiments that probe simpler dynamic properties of proteins than allostery can be difficult to explain. An example is the NMR order parameter, which is a measure of the mobility of a single backbone or side chain group. CaM has anomalous order parameter data such as inverted temperature dependence for some residues (the order parameter increases with  $T$ ), and an unexpectedly large variation in methionine order parameters (46, 47). These observations imply significant correlation between motions of specific residues, but these correlations are not evident in covariance fluctuation matrices or standard frequency analysis (17, 48). Mayer et al. recently obtained detailed residue–residue correlation matrices in protein G from analysis of the correlation in NMR order parameter changes induced by mutations (49). A direct comparison with residue–residue correlations obtained by covariance matrix analysis of MD simulations found almost no relationship to the experimental correlations (50). These and other studies illustrate shortcomings with the available tools for conformational fluctuation analysis in trying to explain experimental data on protein dynamics. This led us to devise a new technique, pump-probe molecular dynamics (PPMD), to address these types of questions and to fill a gap in the simulation analysis toolbox.

## 2. Methods

### 2.1 PPMD

The method called here pump-probe molecular dynamics (PPMD) can be applied within any standard MD simulation using existing force field parameters and most simulation conditions. The basic PPMD protocol is as follows:

1. An atom or set of atoms to be pumped is selected.
2. An oscillating force of a specified magnitude, direction, and frequency  $\nu_0$  (period  $\tau = 1/\nu_0$ ) is applied to the pumped atom(s).
3. Coordinate snapshots are saved throughout the simulation.
4. After the simulation the fluctuation power spectra, or spectral density,  $W(\nu)$  of the motions of particular atoms or groups of interest (probe atoms) are obtained via Fast Fourier Transform (FFT) of their Cartesian coordinate trajectories. It should be noted that the power  $W(\nu)$  in this context is proportional to the contribution to the mean squared displacement of that atom from motions at frequency  $\nu$ : summation over the entire power spectrum yields total mean squared displacement of an atom over the simulation period.
5. The pump frequency region of each probe spectrum is examined to see how much the fluctuation power spectrum is increased by the pump, by comparing with the same region of  $W(\nu)$  from a control simulation without pumping force.

The basic PPMD protocol can be applied repeatedly with different pump atoms, force magnitudes, and pump periods as appropriate, to build up a detailed picture of how the pump impulses are transmitted throughout the protein. In the current implementation, pumping of atoms is done in a circular motion around the  $Z$ -axis with the same phase and direction of force for all pumped atoms. This application of the force had no particular rationale other than its simplicity, and it was chosen merely for the initial implementation and exploration of the technique. Obviously PPMD is not restricted to this way of applying the force, and other ways of applying it could be selected using some other physical considerations. For example, the axis of the applied circular force could be varied in case the  $Z$ -axis direction is atypical for any particular protein/group.

Pumped dynamics are not energy conservative. Elementary considerations show that the power deposited by a particular magnitude of pumping increases as the square of the period  $\tau$ . Pumping force magnitudes are communicated to the MD program CHARMM (51) in its units (kcal/mole/Å), but since we had no a priori information, suitable magnitudes were determined by experimentation, and typically were 1–3 in these units for the range of frequencies examined here. With the range of pumping forces applied in this work, we have found that the temperature control algorithms in standard MD simulation packages can handle the increase in energy, as judged by a stable value of  $T$  during the simulation, and negligible structure distortion.

In practice we found that the increase in power (contribution to rms fluctuation) at the pumped frequency was usually clear enough in the probe atom's power spectrum that comparison of a control power spectrum (no pumping force) of the same atom was not necessary. The increase in fluctuation power spectrum could be detected relative to the

baseline power spectrum to either side of  $\nu_0$  by a suitable peak finding algorithm, obviating the increase in noise inherent in any spectrum subtraction procedure.

## 2.2 Quantifying the Coupling, or Effectiveness of Transmission of the Pumped Motion

The displacement of an atom in response to a given force will vary depending on how stiff that region of the protein is. Thus, each fluctuation power spectrum is normalized by the sum over  $W(\nu)$  (which is just the mean square deviation of the atom) before comparison. After normalization of the power spectra, one can quantify coupling between atoms. If the increase in power of the pumped atom  $i$  at frequency  $\nu$  compared to the control simulation is designated as  $\delta W_i(\nu)$ , and the corresponding increase in power of a probed atom  $j$  is  $\delta W_j(\nu)$ , then the coupling constant at a particular pump frequency  $\nu_0$  can be defined as the ratio of total increase in power (relative contribution to rms fluctuation) of probe to pump atom.

$$C_{ij}(\nu_0) = \int_{\nu_0-\delta}^{\nu_0+\delta} \delta W_j(\nu) d\nu / \delta W_i(\nu_0) \quad (2.1)$$

$C$  is written in terms of integration over a frequency range centered on the pump frequency large enough to capture any frequency shifting in transmission to the probe atom due to the nonlinear nature of protein force fields. We allow for the possibility of frequency shifting so as not to miss the coupling, although we have encountered no detectable frequency shifting in applications thus far. If a group of atoms are pumped, the denominator contains the sum of  $\delta W_i(\nu_0)$  terms over the pumped atoms. Using this

coupling metric, one can compute a coupling profile through the sequence of a protein for any pumped frequency. Since  $C$  is a dimensionless quantity, and for a given protein/simulation condition only relative values across the sequence convey information,  $C$  in plots presented here have been “normalized” and shifted for clarity in graphing. Analysis of residue–residue coupling constants showed that very similar results were obtained if they were computed for the entire backbone of the probe residue (N, C, O, and  $C_\alpha$ ), its side chain atoms, or just its  $C_\alpha$  atom. Results presented here are for coupling constants using just the  $C_\alpha$  atom unless otherwise stated.

## **2.3 Comparison of Coupling Profiles**

To compare two coupling profiles of a protein, obtained for example before and after a mutation or other perturbation, or with different simulation conditions, two types of analysis were employed. These were developed bearing in mind that only relative values of coupling constant within a single profile have significance, and that most of a typical coupling profile is baseline, with relatively few peaks of interest.

### **2.3.1 Outlier analysis**

One set of coupling constants is treated as the independent variable  $x_i$ , the other set as a dependent variable,  $y_i$ , where  $x_i$  and  $y_i$  are the coupling values of the  $i$ th residue say before and after the perturbation. The two sets are subject to linear regression to the equation  $y = ax + b$ , yielding a minimized root mean squared deviation  $\sigma$  for the fit. The

scaled residual  $r_i = (y_i - (ax_i + b))/\sigma$  is computed for each residue. A residual of less than  $-1.5\sigma$  or greater than  $1.5\sigma$  indicates an outlier from the regression, i.e. a significant decrease or increase in coupling, respectively, at that residue due to the perturbation.

### 2.3.2 Percentile analysis

One set of residue coupling constants ( $y_i$ ) is plotted against the other ( $x_i$ ). The resulting scatter plot typically shows the majority of points in a cluster with low coupling constant in both simulations (Fig. 2.7). This cluster contains the “non-interesting” coupling constants in any profile that are close to baseline, i.e., have coupling that is weak or below the noise level in both simulations. The more interesting coupling constants are the stronger ones corresponding to peaks in the coupling profile. The values that divide the upper 10th percentile from the lower 90th percentile are calculated for both sets of coupling constants, designated  $x_{0.1}$  and  $y_{0.1}$ , respectively. Vertical and horizontal lines are drawn at  $x = x_{0.1}$  and  $y = y_{0.1}$ , respectively. This divides the plot into four regions: lower left, containing residues with low, baseline level coupling in both simulations; upper right, containing residues that have stronger (upper 10th percentile) coupling in both simulations; upper left and lower right regions containing residues that show a coupling peak in one simulation but not the other, i.e. with a significant difference. If for example all the points fell into either the lower left “baseline” region or the upper right “peak” region there is complete agreement between the two coupling profiles at a 10% significance level. Put another way, there is a peak in one simulation if and only if there is a peak in the other. The 10%/90% division was chosen to reflect the

typical ratio of residues in peak vs. baseline regions of the coupling profiles analyzed here. The ratio would be adjusted for profiles with less/more peak regions.

Both forms of comparison have the advantage that they are sensitive to the shape of the coupling profile, i.e. the number and location of peaks. They are insensitive to a difference in average coupling between two simulations, or a difference in the range of coupling values between low and high that might occur through differences in scaling, or systematic but nonspecific differences arising from simulation conditions. We note that the linear regression  $R^2$  value alone does not provide a good way either to measure the similarity of profiles or detect differences in peaks: first, the  $R$  value is typically always low, dominated as it is by the uncorrelated cluster of baseline points. Second, a low  $R$  would occur even if all the peaks occupy in the same positions in the two profiles but they are of different heights.

## **2.4 Frequency and Correlation Analysis**

FT, time autocorrelation functions, and time cross-correlation functions for atomic motions were computed from CHARMM format output trajectories using the FORTRAN versions of the routines FFT and CORREL described in Numerical recipes (52). Time autocorrelation functions for the Brownian harmonic oscillator (BHO) model were computed from standard analytical expressions (53). The behavior of a BHO is governed by the dimensionless ratio of friction coefficient to harmonic force constant  $G = \gamma/2\omega_0$ , where  $\gamma$  is the friction coefficient per unit mass, and  $\omega_0$  is the harmonic oscillator frequency. Values of  $G < 0$ ,  $= 0$ , and  $>0$  correspond to under-damped, critically damped,

and over-damped conditions, respectively. To match the approximate time scale of relaxation seen in time autocorrelation functions derived from explicit atom simulations of proteins, a constant oscillator frequency of  $\omega_0 = 0.05$  radian/ps (oscillation period of 125.7 ps) was chosen with friction coefficients of 0.033, 0.1, or 0.2 ps<sup>-1</sup> yielding  $G = 0.33$  (under-damped), 1 (critically damped), and 2 (over-damped), respectively.

## 2.5 MD Protocol

The PPMD method has been implemented in a Fortran77 subroutine that is called by CHARMM (51). For the purposes of developing and testing the PPMD methodology, we used the following standard simulation parameters/conditions: CHARMM version 27 force field with all atoms (54), the Verlet algorithm with a time step of 1 fs, a temperature of 298 K held constant by periodic velocity reassignment, a nonbond cutoff of 14 Å with force shifting, and a total simulation time of at least 10 times the pump frequency period, usually longer than 1 ns. The number of simulation steps and frequency of coordinate saving were adjusted so that exactly  $2^n$  snapshots were generated for FFT. This obviates padding or truncation, and maximizes precision. For most of the simulations we used an approximate but rapid solvent treatment by using the distance dependence dielectric option in CHARMM with a constant of 4. While this model has well-documented shortcomings, it is rapid, and given the large number of simulations required to develop and test the PPMD method, this is an acceptable trade off. The impact of this on our method is assessed with several control simulations as described in results.

For simulations using Langevin dynamics (LD), the Langevin MD command in CHARMM was used. Friction coefficients  $\gamma$  were assigned to all non-hydrogen atoms using

$$\gamma = 6\pi\eta\alpha f/m \quad (2.2)$$

where  $\eta = 0.89$  cP is the viscosity of water,  $m$  is the atomic mass,  $a$  is the atomic radius in the CHARMM forcefield, and  $f$  is the fraction of exposed surface area of that atom, calculated using the program SURFCV (55). Thus, buried atoms experience no solvent friction/fluctuation force, while exposed atoms typically have friction coefficients in the range 50–90 ps<sup>-1</sup>. Hydrogens are assigned a friction coefficient of zero since they are constrained to the heavy atom positions using the SHAKE algorithm (56).

For simulations using NMR type nuclear Overhauser effect (NOE) type distance restraints, the CHARMM command NOE was used. A list of restraints was generated for each pair of C $\alpha$  atoms within 10 Å of each other in the starting structure. For each pair of C $\alpha$  atoms an NOE type restraint was applied using a harmonic potential with force constant 1 kcal/mole/Å. The force was applied if the distance varied by more than 2.5 Å from the starting structure, other wise no force was applied, i.e., a 5 Å zone of fluctuation with no restraining force was allowed.

## 2.6 Additional PPMD Enhancements

In our initial application of PPMD the pumped frequency does not correspond to any special frequency of natural motion in the protein, (like a prominent normal mode).

Thus, it is necessary to scan pumping over as wide a range of frequencies as possible to build up a complete picture of the coupling. To reduce the amount of simulation we investigated pumping at several frequencies simultaneously. Provided the frequencies are well separated, we found that the effects were independent (additive). All results presented here were thus obtained using four simultaneous pumping frequencies: a base frequency with period  $\tau$ , and three others with periods  $3\tau$ ,  $5\tau$ , and  $7\tau$ , respectively.

To explore the sensitivity of the PPMD method simulations of increasing pump magnitude were performed. It was found, not surprisingly, that larger magnitudes eventually resulted in structural distortions in the protein by the end of the simulation. However, by using the NOE-type restraint facility in CHARMM applied to the  $C\alpha$  atoms we could pump with larger magnitudes and so increase sensitivity. NOE-type restraints are convenient for this purpose because they apply no force as long as the distances stay within the upper and lower bounds. These restraints thus can keep a simulation conformation within reasonable bounds while providing minimal bias to the dynamics. PPMD results with and without NOE-type restraints described below show very similar coupling behavior.

## **2.7 Proteins Studied**

For the initial development and testing of PPMD, we selected two proteins for which good structures were available, and for which detailed experimental data related to function and allostery was available, CaM and the PDZ domain protein. More

specifically, for each protein a variety of experiments and analysis has demonstrated specific residue–residue couplings of functional importance.

### **2.7.1 CaM (PDB entry 1CDL (57))**

CaM is a calcium-regulated protein involved in signal transduction, trafficking, muscle contraction, and many other cellular processes. It is known to recognize over 200 targets (58), and so it is central in cell regulation and signaling. CaM binds four calcium ions, forming a dumb-bell like structure with two globular domains each containing two ions separated by a long helical domain. Studies of the CaM system include the structural basis for its protein target recognition (38, 59), the thermodynamics of calcium binding (60), the thermodynamics of smMLCK target peptide binding (61), and the concerted conformational change upon peptide (62). Wand et al. have an extensive set of dynamic information for calcium-loaded CaM/smMLCK peptide complex, including NMR order parameters ( $S^2$ ) and relaxation times ( $\tau$ ) for almost all the amide and methyl groups over the range 15–73°C, and additional order parameter data for the uncomplexed CaM, and for CaM mutants. This work has resulted in an unprecedented amount of site resolved dynamics data on a protein, and interesting dynamic behavior that we have only partly been able to explain with standard MD simulation analysis (17, 48). There is also an extensive series of studies systematically exploring the relationship between fluctuations, conformational changes, binding, and cooperativity (37–40, 63). This experimental work has elucidated specific residue interactions involved in allostery, for example between residue Y138 and helix E cap residues Glu82, and between E78, Q79,

and D80 of linker region and residues on helix A. CaM shows two levels of allosteric interaction: between calcium and peptide binding and peptide binding and large binding related changes in conformation.

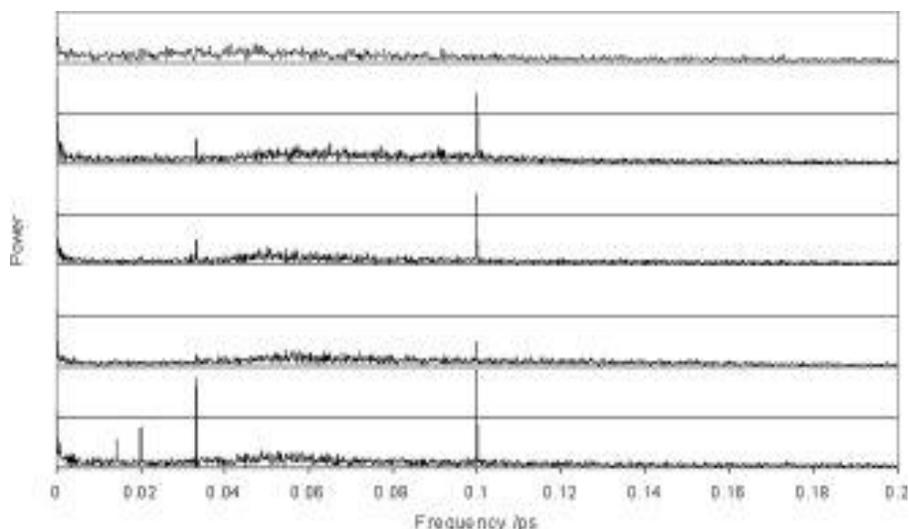
### **2.7.2 PDZ domain protein (PDB entry 1BE9 (64))**

PDZ domain proteins are a family of modular peptide binding domains found in many cytosolic signalling proteins (65). Extensive sequence data and several high resolution structures are available. Lockless and Ranganathan (41) have used evolutionary sequence analysis and mutation/function analysis to identify specific coupling at the level of residue–residue interactions. These interactions form a network that spans a significant distance in space (i.e. they are allosteric in nature). By examining the spatial sequence of pairs of couplings they can trace pathways of communication. For example residue H76, which is crucial for determining peptide binding specificity, is coupled, through residues F29 and E57 to A51 on the opposite side of the protein. Their combined genomics and experimental analysis also identifies other couplings between some, but by no means all, close residues. The fact that in their analysis not all close residues are coupled forms an important control for the concept of specificity in residue–residue interactions. Recently, Fuentes et al. have shown that very similar coupling pathways are revealed by changes in NMR-derived side chain and backbone order parameters upon peptide binding to PDZ (42), a direct demonstration of long range propagation of signals.

Starting structures of the single site mutants G33A, G34A, F29K, F44A, S75T, A80V, K84A, and T89S were generated from the wild-type structure (pdb entry 1BE9) using CHARMM by changing the residue template, rebuilding the atoms of the mutated residue and then minimizing the structure using the adaptive basis Newton–Raphson minimizer for 1000 steps. PPMD simulations were then run on the mutants under the same conditions as the wild type.

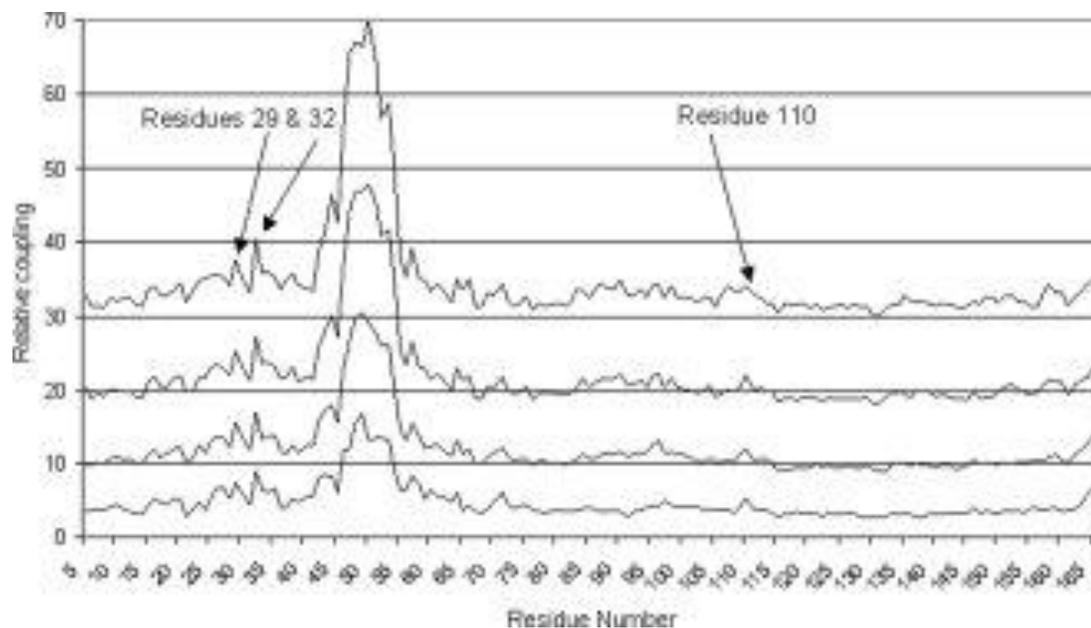
### **3. Results**

PPMD simulations were run on CaM, pumping residues 46–53 in helix C with oscillating forces of 10, 30, 50, and 70 ps period. Figure 2.1 shows typical fluctuation power spectra obtained from a 5 ns simulation. The upper plot shows a typical power spectrum of an atom from a control simulation with no pumping. The lower plot shows the spectrum of a pumped atom in residue Gln49. Against the rather featureless background four sharp peaks at the pumping frequencies can easily be distinguished as one would expect since this is a pumped atom. The spectra at selected probe atoms in residues T29 and L32 in the neighboring helix and a more distance residue T110, which neighbors residues 29 and 32 but not the pumped helix, show similar features except the spikes have different relative intensities, some being undetectable altogether. This indicates different degrees of transmission to different atoms and at different frequencies.



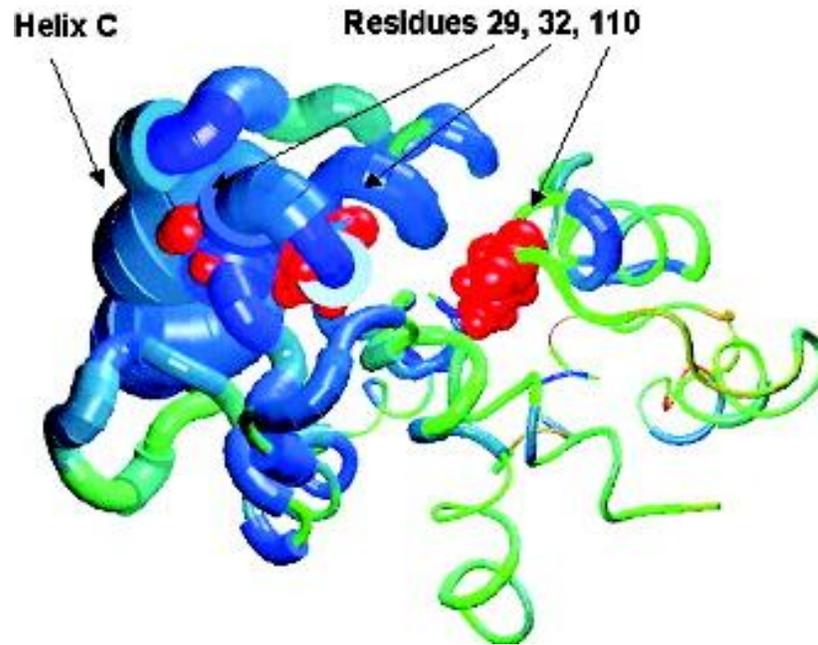
**Figure 2.1** Fluctuation power spectra of some atomic motions in CaM with 10, 30, 50, and 70 ps period pumping forces applied to helix C, over a total simulation time of 5 ns. Traces are displaced vertically for clarity, and from the bottom up are for the  $\text{Ca}$  atom in residues Q45 (a pumped residue), T110, L32, and T29. Top trace: Fluctuation power spectrum of  $\text{Ca}$  atom of Q54 in a control simulation with no pumping force.

Figure 2.2 shows the coupling profile for the simulation in Figure 2.1, obtained from analyzing the fluctuation power spectra for all the residues, extracting the relative intensities at the pumped frequencies, and applying Eq. 2.1. The coupling profiles exhibit a large peak at the pumped residues, indicating trivial coupling of the pumped residues with themselves. More significantly, there are peaks in coupling at more distant residues in both sequence and space, e.g. at residues 29, 32, and 110, indicating nonhomogeneous spread of the pumped energy and specific couplings.

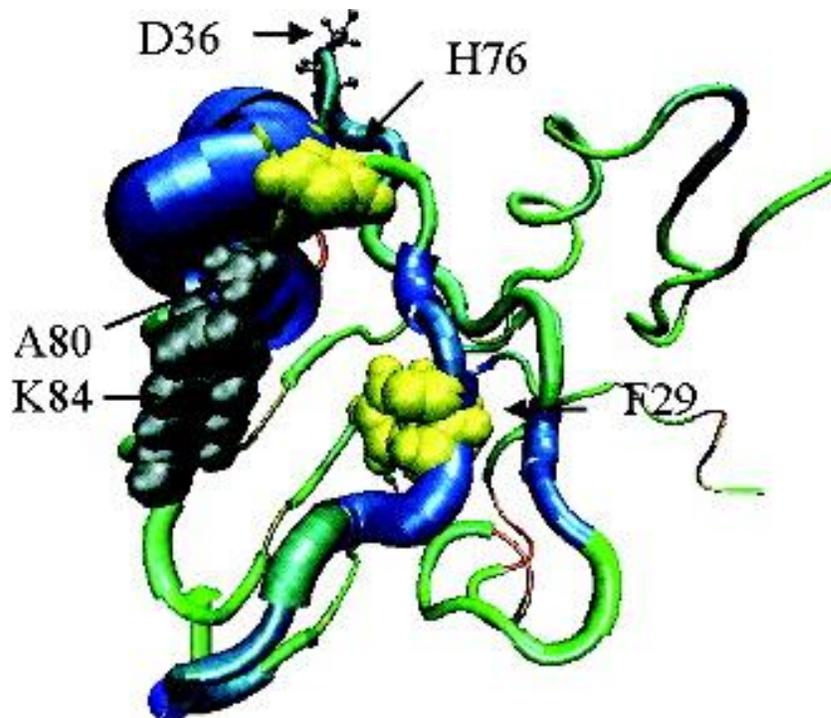


**Figure 2.2.** Coupling profiles for CaM with the C helix pumped. Traces are displaced vertically for clarity. From bottom up, couplings for 10, 30, 50, and 70 ps period pumping. Probe residues whose fluctuation power spectra are shown in Figure 2.1 are labeled.

An alternative way to display the coupling is through a coded structural representation of the protein (Figs. 2.3 and 2.4). Here the thickness of the backbone worm indicates the degree of coupling, and enables one to see which parts of the protein are more coupled, and where they are. Figure 2.3 shows a PPMD simulation on CaM where helix C was pumped at 10, 30, 50, and 70 ps. The coupling profile at 10 ps is coded on the figure. The pumped helix C shows strong coupling to the B helix, which contains residues 29 and 32 and forms one flap closing over the peptide. There is “follow-on” coupling to the helix turn containing residue 110 on the opposite flap. Residues 29, 32, and 110, whose fluctuation power spectra are shown in Figure 2.1, are rendered in CPK.



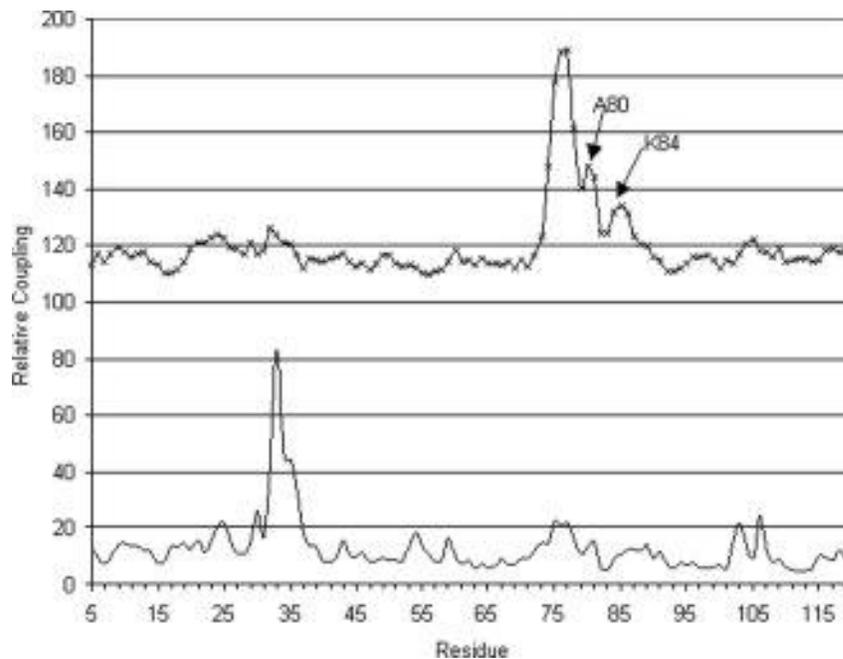
**Figure 2.3** Coded structural representation of CaM. Worm thickness is proportional to coupling to Helix C at the 10 ps oscillation period. Probe residues whose fluctuation power spectra are shown in Figure 2.1 are rendered in CPK.



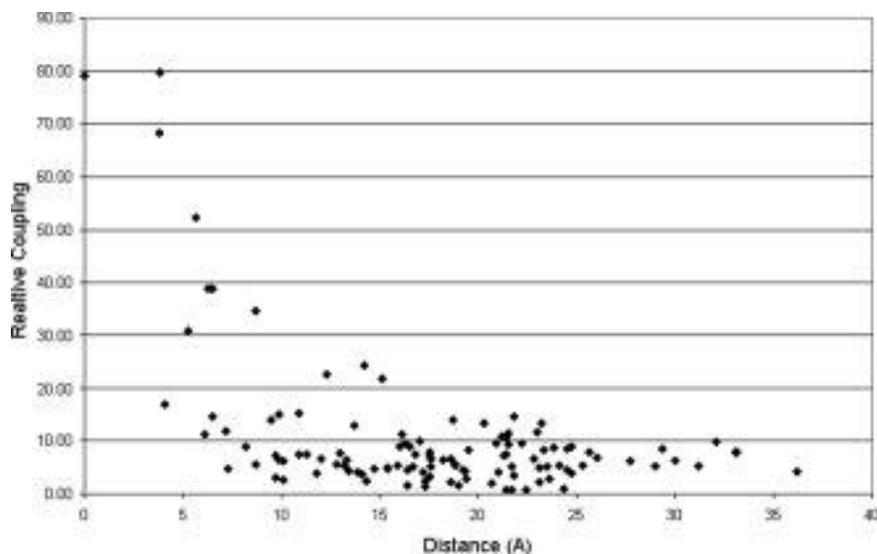
**Figure 2.4** Coded structural representation of PDZ domain protein. Residue H76 was pumped. Worm thickness is proportional to coupling to H76 at the 10 ps oscillation period. Key coupling pathway residues are shown in CPK.

Figure 2.4 shows a similar representation for a PPMD simulation of the PDZ domain protein. In this case residue H76, a key coupled residue identified by Lockless and Raganathan (41) was pumped at 10, 30, 50, and 70 ps. The coupling at 10 ps is shown on the figure. This figure shows that coupling from the pumped residue extends down the helix containing H76 and also across to the  $\beta$  strand containing F29. Interestingly, coupling is much less effective in the other strand direction from H76: round the turn. Heterogeneous coupling to other regions is also apparent. Since the coupling is frequency dependent, average coupling profiles using data from multiple frequencies provide a better picture of the coupling through the protein and increase the sensitivity. To obtain average coupling profiles, profiles from eight different pumping periods, 1, 3, 5, 7, 10, 30, 50, and 70 ps, were averaged. Figure 2.5 shows two such coupling profiles for the PDZ protein. Each coupling profile represents a set of simulations in which either residue H76 or G33 was pumped. These two residues were shown experimentally to be coupled (41). Each profile shows a peak in the region of the pumped residue, and other peaks. In the simulations where H76 is pumped, coupling down one side of the helix to the residues one and two helical turns down (A80 and K84) is clearly seen, recapitulating the pathway seen by Lockless and Ranganathan (41). Significantly, one sees a peak at G33 when H76 is pumped, and a peak at residue H76 when pumping G33, i.e. the coupling is reciprocal. To better illustrate the heterogeneous nature of the coupling, in Figure 2.6 the residue coupling constants from the H76 pump simulation of PDZ domain illustrated in Figure 2.5 (upper profile) are plotted against the distance of each residue from the pumped residue H76. The figure illustrates a general decrease in coupling with distance, as one would expect as energy is dissipated across the

protein. However, examining the couplings at a given distance in the 5–15 Å region shows a wide range of coupling strengths at each distance, notably even at the very shortest distances that represent neighboring residues.



**Figure 2.5** PDZ domain coupling profiles for pumping at H76 (-x-, trace displaced upward for clarity) and G33 (no symbol). Arrows indicate residues one and two helical turns down from pumped residue H76, as shown in Figure 2.4. Profiles are profiles averaged over 1, 3, 5, 7, 10, 30, 50, and 70 ps pumping periods.

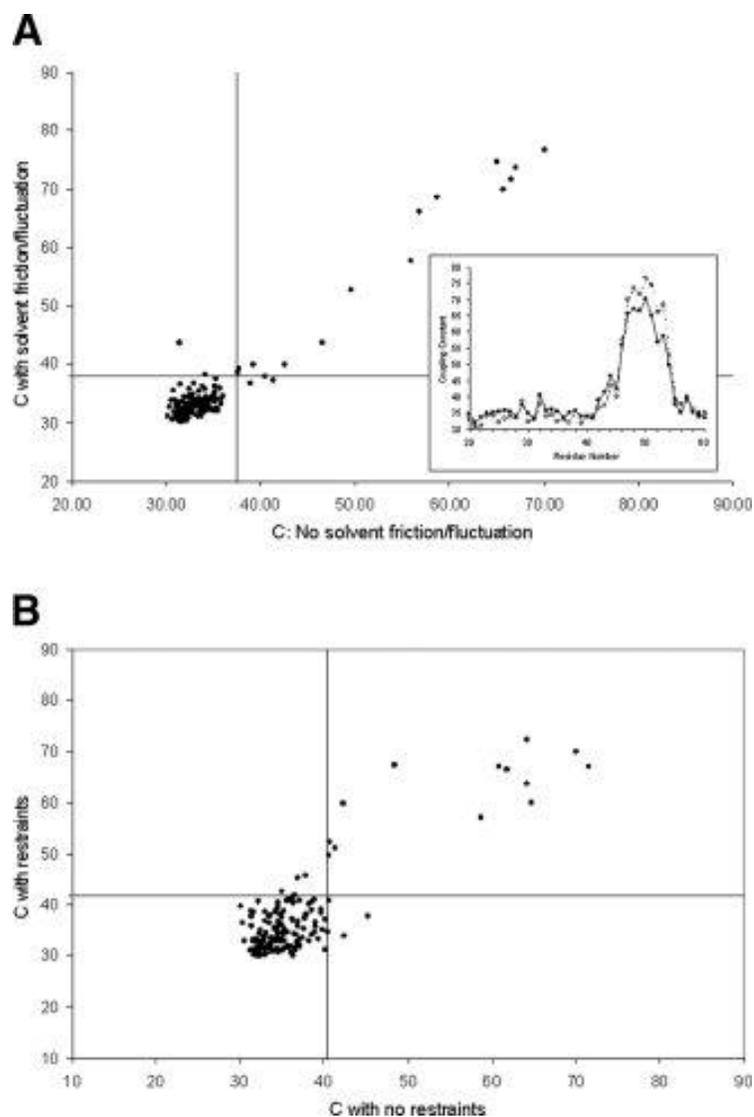


**Figure 2.6** Plot of coupling strength vs.  $C\alpha$ - $C\alpha$  distance for PDZ domain pumped at H76. Coupling strengths are taken from the profile averaged over 1, 3, 5, 7, 10, 30, 50, and 70 ps pumping periods (Fig. 5.5). Distances are measured from the pumped residue H76  $C\alpha$ .

These coupling profiles illustrate several key features revealed by the PPMD simulations:

- i. Heterogenous coupling in space. Not all residues close to the pumped residue(s) are coupled to the same extent (Fig. 2.6).
- ii. Long range coupling. Significant transmission is observed, in some cases from one secondary structure element across another to a third as illustrated in Figures 2.2 and 2.3.
- iii. Coupling does not necessarily occur via the shortest pathway, as illustrated by the coupling of F29 to H76 via A80 and K84 in PDZ domain protein (Fig. 2.4).
- iv. Reciprocity of coupling between pump and probe residues, as illustrated in Figure 2.5.

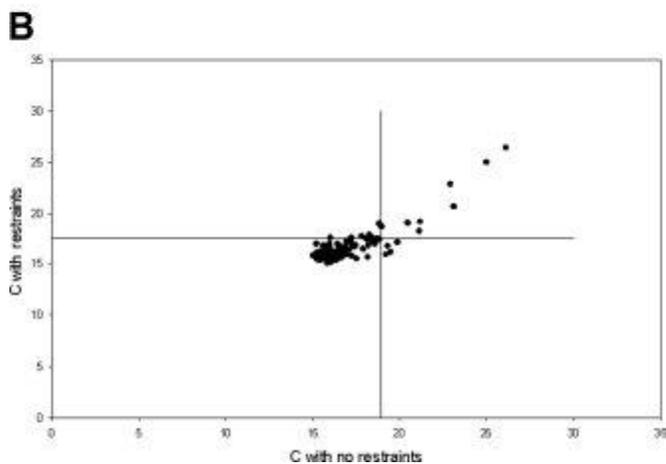
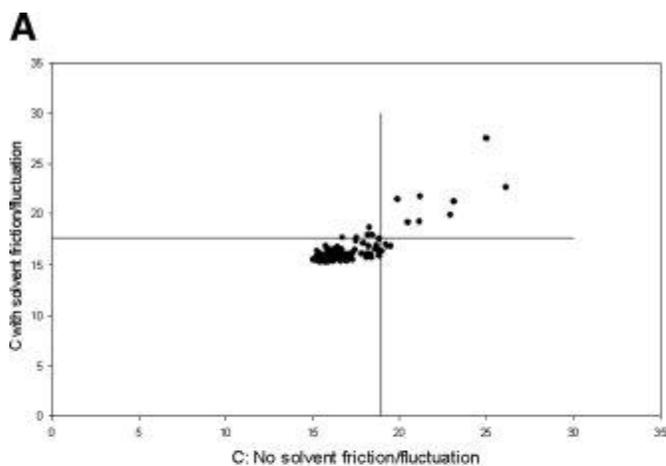
The simulation protocol used for the initial development and testing of the PPMD method was chosen for its speed and to maximize the sensitivity to the pumping signal. However, there are two simulation conditions used here that could potentially affect the coupling profiles. The first is the use of an implicit solvent model. The most serious concern with this model for the PPMD method is the lack of solvent friction and fluctuations. Solvent friction/fluctuation could more rapidly damp the PPMD signal, and so reduce the sensitivity of the method and its ability to detect long range coupling. Also, since solvent does not act uniformly on the protein, but preferentially on surface atoms, it could potentially alter the coupling profiles in a qualitative way. To examine these two possibilities we ran control simulations in which solvent friction/fluctuation on solvent exposed atoms was included via CHARMM's LD facility. Since the solvent friction forces in the LD method would actually be more random than those from an explicit solvent simulation (in which there is a degree of correlation between solvent and protein atom motions due to force reciprocity), the LD control should provide an upper estimate of the effect of neglecting solvent friction/fluctuation. Figure 2.7A shows a comparison of the coupling constants obtained at a pump period of 70 ps applied to the C helix of CaM. The simulation with no solvent friction/fluctuation is the same as that depicted in Figure 2.2. The simulation with solvent friction was run with a factor of three larger pumping force. The inset to the figure shows an expanded section of the coupling profile between residues 20 and 60, showing that the profiles from the no-friction simulation and the control LD simulation with solvent friction are very similar.



**Figure 2.7** Percentile analysis of similarity of coupling profiles for CaM pumped at Helix C with a 70 ps period. **(A)** No solvent friction (abscissa) vs. solvent friction (ordinate). Inset shows part of both coupling profiles overlaid to illustrate similarity. **(B)** With NOE restraints vs. no restraints (ordinate).

To better compare the two coupling profiles we plotted the “with friction” coupling constants against the “no friction” coupling constants for the 70 ps pump period, and used the method of percentile analysis as described in the methods section. The upper 10th percentile lines divide the graph into four regions. Using this analysis we see that Figure 2.7A shows excellent correspondence in coupling profiles with and without

solvent friction, with most of the points falling in the lower left baseline regions, or the upper right region indicating a peak found in both simulations. Only one or two residues fall in each of the upper left and lower right regions that indicate a peak in one case but not another. Figure 2.8A shows the same comparison of friction/no friction for a simulation of the PDZ domain, pumping residue H76 at 10 ps. Again, relatively few points fall in either of the upper left or lower right regions that would indicate differences in peaks. Similar levels of agreement are obtained for coupling at the other three pump periods in each simulation. We conclude that solvent friction does not change the coupling profiles qualitatively, although it requires a larger pumping force to get the same signal to noise ratio which increases the risk of structure distortion.



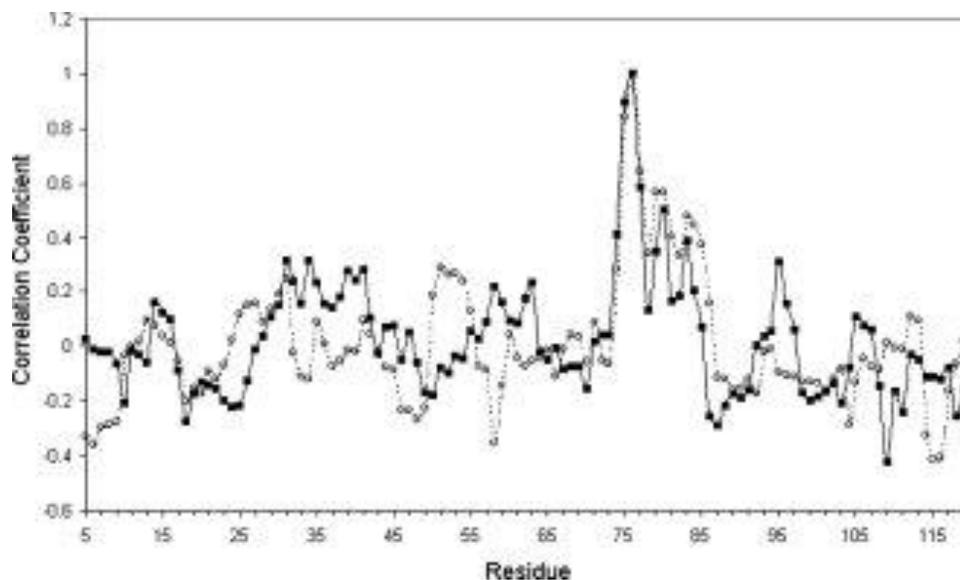
**Figure 2.8** Percentile analysis of similarity of coupling profiles for PDZ domain protein pumped at residue H76 with a 10 ps period. **(A)** No solvent friction (abscissa) vs. solvent friction (ordinate). **(B)** With NOE restraints vs. no restraints (ordinate).

The second compromise in the simulation protocol was to use NOE-type restraints on the protein C $\alpha$  atoms so as to use a larger pumping force to improve signal to noise without distorting too much, or unfolding the protein. While NOE constraints have no force bias when they are satisfied (the majority of the time), they do apply a harmonic restoring force to atoms outside the distance bounds, and this harmonic force may introduce its own coupling behavior and so distort the coupling profile. Figure 2.7B shows a comparison of the coupling profiles obtained at a pump period of 70 ps applied to the C helix of CaM with and without NOE type restraints. Figure 2.8B shows the same comparison for PDZ domain with residue H76 pumped with a 10 ps period. The percentile analysis again shows that there is very good correspondence between the peaks in the two profiles for each protein. The simulations without restraints were done with a factor of 5 lower pumping force to avoid protein shape distortion, i.e. they have a poorer signal to noise ratio which requires a longer simulation. Thus as long as mild NOE type restraints do not qualitatively affect the coupling profile, they have the advantage of better signal to noise and shorter simulation time. There may however be situations in which the type of coupling that is being probed requires rather large excursions of the protein to occur. In this case NOE-type restraints would tend to oppose or mask the coupling, and so they would not be desirable.

In Figure 2.9 the standard fluctuation cross-correlation coefficients

$C_{ij} = \langle \Delta x_i \Delta x_j \rangle^2 / \langle \Delta x_i^2 \rangle \langle \Delta x_j^2 \rangle$  between the C $\alpha$  of H76 and other residue C $\alpha$ 's are shown for

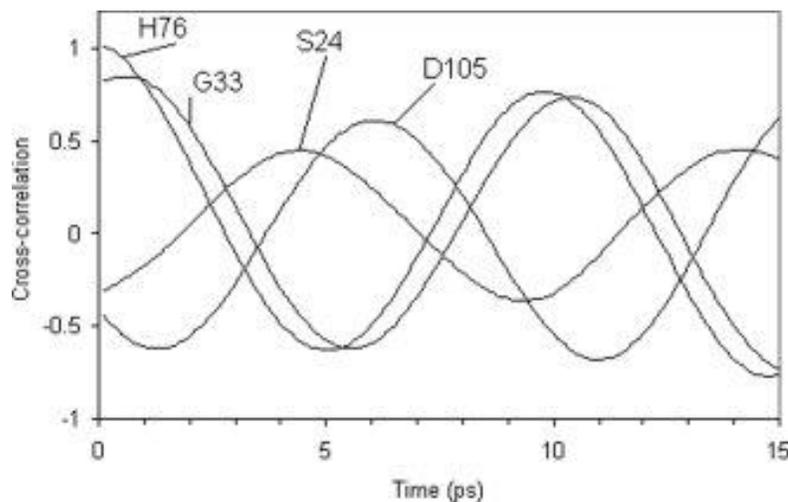
the PDZ domain protein. The plot shows the H76 row extracted from the  $C_\alpha$  correlation matrix from a control simulation (no pumping) performed for exactly the same length and conditions as a pumped simulation. The simulation was divided into two halves, and the correlation coefficients calculated for each half to assess the variability. The result is typical of fluctuation correlation analysis, i.e. the residue in question H76, is perfectly correlated with itself, and there is significant coupling to its immediate neighbors. The large majority of correlation coefficients have magnitudes less than 0.5, and moreover most of these vary widely between the first and second halves of the simulation indicating that they are not significant at this length of simulation, at least. The pattern of PPMD couplings, in contrast, persist across different frequencies (Fig. 2.2) and different simulations (Figs. 2.7 and 2.8). Coupling between more distant (in sequence) residues, e.g. H79 and F29, is not apparent from the correlation analysis, nor is any systematic pathway of coupling. Thus, at least for the length of simulation required for PPMD (1–2 ns) this technique has better signal to noise, and is more revealing of coupling pathways than conventional correlation coefficient analysis.



**Figure 2.9** Fluctuation cross-correlation coefficient  $C_{ij} = \langle \Delta x_i \Delta x_j \rangle^2 / \langle \Delta x_i^2 \rangle \langle \Delta x_j^2 \rangle$  between the C $\alpha$  of H76 and other residue C $\alpha$ 's for PDZ domain protein from a control simulation (no pumping). From batches 1 to 8 (filled squares), and from batches 9 to 16 (open circles).

It should be stressed that PPMD is not measuring the mobility of atoms or residues per se i.e. their root mean square (rms) deviation, but the change in mobility due to coupling, or transmission of energy/motions from the pumped residue(s) to the probed residue(s) as measured by the increase in fluctuation power at particular frequencies. Thus, a coupling profile *does not* reflect differences in intrinsic flexibility along the polypeptide chain. The effect of intrinsic differences is removed by the normalization of each atom's fluctuation power spectrum by that atom's rms deviation (which is given by the sum over the power spectrum) prior to the calculation of coupling via Eq. 2.2. The coupling profile does provide a measure of the interaction between specific atoms or groups at the level of dynamics. To explore the transmission characteristics further, we used time cross-correlation analysis. The cross-correlation function quantifies the correlation in motion between two atoms or groups for different time lags. The same FFT

transforms of coordinate trajectories used for the fluctuation power spectra can be used to obtain these time cross-correlation functions in a straightforward manner (52). From the time axis offset one can obtain the phase delay at different points in the protein. An example is shown in Figure 2.10, which illustrates the transmission of pumped impulses applied at residue H76 of PDZ domain protein. For the time range shown in the figure, the cross-correlation functions are dominated by the highest frequency pump applied in this simulation, the 10 ps period pump. The profile for the pumped residue is shown for reference, and has a phase shift of zero. Various phase shifts are observed at other atoms. While the time delay generally increases with distance it is not just dependent on distance, as Table 2.1 shows. For example the impulse arrives at residue S24 before residue D105, even though the former residue is farther away. Thus, there is heterogeneous propagation speed through different parts of the protein. These results suggest that variation in distance-normalized transmission speed is another way to quantify coupling strength in the PPMD analysis. This will be examined further in a future study.



**Figure 2.10** Time cross-correlation functions for C $\alpha$  atoms in PDZ domain protein pumped at H76. Distances from the pumped residue are given in Table 2.1.

**Table 2.1** Phase Delay of a 10 ps Pump Applied to H76 of PDZ. Distance is from H76 C $\alpha$  to residue C $\alpha$

Residue	Lag (ps)	Distance (Å)
H76	0	0
G33	0.7	6
S24	4.5	22
D105	6	16

The other information obtained from the time cross-correlation analysis is the general rate of transmission of impulses through a protein. The effect propagates at speeds of up to 5 Å/ps, and so it could cross a medium size protein within several picoseconds. It should be noted that this is well within the time scale of the simulations. The upper limit for transmission of any mechanical impulse in a medium is the speed of sound,  $V_s$ . While this is not known exactly for proteins, and it also depends on whether the waves are longitudinal or transverse, and how homogenous the mechanical properties are,  $V_s$  may be estimated from the relationship:

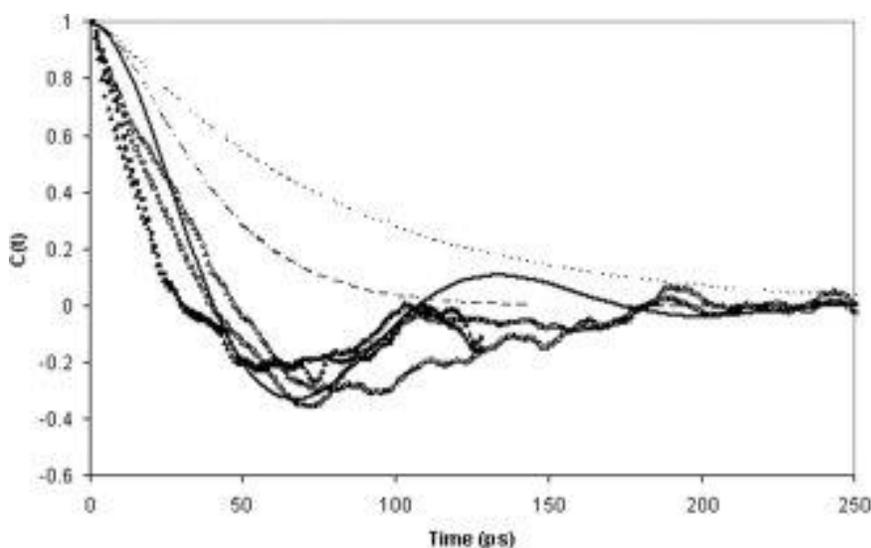
$$V_s = \sqrt{\frac{Y}{\rho}} \quad 2.3$$

where  $Y$  is the Young's or bulk modulus and  $\rho$  is the density. Since  $Y$  is the inverse of the compressibility, it may be estimated from experimental measurements of the latter as  $Y \approx 1/(5 \times 10^{-6}/\text{Atm}) = 2 \times 10^{11}$  dynes/cm<sup>3</sup> (66). With a typical protein density of 1.3 g/cm<sup>3</sup>, Eq. 2.3 gives  $V_s = 14,000$  km/h or 39 Å/ps. This estimate lies somewhere between that of water (14 Å/ps) and steel (50 Å/ps), which are accurately estimated by Eq. 2.3 (67), and it thus seems reasonable. From this we conclude that the transmission rates derived from

the data in Table 2.1 are similar enough to  $V_s$ , but far enough below the estimated upper limit to be reasonable.

It should be noted that the time cross-correlation functions are normalized by the mean squared motions of the two atoms, and so Figure 2.10 only shows the time lag in transmission and not the decay in intensity with distance. The latter is more accurately depicted by the coupling profiles in Figures 2.2 and 2.5. To better understand how transmission of an impulse such as that applied by PPMD might be dissipated in a dense, fluctuating medium such as a protein, we compared the general dynamic behavior of the protein to the BHO model. The BHO model is an analytically treatable stochastic model of a system undergoing fluctuation and frictional dissipation around some minimum energy state, with a harmonic restoring force (53). The BHO model is effectively a one-dimensional analogue of a stable protein undergoing fluctuations around the minimum energy “native state” structure. The time autocorrelation functions for atomic motions in non-PPMD simulations of CaM with implicit and explicit solvent were calculated from the MD trajectories again using FFTs. Characteristic profiles are shown in Figure 2.11. These are compared to the BHO model using friction and harmonic force constant parameters chosen to match the protein dynamics time-scale. The BHO model has three regimes: where friction forces dominate and there is no oscillatory behavior (over-damped); where the restoring force dominates and the system shows oscillatory behavior (under-damped), and an intermediate regime (critically damped); and where the system has the fastest relaxation and which also shows no oscillatory behavior. As can be seen from Figure 2.11, the protein behaves as an under-damped system with a characteristic time scale in the 10–100 ps range. Interestingly, the protein motion has strong oscillatory

characteristics even with explicit solvent, even though one might expect the frictional effect of the solvent to make the system behave as an over/critically damped one. The under-damped, oscillatory character of protein dynamics is consistent with the detectable transmission of pumped oscillatory impulses over the same length scale as the protein revealed by the PPMD simulations.



**Figure 2.11** Time autocorrelation functions for atomic motions of  $C\alpha$  of Asp58 ( $\blacksquare$ ,  $\square$ ) and Glu 67 ( $\blacktriangle$ ,  $\triangle$ ) in CaM simulated with implicit solvent (empty symbols) or explicit solvent (filled symbols). Brownian harmonic oscillator that is under-damped (—), over-damped (...), or critically damped (- -).

A key experimental tool to detect residue–residue coupling is site directed mutagenesis. For example a series of PDZ mutants examined by Lockless and Raganathan (41) revealed very similar coupling to that obtained from their multiple sequence analysis. We examined eight of their mutants using PPMD simulations. Three of these mutants, F44A, S75T, and T89S, involved residues that were determined by sequence analysis and mutation not to be coupled to H76. The other five mutants, G33A, G34A, F29K, A80V, and K84A, involved residues that were found experimentally to be

coupled to H76. PPMD simulations were performed on each of the mutants by pumping at 1, 3, 5, 7, and 10 ps at either H76 or F29, and the average coupling profiles were computed. Each mutant coupling profile was compared to the corresponding wild-type coupling profile (pumped at either H76 or F29), and significant changes in coupling peaks caused by mutations were determined by the outlier analysis. The results are summarized in Table 2.2. The most notable result is for residue 36, where significant changes in coupling peak are seen in all the mutations involving “on-pathway” residues, but for none of the mutations involving off pathway residues.

**Table 2.2** Summary of changes in coupling due to mutations in PDZ domain protein. X indicates a peak at that residue that changed due to that mutation (increased or decreased) by  $>1.5\sigma$  as determined by outlier analysis. Mutants in italics are not coupled to H76. Residue numbering is according to PDB 1BE9.

Peak position	<i>F44A</i>	<i>S75T</i>	<i>T89S</i>	G33A	G34A	F29K	A80V	K84A
24	X	X			X			X
36				X	X	X	X	X
46	X	X					X	
49	X			X	X		X	
54				X				X
75		X	X		X	X	X	X
81	X		X		X		X	
85	X		X				X	X

#### 4. Discussion

We present here the development and initial implementation of a method for analyzing the dynamics of proteins using MD simulations, the PPMD method. The method is straightforward to implement and very flexible. It was designed with the goal of helping to interpret experiments that probe residue–residue coupling, particularly in the context of modern NMR measurements of site resolved protein dynamics, and more

generally in phenomena that involve long range interactions in proteins, such as allosteric effects. The method is not, however, restricted to these applications. The ability to perturb a protein in a defined way at a defined place and analyze the resulting change in dynamics should be applicable to the understanding of proteins in general as complex dynamical objects. The idea of manipulating MD trajectories in a controlled way is not new, and various methods have been developed.(23–25, 68, 69) However, the PPMD method described here has significant differences from these earlier methods.

Temperature pulse methods (68, 70, 71) manipulate the total kinetic energy. In effect they operate over the entire frequency regime simultaneously, and usually are applied over the entire protein. In contrast the PPMD method selectively applies a force at a few frequencies at a time (although these can be scanned over the entire spectrum), and to selected atoms of residues. Steered or targeted MD simulations use a force along some “pathway” applied to some part of the system (72). Here the goal is different from PPMD: it is to drive the system through some specific transition or sequence of events in the MD time scale, events that presumably happen over much longer time scales naturally. Temperature pulse methods are also used with a similar goal, for example to drive protein unfolding on the MD time scale (70, 71).

A method that more closely relates to PPMD is one that manipulates atomic velocity components in a frequency specific way during MD simulation (24, 25). PPMD differs from the methods described in these papers in several key aspects. First, these methods were applied to the atomic velocities. Since the causal sequence in an MD algorithm is Force → Acceleration → Velocity → Position, intervening at the velocity stage will cause atomic velocities and positions to get out of phase. This limits the extent

and magnitude of the pulses one can apply (25). Special precautions such as pulsed applications of the perturbation and quenching can be used to get around this (25), but this complicates the method, and it can only be applied to a single frequency at a time. In contrast, PPMD avoids this by intervening at the force step. The PPMD method is straightforward, and can be applied at multiple frequencies simultaneously. This is an advantage if a wide range of frequencies is to be explored. Second, the goal of the velocity frequency manipulation method was somewhat different: it was applied principally to drive low frequency motions and torsional motions so as to cause some transition (24, 25). Previously Bowman et al. developed a method called driven molecular dynamics (DMD), which also applies harmonic forces to a protein during an MD simulation (26, 27). In DMD a centro-symmetric force of frequency  $\nu$  is applied between each pair of atoms, whose strength is proportional to the distance between the atoms. The goal of applying this “non-specific dilating” force uniformly to all the atoms is to excite normal mode vibrations, which are detected by the amount of energy absorbed at each frequency. The goal of DMD is to provide an alternative to the standard normal mode analysis using Hessian matrix diagonalization (26, 27).

The PPMD method thus differs from previous velocity and force driven MD methods in two respects. First, in the way the frequency-based perturbation is applied—as a force—but only to specific residues. Second, in what the method is designed for: the PPMD method is not designed to drive the protein through some conformational change, low frequency or otherwise, or to excite normal modes, but to examine residue–residue couplings and coupling “pathways.”

Another method that was recently applied to the PDZ domain protein so as to detect differential coupling between different parts of the protein is the anisotropic thermal diffusion (ATD) method of Ota and Agard (69). In this innovative method, the MD simulation is quenched periodically by reducing the simulation temperature to 10 K, then the system is rethermalized with just one residue coupled to a heat bath. The transmission of kinetic energy from this residue to other residues as a function of time is used to extract differences in thermal conductivity. This cycle can be repeated and the results combined to improve the precision. Differences in the rate of thermal diffusion are then interpreted as different couplings between residues. Ota and Agard found that residues H76 and F29 of PDZ domain were coupled, in agreement with the results of Lockless and Raganathan and our analysis. This coupling occurred through the residue I31 that, although not coupled to either H76 or F29 in Lockless and Raganathan's experimental analysis (41), lies directly between H76 and F29, i.e. the coupling occurs by the shortest pathway. In contrast, in the PPMD analysis presented here, the coupling between H76 and F29 occurs along a helix, through residues A80 and K84, which are coupled to H76 in Lockless and Raganathan's experimental analysis, and then across to F29 on a neighboring strand. This is a more indirect path in space. While the aim of the ATD and PPMD methods applied to the PDZ domain protein is the same, they probe different modes of energy transmission under somewhat different conditions, which may account for these differences in the coupling pathway. In the ATD method it is the transmission of kinetic or heat energy, which comes from all types of motion, whether they can be characterized by a frequency or not. In the PPMD method it is transmission of “acoustic” or mechanical motions at specific frequencies. The other difference is that

in the ATD method the protein is quenched and rethermalized, i.e. the method is intermittent and nonequilibrium, while the PPMD method obtains couplings under continuous, equilibrium conditions. The two methods may, therefore, reveal different aspects of coupling, and they could provide complementary information. Further comparison of the two methods on more proteins would be valuable. In this regard, Leitner et al. have done a detailed examination of the thermal conductivity of proteins using normal mode methods (73), although this study was not undertaken from the perspective of examining specific residue–residue couplings.

In the initial applications of PPMD described here we have shown that nontrivial residue–residue couplings can be detected by the method, and that they reveal long range, nontrivial pathways of interaction. In the PDZ domain for example, these coupling pathways recapitulate those seen experimentally. We emphasize although that the genomics analysis, mutation experiments, and simulations are each examining different properties of the protein, the sequence analysis examines the tolerance of different sites to coupled changes in amino acid residues while preserving biological function. The mutation experiments are probing the contribution of different side-chains to peptide binding energetics. The pump-probe simulations measure the transmission of motional energy between residues. In addition, we are using one wild-type sequence alone, while the statistical analysis extracts couplings from many different sequences. Thus, one would not expect exact correspondence between the three lines of analysis. Indeed it is encouraging that the mutation/binding, sequence analysis, and PPMD agree to the extent they do. However, there are differences in detail. In particular, one expects PPMD to reveal more couplings than those seen in experiment, since the simulations provide

atomic level detail on every residue, whereas not all couplings would be important for the correct sequence-to-structure relationship or for binding energetics. An obvious way to more closely mimic the PDZ mutation experiments would be to do pump-probe simulations on PDZ-peptide complexes and analyze binding related changes.

An important consideration for the usefulness of PPMD is distance over which one can detect couplings. The results in Figures 2.2 and 2.3 show that one can pick up an interaction between residues on two secondary structure units separated by a third (a helix). Nevertheless, it would be naïve to expect that one simply pump a residue on one side of a protein and always see an effect on the other side. To map out long pathways one might have to “walk” the pump-probe across the protein in shorter steps. This actually parallels how sequences of specific residue–residue interactions that form pathways have been elucidated in experiments on PDZ domain, GPCR proteins, and CaM (38–41, 43). In spite of limitations in the distance over which one can detect couplings, it is important to point out that the PPMD method detects specific residue–residue couplings that we simply cannot detect in standard residue–residue or atom–atom fluctuation covariance matrix analysis. Such couplings, if present in these matrices, are either swamped by the large amount of noise and statistically spurious correlation coefficients likely in the large number ( $N \times N$ ) of terms, or they are not manifest at a simple pair-wise correlation level. In contrast our analysis effectively reduces the  $N \times N$  dimensionality in the covariance matrix analysis to a series of  $N$ -dimensional analyses, i.e. between each pumped residue and the  $N - 1$  remaining residues, and it also increases signal to noise by pumping, rather than trying to use ambient fluctuations. Of course there is an increased computational cost, as a full series of couplings would require additional

simulations pumping at the other  $N - 1$  residues. However, specific applications are unlikely to require pumping at every residue. Typically only key biologically important residues would need to be pumped to gain useful information.

We examined several of the mutants of PDZ domain protein using PPMD that were studied by Lockless and Raganathan in elucidating the coupling between H76 and F29, namely three “off-pathway” mutations F44A, S75T, and T89S, and five “on-pathway” mutants, G33A, G34A, F29K, A80V, and K84A. The goal of this analysis was to use changes in coupling profile to identify residues that might be involved in the H76-F29 coupling pathway that are not apparent from the sequence analysis/mutation experiments. The results in Table 2.2 show, for example, that at residue 36 significant changes in the coupling are seen in all the mutations involving on-pathway residues, but for none of the mutations involving off-pathway residues. Interestingly, residue D36 sits on the loop directly above residue H76 (as seen at the top of Fig. 2.4). We tentatively identify this residue as having a secondary role in the pathway. We speculate that it may play a role in “directing” the coupling down the helix from H76 through A80 and K84, rather than in the other strand direction from H76. No other peak shows such a clear distinction between on- and off-pathway mutations, although the peak at residue 49 is prominently perturbed by just one off-pathway mutations, but by the majority of on-pathway mutations. The peak at position 46 is effectively the converse, being perturbed in two of three off-pathway mutations, while being affected by just one on-pathway mutation. This presumably indicates a residue that is quite decoupled from the pathway. It is significant that one should see some such decoupled residues, since if every residue in the protein were involved in a coupling it would make little sense to talk about a

coupling pathway at all. Our mutant analysis is far from exhaustive, and clearly one would not expect a clean binary distinction between “on” and “off” pathway cases for every position/mutation, given the complexity of the dynamics of PDZ and the various experimental and simulation uncertainties and limitations. However, this analysis does indicate how one can use PPMD to provide a more mechanistic view of experimentally determined couplings, and how one might use this analysis of MD simulations to examine residue and atomic level details of coupling inaccessible to experiment.

Characterization of CaM dynamics by the time autocorrelation functions revealed that it behaves as an under-damped system, i.e. it has a significant oscillatory characteristic. Other proteins we have examined show the same general behavior. This is significant for the application of the PPMD method. It means that oscillations, whether intrinsic or applied externally as here, will propagate significantly through the protein, as seen here. These oscillations are thus in principle capable of mediating long range coupling, of being detected and analyzed. If a protein acted as an over-damped system, one would expect much less transmission of the PPMD pulse, making the method less sensitive. Indeed, the consequences of a protein being an over-damped dynamic system would be problematic for any frequency-based analysis such as mode analysis or frequency filtering. If protein dynamics had no oscillatory character, it would make little sense to talk about modes or frequencies at all. The oscillatory character of protein dynamics, while clear in the time correlation analysis, is not obvious from the atomic fluctuation power spectra. In a non-PPMD simulation the fluctuation power spectrum is rather featureless with a low, broad peak extending from about 0.02–0.08 ps<sup>-1</sup> (Fig. 2.1, top trace). Examination of such profiles for different atoms, or for the same atom from

different simulation batches or conditions shows that these profiles are generally similar. No clearly identifiable modes both stand out from their neighbors and are persistent across a group of atoms or between different simulation batches. Put another way, it is difficult from examination of the fluctuation power spectrum alone to determine whether protein dynamics is occurring in either the under-damped or over-damped regime. Moreover, the density of vibrational states determined both experimentally, and in harmonic and quasi-harmonic analyses of proteins is also dense enough to approximate a continuum.(74, 75) The difficulty in identifying specific and persistent low frequency modes from principal component analysis (13–16) is partly due to how closely spaced in frequency they are, and how variable many modes are. So selection of particular native modes from harmonic analysis so as to analyze coupling or putative allosteric conformational changes is difficult and somewhat subjective. As demonstrated here, the PPMD provides another approach to this problem, where one specifically probes the coupling between residues at a wide range of frequencies. The couplings obtained from PPMD show physically reasonable properties, such as reciprocity, decay with distance, and transmission times consistent with known physical behavior of proteins. The couplings are also robust with respect to the presence of solvent friction and mild restraints on the protein. This is useful since it allows one to use larger pumping forces and get larger signal to noise ratios in the PPMD simulations.

## 5. Future Directions

The PPMD method is very flexible in terms of how and where the pumping force is applied. We chose only quite simple pumping schemes at a restricted set of frequencies so as to demonstrate the usefulness of the method. We pumped at periods from 1 to 70 ps. The full range, conservatively, would be from about 5 times the MD time step (5 fs) to about one-fifth of the total simulation length (anywhere from 1 ns upwards with current computers). In addition to the pumping frequency, the selection of pumped atoms, magnitude, type, and phase of force can be varied widely. These features of PPMD can be exploited to maximize control and sensitivity. As PPMD is a new technique, these options need to be explored in a systematic fashion so as to get the most out of it. Some specific considerations include:

- How finely to scan the pump frequency: scanning requires multiple simulations, and so one wants to be effective but frugal. Pumping at four frequencies yields a proportional speed up in this regard. The practical upper limit to the number of simultaneous pumping frequencies is not known.
- The largest magnitude pumping force that can be applied without causing too much heating or structural distortion so as to maximize sensitivity.
- How best to apply the force: options include applying it to a single atom or a collection of atoms, as a planar or circular oscillation, as torsional force around a bond, etc. To increase sensitivity one can also pump more atoms, such as parts of/whole secondary structure units, reflecting the tendency of longer time-scale

motions to involve more atoms. The pumping of a portion of Helix C of CaM illustrates this.

PPMD analysis has a close correspondence to NMR experiments that measure propagation of changes in order parameters ( $S^2$ ) of residues surrounding a site of single residue mutations (42, 44, 45). Both provide a measure of residue–residue coupling through dynamics on the same spatial and time scales. The experiments also reveal inhomogeneous transmission of effects. For example in CaM the M119L mutant produces large changes that are confined to the region of the mutation, whereas D58N produces an apparent pathway of changes in methyl order parameters up to 15–20 Å away (45). While order parameters can be obtained directly from MD simulations and compared to NMR data (29), both experimental and calculated quantities are single residue values: there is no specific information about correlation between sites. Here the explicit coupling information obtained from PPMD provides an important extra dimension. Analysis of such NMR experiments is a natural future application for PPMD. More generally, we envision PPMD as a flexible method for analyzing protein dynamics with different, but complementary properties to existing analysis techniques, not as a replacement. It is in fact possible to use PPMD in combination with existing techniques. For example, normal mode and quasi-harmonic analysis provide eigen-values and eigen-vectors that describe the frequency and identity of effective modes. In PPMD it would be possible to pump certain combinations of atoms with the right magnitudes and directions to mimic a mode, thus combining elements of both approaches. Either alone or in combination with other techniques, the PPMD method can help address specific questions about protein dynamics such as how much one residue is coupled to another,

how a perturbation at one site affects another site, and whether there is a pathway of coupling between the two sites.

## 6. Bibliography

1. Sharp, K., and J.J. Skinner. 2006. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins*. 65: 347-361.
2. Englander, S.W., and N.R. Kallenbach. 1983. Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids. *Quarterly Reviews of Biophysics*. 16: 521-655.
3. Palmer, A.G., 3rd, C.D. Kroenke, and J.P. Loria. 2001. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Meth. Enzymol*. 339: 204-238.
4. Gō, N., and H. Scheraga. 1969. Analysis of the Contribution of Internal Vibrations to the Statistical Weights of Equilibrium Conformations of Macromolecules. *The Journal of Chemical Physics*. 51: 4751.
5. Case, D.A. 1994. Normal mode analysis of protein dynamics. *Current Opinion in Structural Biology*. 4: 285-290.
6. Levy, R.M., A.R. Srinivasan, W.K. Olson, and J.A. McCammon. 1984. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers*. 23: 1099-1112.

7. Levy, R.M., M. Karplus, J. Kushick, and D. Perahia. 1984. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an  $\alpha$ -helix. *Macromolecules*. 17: 1370-1374.
8. McCammon, J.A., B.R. Gelin, M. Karplus, and P.G. Wolynes. 1976. The hinge-bending mode in lysozyme. *Nature*. 262: 325-326.
9. Karplus, M., and J.N. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules*. 14: 325-332.
10. Ichiye, T., and M. Karplus. 1991. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Bioinformatics*. 11: 205-217.
11. Hayward, S., A. Kitao, and N. Gō. 1995. Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis. *Proteins: Structure, Function, and Bioinformatics*. 23: 177-186.
12. Amadei, A., A.B.M. Linssen, and H.J.C. Berendsen. 1993. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*. 17: 412-425.
13. Hünenberger, P.H., A.E. Mark, and W.F. van Gunsteren. 1995. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* 252: 492-503.

14. Van Aalten, D.M.F., B.L. De Groot, J.B.C. Findlay, H.J.C. Berendsen, and A. Amadei. 1997. A comparison of techniques for calculating protein essential dynamics. *Journal of Computational Chemistry*. 18: 169-181.
15. Abseher, R., and M. Nilges. 1998. Are there non-trivial dynamic cross-correlations in proteins? *J. Mol. Biol.* 279: 911-920.
16. Balsera, M.A., W. Wriggers, Y. Oono, and K. Schulten. 1996. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* 100: 2567-2572.
17. Prabhu, N.V., A.L. Lee, A.J. Wand, and K.A. Sharp. 2003. Dynamics and Entropy of a Calmodulin–Peptide Complex Studied by NMR and Molecular Dynamics †. *Biochemistry*. 42: 562-570.
18. Tirion. 1996. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* 77: 1905-1908.
19. Kim, M.K., G.S. Chirikjian, and R.L. Jernigan. 2002. Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.* 21: 151-160.
20. Miyashita, O., J.N. Onuchic, and P.G. Wolynes. 2003. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 100: 12570-12575.
21. Zheng, W., B.R. Brooks, S. Doniach, and D. Thirumalai. 2005. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure*. 13: 565-577.

22. Sessions, R.B., P. Dauber-Osguthorpe, and D.J. Osguthorpe. 1989. Filtering molecular dynamics trajectories to reveal low-frequency collective motions: Phospholipase A2. *Journal of Molecular Biology*. 210: 617-633.
23. Levitt, M. 1991. Real-time interactive frequency filtering of molecular dynamics trajectories. *J. Mol. Biol.* 220: 1-4.
24. Dauber-Osguthorpe, P., C.M. Maunder, and D.J. Osguthorpe. 1996. Molecular dynamics: deciphering the data. *J. Comput. Aided Mol. Des.* 10: 177-185.
25. Phillips, S.C., J.W. Essex, and C.M. Edge. 2000. Digitally filtered molecular dynamics: The frequency specific control of molecular dynamics simulations. *The Journal of Chemical Physics*. 112: 2586.
26. Bowman, J.M., X. Zhang, and A. Brown. 2003. Normal-mode analysis without the Hessian: A driven molecular-dynamics approach. *The Journal of Chemical Physics*. 119: 646.
27. Kaledin, M., A. Brown, A.L. Kaledin, and J.M. Bowman. 2004. Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules. *J Chem Phys*. 121: 5646-5653.
28. Akke, M., R. Brueschweiler, and A.G. Palmer. 1993. NMR order parameters and free energy: an analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k. *J. Am. Chem. Soc.* 115: 9832-9833.

29. Chatfield, D.C., A. Szabo, and B.R. Brooks. 1998. Molecular Dynamics of Staphylococcal Nuclease: Comparison of Simulation with  $^{15}\text{N}$  and  $^{13}\text{C}$  NMR Relaxation Data. *J. Am. Chem. Soc.* 120: 5301-5311.
30. Wong, K.-B., and V. Daggett. 1998. Barstar Has a Highly Dynamic Hydrophobic Core: Evidence from Molecular Dynamics Simulations and Nuclear Magnetic Resonance Relaxation Data†. *Biochemistry.* 37: 11182-11192.
31. Zabell, A.P.R., and C.B. Post. 2002. Docking multiple conformations of a flexible ligand into a protein binding site using NMR restraints. *Proteins: Structure, Function, and Bioinformatics.* 46: 295-307.
32. Ma, J., and M. Karplus. 1998. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 95: 8502-8507.
33. Cooper, A., and D.T. Dryden. 1984. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* 11: 103-109.
34. Perutz, M.F. 1989. Mechanisms of cooperativity and allosteric regulation in proteins. *Q. Rev. Biophys.* 22: 139-237.
35. Perutz, M.F., G. Fermi, B. Luisi, B. Shaanan, and R.C. Liddington. 1987. Stereochemistry of cooperative mechanisms in hemoglobin. *Acc. Chem. Res.* 20: 309-321.

36. Freire, E. 1999. The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 96: 10118-10122.
37. Sun, H., D. Yin, L.A. Coffeen, M.A. Shea, and T.C. Squier. 2001. Mutation of Tyr138 disrupts the structural coupling between the opposing domains in vertebrate calmodulin. *Biochemistry.* 40: 9605-9617.
38. Jaren, O.R., J.K. Kranz, B.R. Sorensen, A.J. Wand, and M.A. Shea. 2002. Calcium-Induced Conformational Switching of Paramecium Calmodulin Provides Evidence for Domain Coupling †. *Biochemistry.* 41: 14158-14166.
39. Sorensen, B.R., L.A. Faga, R. Hultman, and M.A. Shea. 2002. An interdomain linker increases the thermostability and decreases the calcium affinity of the calmodulin N-domain. *Biochemistry.* 41: 15-20.
40. Faga, L.A., B.R. Sorensen, W.S. VanScyoc, and M.A. Shea. 2003. Basic interdomain boundary residues in calmodulin decrease calcium affinity of sites I and II by stabilizing helix–helix interactions. *Proteins: Structure, Function, and Bioinformatics.* 50: 381-391.
41. Lockless, S.W., and R. Ranganathan. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science.* 286: 295-299.
42. Fuentes, E.J., C.J. Der, and A.L. Lee. 2004. Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J. Mol. Biol.* 335: 1105-1115.

43. Süel, G.M., S.W. Lockless, M.A. Wall, and R. Ranganathan. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10: 59-69.
44. Clarkson, M.W., and A.L. Lee. 2004. Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c. *Biochemistry.* 43: 12448-12458.
45. Igumenova, T.I., A.L. Lee, and A.J. Wand. 2005. Backbone and side chain dynamics of mutant calmodulin-peptide complexes. *Biochemistry.* 44: 12627-12639.
46. Lee, A.L., S.A. Kinnear, and A.J. Wand. 2000. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat. Struct. Biol.* 7: 72-77.
47. Lee, A.L., and A.J. Wand. 2001. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature.* 411: 501-504.
48. Lee, A.L., K.A. Sharp, J.K. Kranz, X.-J. Song, and A.J. Wand. 2002. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry.* 41: 13814-13825.
49. Mayer, K.L., M.R. Earley, S. Gupta, K. Pichumani, L. Regan, and M.J. Stone. 2003. Covariation of backbone motion throughout a small protein domain. *Nat. Struct. Biol.* 10: 962-965.

50. Lange, O.F., H. Grubmüller, and B.L. de Groot. 2005. Molecular Dynamics Simulations of Protein G Challenge NMR-Derived Correlated Backbone Motions. *Angewandte Chemie International Edition*. 44: 3394-3399.
51. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*. 4: 187-217.
52. Press, W., B. Flannery, S. Teukolsky, and W. Vetterling. 1986. *Numerical recipes*. Cambridge: Cambridge University Press.
53. van Kampen, N. 1992. *Stochastic processes in physics and chemistry*. Amsterdam: Elsevier.
54. MacKerell, A., B. Brooks, C.L. Brooks, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998. CHARMM: the energy function and its parameterization with an overview of the program. In: Schleyer P, editor. *The encyclopedia of computational chemistry*. Chichester: Wiley. pp. 271-277.
55. Sridharan, S., A. Nicholls, and K.A. Sharp. 1995. A rapid method for calculating derivatives of solvent accessible surface areas of molecules. *Journal of Computational Chemistry*. 16: 1038-1044.
56. Ryckaert, J.-P., G. Ciccotti, and H.J.. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*. 23: 327-341.

57. Meador, W., A. Means, and F. Quijano. 1992. Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science*. 257: 1251-1255.
58. Yap, K., J. Kim, K. Truong, M. Sherman, T. Yuan, and M. Ikura. 2000. Calmodulin Target Database. *Journal of Structural and Functional Genomics*. 1: 8-14.
59. Babu, Y., C. Bugg, and W. Cook. 1988. Structure of calmodulin refined at 2.2 Å resolution. *Journal of Molecular Biology*. 204: 191-204.
60. Pedigo, S., and M.A. Shea. 1995. Discontinuous Equilibrium Titrations of Cooperative Calcium Binding to Calmodulin Monitored by 1-D <sup>1</sup>H-Nuclear Magnetic Resonance Spectroscopy. *Biochemistry*. 34: 10676-10689.
61. Wintrobe, P.L., and P.L. Privalov. 1997. Energetics of target peptide recognition by calmodulin: a calorimetric study. *J. Mol. Biol.* 266: 1050-1062.
62. Kranz, J.K., P.F. Flynn, E.J. Fuentes, and A.J. Wand. 2002. Dissection of the Pathway of Molecular Recognition by Calmodulin †. *Biochemistry*. 41: 2599-2608.
63. VanScyoc, W.S., and M.A. Shea. 2001. Phenylalanine fluorescence studies of calcium binding to N-domain fragments of Paramecium calmodulin mutants show increased calcium affinity correlates with increased disorder. *Protein Science*. 10: 1758-1768.

64. Doyle, D. 1996. Crystal Structures of a Complexed and Peptide-Free Membrane Protein? Binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell*. 85: 1067-1076.
65. Songyang, Z., A.S. Fanning, C. Fu, J. Xu, S.M. Marfatia, A.H. Chishti, A. Crompton, A.C. Chan, J.M. Anderson, and L.C. Cantley. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*. 275: 73-77.
66. Dadarlat, V.M., and C.B. Post. 2003. Adhesive-cohesive model for protein compressibility: an alternative perspective on stability. *Proc. Natl. Acad. Sci. U.S.A.* 100: 14778-14783.
67. 1990. CRC handbook of chemistry and physics. 71st ed. Boca Raton: CRC.
68. Becker, and Karplus. 1993. Temperature echoes in molecular dynamics simulations of proteins. *Phys. Rev. Lett.* 70: 3514-3517.
69. Ota, N., and D.A. Agard. 2005. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.* 351: 345-354.
70. Paci, E., and M. Karplus. 2000. Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc. Natl. Acad. Sci. U.S.A.* 97: 6521-6526.
71. Snow, C.D., L. Qiu, D. Du, F. Gai, S.J. Hagen, and V.S. Pande. 2004. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 101: 4077-4082.

72. Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75: 662-671.
73. Yu, X., and D.M. Leitner. 2005. Heat flow in proteins: Computation of thermal transport coefficients. *The Journal of Chemical Physics.* 122: 054902.
74. Cusack, S., J. Smith, J. Finney, B. Tidor, and M. Karplus. 1988. Inelastic neutron scattering analysis of picosecond internal protein dynamics. Comparison of harmonic theory with experiment. *J. Mol. Biol.* 202: 903-908.
75. Balog, E., T. Becker, M. Oettl, R. Lechner, R. Daniel, J. Finney, and J. Smith. 2004. Direct Determination of Vibrational Density of States Change on Ligand Binding to a Protein. *Physical Review Letters.* 93.

## CHAPTER 3

### Enhancing the Accuracy of Hydrogen Exchange Rates

#### Determined by $^1\text{H}$ - $^1\text{H}$ Measurements

##### 1. Introduction

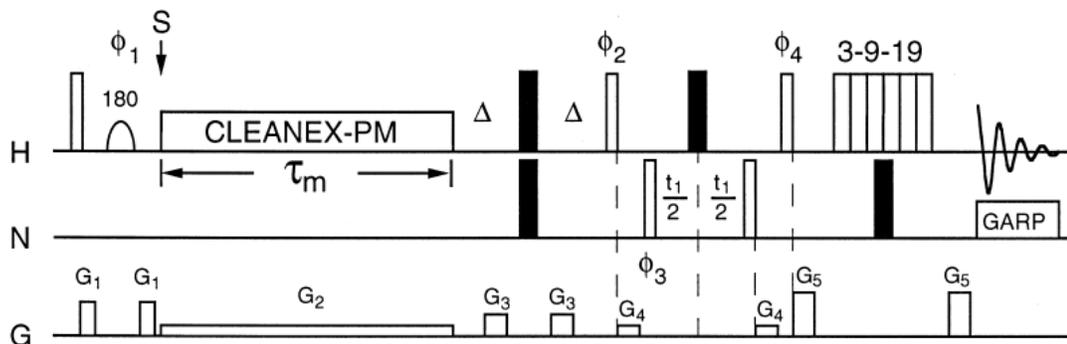
Protein hydrogen exchange (HX) rates span more than a billion-fold range (1). HX rates over most of this range can be measured by mixing protein with  $^2\text{H}_2\text{O}$  and measuring HX either by NMR (2) or by mass spectroscopy (3). However, this method is incapable of measuring rates on very short time scales as the solvent mixing step becomes a limiting factor. In order to overcome this limitation, previous workers have developed methods for measuring  $^1\text{H}$ - $^1\text{H}$  exchange by NMR by observing the transfer of magnetization from water to the protein (4, 5). This has allowed for HX rates to be measured on the order of  $1\text{-}100\text{ s}^{-1}$ , which corresponds to the HX rates for unprotected and minimally protected amides at physiological pH.

One significant downside of this technique is that magnetization transfers to the amide by multiple pathways. Advances in pulse sequence development have minimized the contributions of these alternative pathways, but have not eliminated them entirely (6). These contributions manifest themselves in a predictable manner, yet the complexity of the formula that describes their behavior has prevented researchers from properly accounting for them when calculating HX rates. Here is described a novel algorithm for accurately fitting magnetization transfer data. This algorithm has been applied to data collected by Woon Lim and the accuracy of the resulting rates was verified in several

ways. This fitting algorithm results in determining rates that are systematically faster than those determined by the previously utilized method.

## 2. Theory - Magnetization Transfer

Magnetization transfer experiments are performed by applying a series of pulses that selectively excite the solvent and then monitoring the  $^1\text{H}$ - $^{15}\text{N}$  signal as magnetization transfers to the protein. Magnetization transfer primarily occurs by three pathways: 1. chemical exchange, 2. NOE transfer from  $\text{C}_\alpha\text{H}$ 's that were magnetized along with solvent, and 3. NOE transfer from exchange-relay that results from rapidly exchanging protons, such as those associated with hydroxyl or amine groups. The CLEANEX-PM pulse module largely eliminates NOE contributions by manipulating magnetization trajectories, which forces spins to spend twice as much time along the z-axis than in the xy-plane, causing NOE and ROE contributions to cancel one another (5, 6). This spin-locking sequence is applied during the HX mixing time and the magnetization transfer is observed by fast HSQC (Fig. 3.1).



**Figure 3.1:** CLEANEX-PM pulse sequence adapted from (6).

The signal arising from the CLEANEX-PM pulse sequence is a 2D  $^1\text{H}$ - $^{15}\text{N}$  spectra in which each peak has a normalized volume ( $V_\tau/V_0$ ) that has been described for a two-spin system (Eq. 3.1; (7, 8))

$$\frac{V_\tau}{V_0} = \frac{fk_{ex}}{k_{ex}+R_{1A}-R_{1B}} \left( e^{-R_{1B}\tau} - e^{-(k_{ex}+R_{1A})\tau} \right) \quad \text{Eq. 3.1}$$

The signal arises from the chemical exchange rate ( $k_{ex}$ ) and from the amide relaxation rate  $R_{1A}$ , which is a combination of longitudinal and transverse relaxation rates (6). The signal is also affected by the fraction of water that recovers between experiments (f) and the water relaxation rate ( $R_{1B}$ ).

Due to the complex and non-linear nature of Eq. 3.1, other groups have avoided fitting to the full equation and have instead approximated  $k_{ex}$  from initial slope based on the fact that when  $\tau = 0$ ,  $d(V_\tau/V_0)/d\tau = f k_{ex}$  (6). However, because the slope deviates from this behavior when  $\tau > 0$ , this fitting method results in systematic errors.

### 3. Methods

#### 3.1 Data Collection

All magnetization transfer data were collected by Dr. Woon Lim at 20 °C using a 500 MHz magnet with Varian cold probe.  $^1\text{H}$  -  $^1\text{H}$  exchange was measured using the CLEANEX-PM pulse sequence (5, 6) with mixing times of 4, 5, 6, 7, 8, 10, 15, and 20 ms over the pH range 4.9-11.26 at approximately half pH increments.

Values for  $f$  (the fraction of H<sub>2</sub>O that recovers between pulses) and  $R_{1B}$  (a residue and pH-independent solvent relaxation rate) were determined experimentally at these conditions to be 0.54 and 0.6 s<sup>-1</sup>, respectively.

### 3.2 Accurate Fitting Algorithm

The rates  $k_{ex}$  and  $R_{1A}$  were first approximated for each residue at each pH by setting the starting value for  $k_{ex}$  equal to the initial slope divided by  $f$ . In the second round,  $k_{ex}$  was fit for each residue at each pH using Eq. 3.1 and setting  $R_{1A} = 0$ . Fitting was then repeated allowing both  $k_{ex}$  and  $R_{1A}$  to float and using  $k_{ex}$  from the previous round as the starting value and 50 s<sup>-1</sup> as the starting  $R_{1A}$  value.  $k_{ex}$  and  $R_{1A}$  were then fit for each residue at all pH's simultaneously based on the knowledge that  $k_{ex}$  increases by 10-fold per pH unit and  $R_{1A}$  is independent of pH. This provides the final  $R_{1a}$  rate for each residue and the starting  $k_{ex}$  rates for a final round of fitting, done for each residue at each pH. All fitting was done using the Matlab v2008a function `lsqnonlin`. An example is described in detail in section 4.1.

### 3.3 Removal of Background Noise

Because eq. 3.1 describes a two spin system it fails to account for magnetization signals arising from cross relaxation. It is therefore necessary to remove the cross relations contribution from the measured signal prior to fitting. At low pH when  $k_{ex} < 1$  s<sup>-1</sup>, a low level pH-independent background signal was observed for most amides. This

behavior is not accounted for in Eq. 3.1. This signal, when it could be measured for each amide at low pH, or the global average when it could not, was subtracted from all pH-dependent time points prior to rate determination.

After background noise was subtracted an additional automated filtering routine was applied to remove poor quality data. Measurements for a given residue at a single pH were discarded if any of the volumes were negative, if the maximum volume was greater than  $f$  times the reference volume, or if the best fit line through the data had a negative slope.

## **4. Results**

### **4.1 Stepwise Fitting**

A multi-step fitting algorithm was applied to CLEANEX-PM data in order to obtain accurate  $k_{ex}$  rates. The stepwise fitting process is illustrated for Gly86 in Fig. 3.2-3.8.

The first step was to remove the background signal. This results in a significant signal change at low pH but very little change at high pH as can be seen by comparing Fig. 3.2 to Fig. 3.3. In this case the pH 7.47 and pH 7.95 time courses were used to determine the background signal and therefore do not appear in Fig. 3.3.

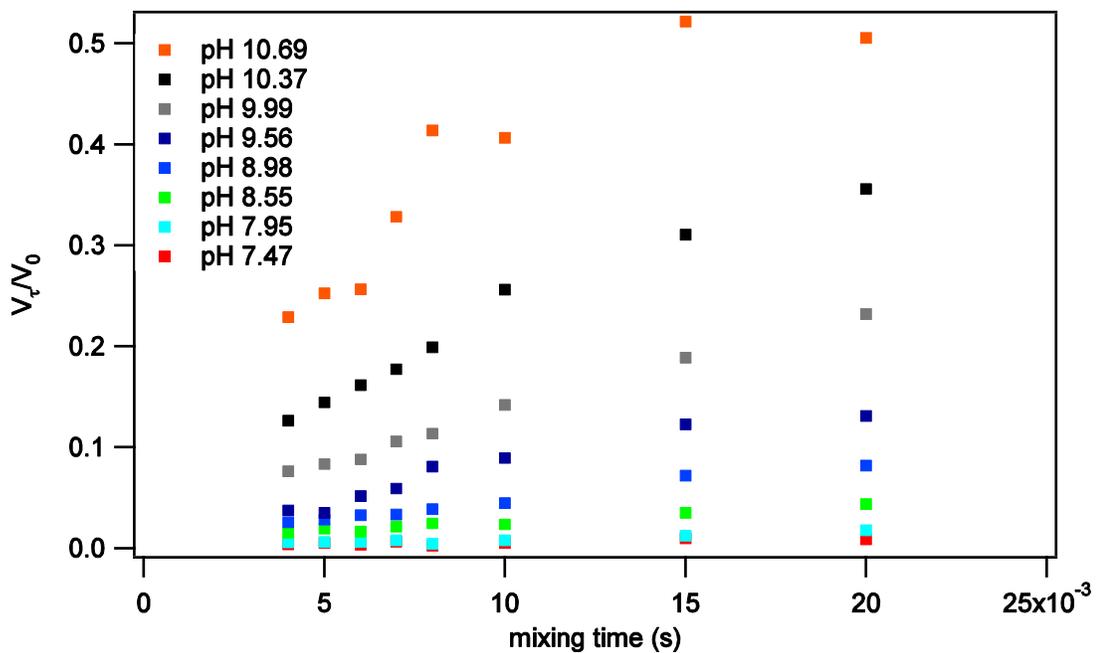


Figure 3.2 Uncorrected peak volumes as a function of mixing time for Gly86.

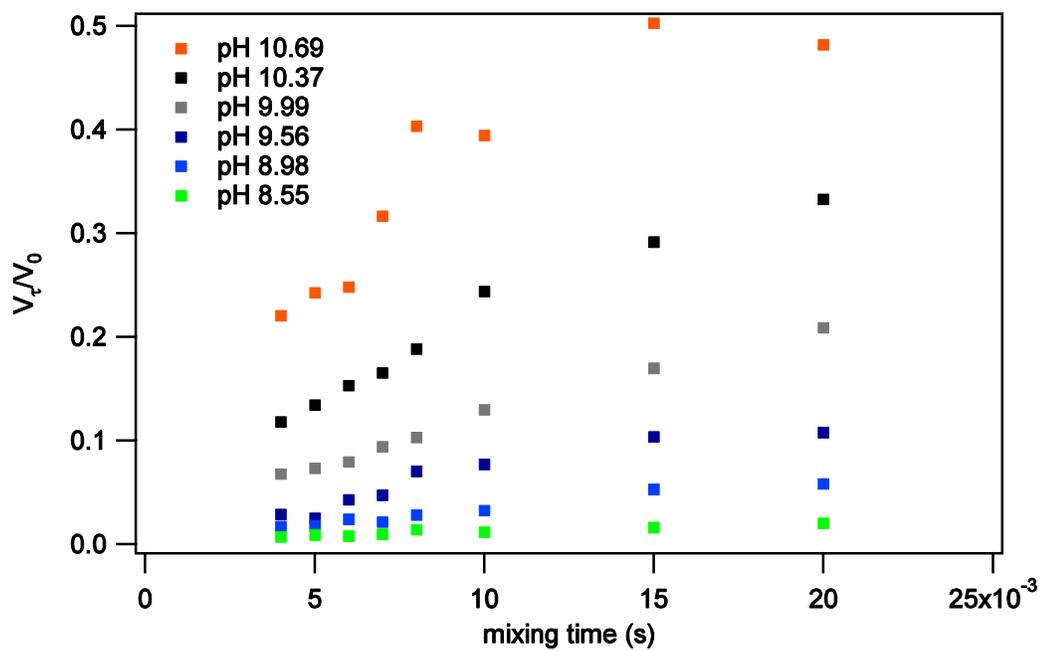
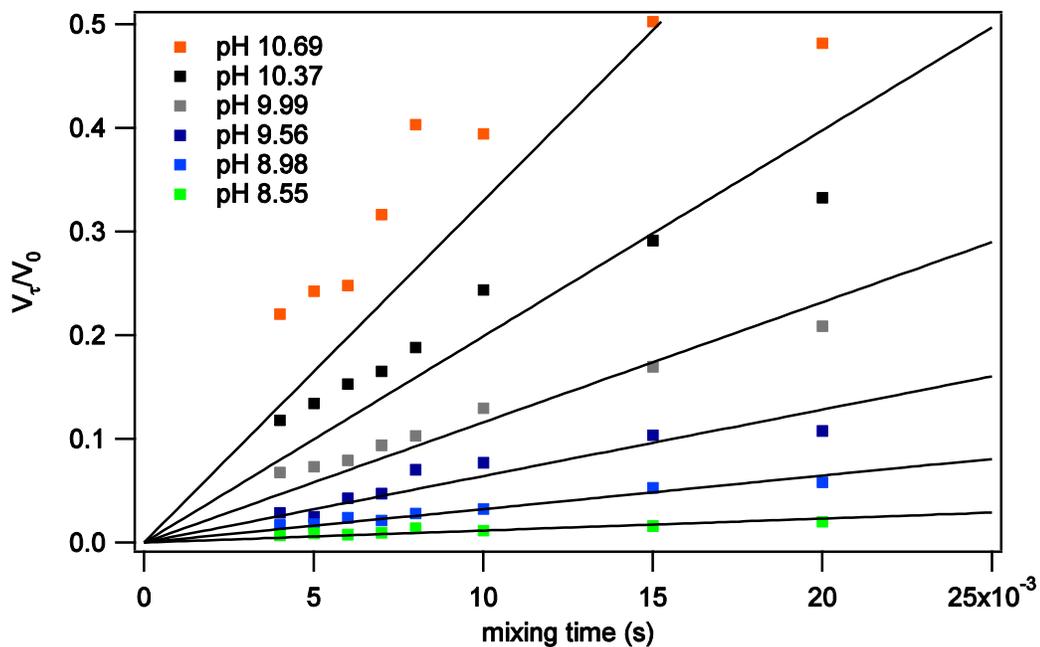


Figure 3.3 Peaks volumes as a function of mixing time for Gly86 after background signal has been subtracted.

The first step involved fitting a line to each time course (Fig. 3.4). The initial  $k_{ex}$  rate was calculated based on the assumption that  $f k_{ex}$  equals the initial slope. The best fit line for the entire time course was used rather than the initial slope because choosing which time points to categorize as initial slope requires an arbitrary cutoff.



**Figure 3.4** Linear fits to time courses for Gly86. The slope is assumed to equal  $f k_{ex}$ .

Eq. 3.1 was used for the next round of fitting. In this step  $R_{1A}$  was set to 0 and the fitting function searched for  $k_{ex}$  only (Fig. 3.5). Even without including  $R_{1A}$ , the resulting fits are superior to linear fits. This fitting step was then repeated, this time seeking the best fit values for both  $k_{ex}$  and  $R_{1A}$  (Fig. 3.6). For this step the initial  $R_{1A}$  value was set to  $50 \text{ s}^{-1}$  so that the starting point would not be identical to the end point of the previous fitting round, which could be a local minimum.

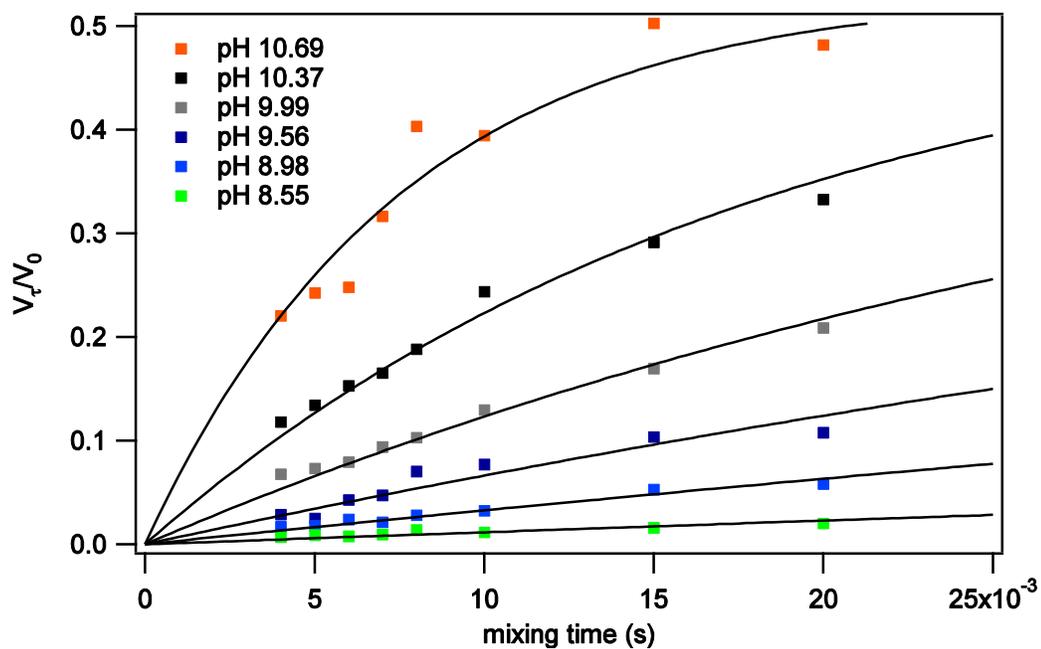


Figure 3.5 Best fit lines using Eq. 3.1 with the approximation  $R_{1A} = 0$ .

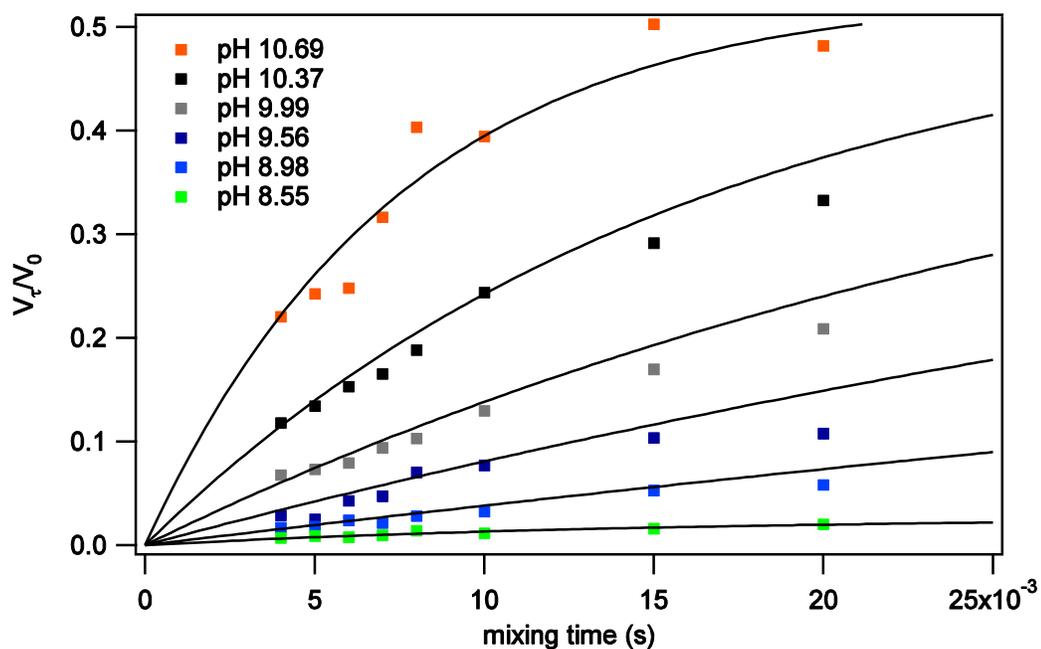
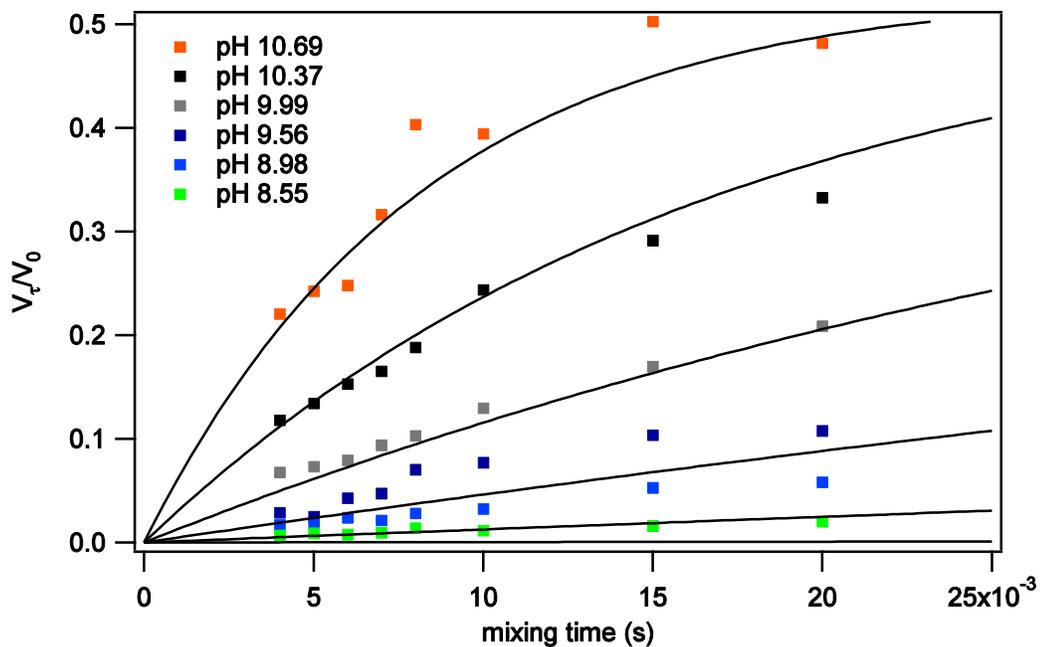


Figure 3.6 Best fit lines using Eq. 3.1 to fit both  $k_{ex}$  and  $R_{1A}$

The next round involved fitting the time courses for all pH's simultaneously by setting  $k_{ex} = k_0 + \text{pH}$ . This was based on the knowledge that  $k_{ex}$  increases by a factor of

10 per pH unit (9), as described in Chapter 1. Starting values for  $k_0$  and  $R_{1A}$  were set to the mean values calculated in the previous round. For Gly86, this reduces the number of fitting parameters from 12 to 2, which results in poorer fits for some time courses (Fig. 3.7).



**Figure 3.7** Fit for all pH's simultaneously

For the final fitting round the time courses for each pH were once again fit separately. The same  $R_{1A}$  value was used for each pH and was kept constant while fitting (Fig. 3.8). The  $k_{ex}$  rates from the final round of fitting were used to calculate pH-independent  $k_0$  rates (Table 3.1).

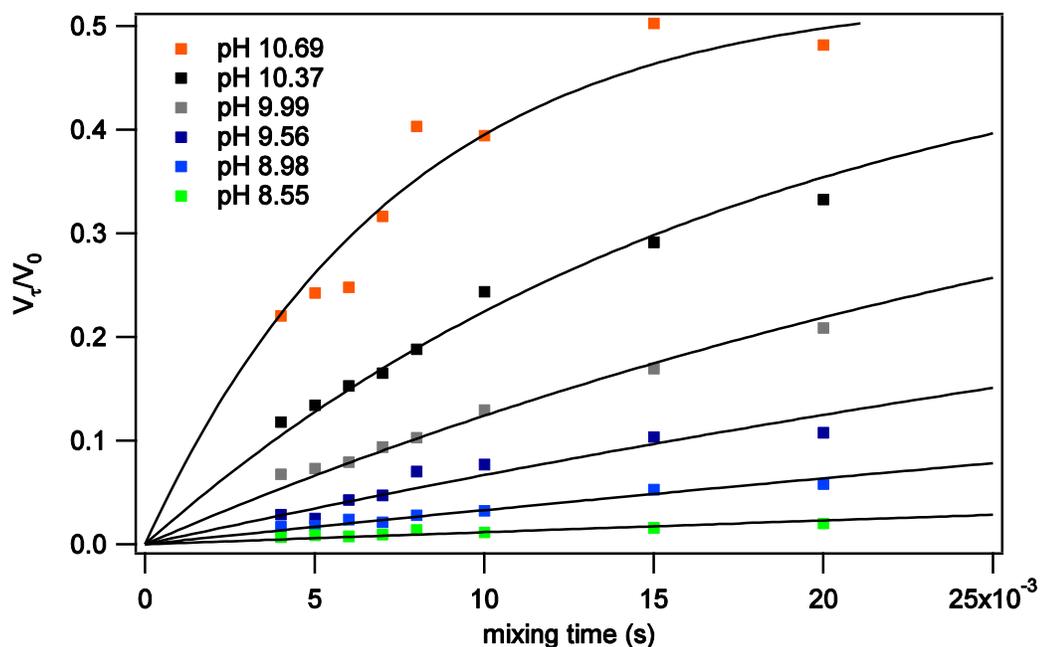


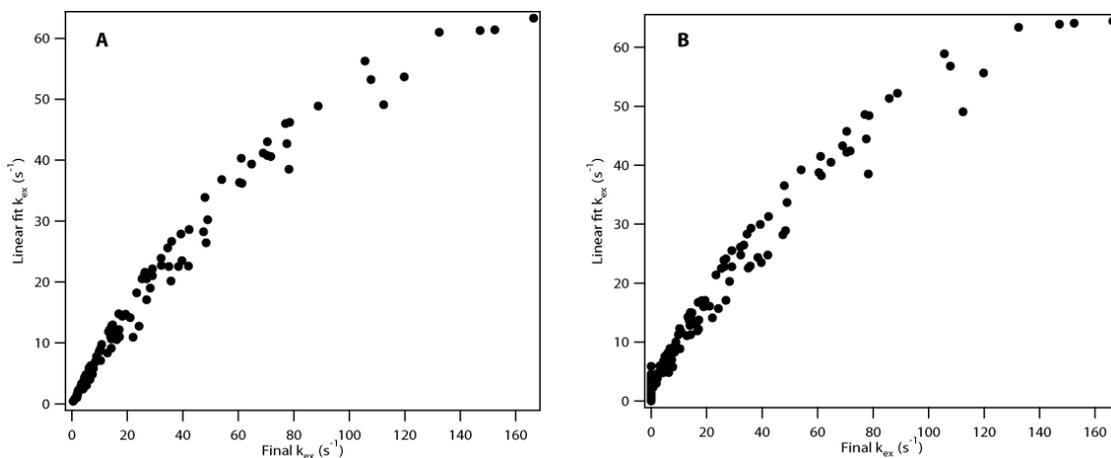
Figure 3.8 Final pH dependent fitting step

Table 3.1 pH-independent  $\log k_0$  rates in  $s^{-1}$

Res	$\log k_0$						
2	10.7	21	4.0	76	3.8	138	5.7
5	7.7	28	7.5	80	8.2	144	7.8
7	7.2	29	7.2	82	5.5	145	8.2
8	7.5	33	7.3	83	4.2	146	8.2
9	7.2	58	5.9	86	5.6	147	8.4
10	4.6	59	5.3	96	6.8	148	8.3
14	6.3	68	6.0	119	7.1	149	6.8
17	7.1	69	4.4	120	5.7		
19	3.7	72	5.8	123	7.7		

On average these rates were 1.4 fold faster than the rates determined by linear fitting alone. A simple average difference between rates determined by our fitting algorithm and simple linear fitting fails to capture the significance of this advancement. Figure 3.9A illustrates that the deviation in rates is non-linear. As  $k_{ex}$  increases the linear fitting underestimates exchange rates even more. Failure to remove background signal

prior to linear fitting results in a similar overall systematic deviation, though with a more pronounced deviation when  $k_{ex}$  is small (Fig. 3.9B).

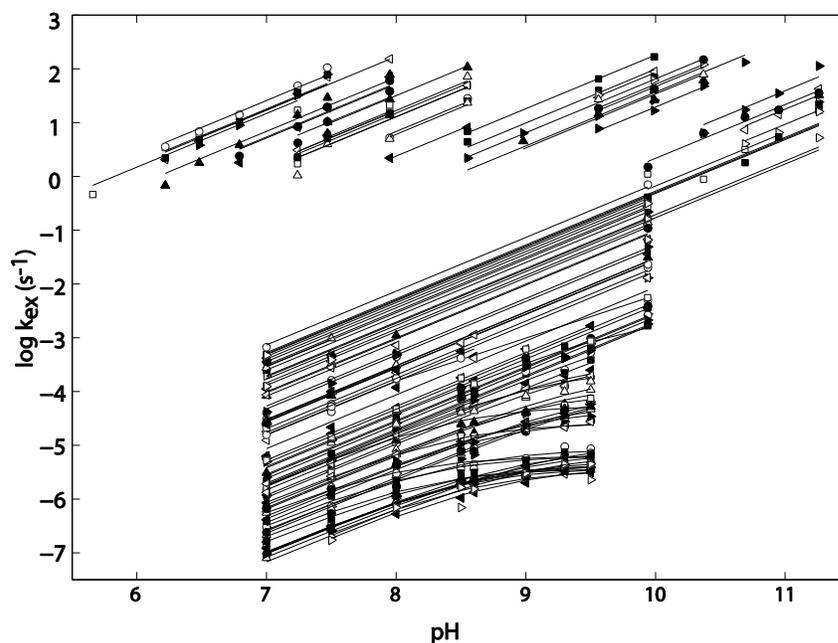


**Figure 3.9**  $k_{ex}$  determined from linear slope vs  $k_{ex}$  determined by the method described here with (A) and without (B) removing background noise.

## 4.2 Rate Validation

### 4.2.1 pH Dependence

At physiological pH HX is catalyzed by hydroxide (9), therefore HX rates are expected to increase by a factor of 10 per pH unit. This behavior can be seen in fig. 3.10 for  $^1H$ - $^1H$  rates determined using the algorithm described above. The systematic deviation that results from standard linear fitting (Fig. 3.9) prevents rates determined by that method from exhibiting expected pH dependence.



**Figure 3.10**  $\log k_{\text{ex}}$  as a function of pH with unity slope. Rates faster than  $0.5\text{s}^{-1}$  were collected by  $^1\text{H}$ - $^1\text{H}$  exchange. Slower rates were collected by  $^2\text{H}$ - $^1\text{H}$  exchange are discussed in Chapter 4.

#### 4.2.2 Comparison to $^2\text{H}$ - $^1\text{H}$ Rates

Seven residues could be measured by both  $^1\text{H}$ - $^1\text{H}$  and  $^2\text{H}$ - $^1\text{H}$  exchange, namely the slowest amides measured by  $^1\text{H}$ - $^1\text{H}$  at high pH and the fastest amides measured by  $^2\text{H}$ - $^1\text{H}$  at low pH. These rates can be compared by applying the expected pH dependence (10) and isotope effects (11). Rates for six of the seven sites agree to within a factor of 4. The exception, Phe61 (much faster), could only be measured by  $^1\text{H}$ - $^1\text{H}$  at pH 11.26 where rates are questionable due to protein destabilization.

### 4.2.3 Comparison to Theoretical Rates

Previous calibrations for main chain amides in unstructured polypeptides make it possible to predict HX rates based on primary amino acid sequence and ambient conditions (10, 11). For seven of the eight amides measured on unstructured SNase segments (residues 2 and 5 at the N-terminus and 144 to 149 at the C-terminus), rates agree with predicted values within a factor of 3. The one exception, Thr2, exchanges faster than predicted. This behavior has been seen before for the +2 residue position (10). It should be noted that Thr2 is an outlier in that its maximum peak intensity exceeds the theoretical maximum according to Eq. 3.1. This suggests that in addition to the rapid HX rate expected for the +2 position, Thr2 also experiences significant magnetization transfer from other sources, perhaps due to nearby rapid exchangers such as the N-terminus and its own sidechain, which have been known to cause additional signal (6).

## 5. Conclusions

The CLEANEX-PM NMR pulse sequence is a powerful tool for measuring HX rates on the order of 1-100 s<sup>-1</sup> (6). Yet the complex nature of magnetization transfer from solvent into the protein had prevented previous workers from utilizing all available data. The step-wise fitting algorithm described here has resulted in more accurate HX rates from the same experimental data. On average the rates determined by this algorithm were 1.4 fold faster than those determined by a simple linear fit. However, this small average difference belies larger systematic deviations that occur when exchange rates are near the fast and slow detection limits (Fig. 3.9).

Rates were also improved by removing a pH-independent background signal from each time course. This signal likely arises from magnetization transfer through NOE pathways. Eq. 3.1 fails to capture this behavior because it describes a simple two spin system. A truly accurate equation describing magnetization transfer would require using a spatial structure of the protein in order to calculate cross relaxation, such as was described by Fan and colleagues (12).

Based on expected pH dependence (Fig. 3.10) and comparison to both theoretical and independently measured rates, we find phenomenological treatment of background signal combined with accurate fitting to the two spin description of magnetization transfer (Eq. 3.1) to markedly improve HX rate determination.

## **6. Bibliography**

1. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2008. A Billion-fold Range in Acidity for the Solvent-Exposed Amides of *Pyrococcus furiosus* Rubredoxin. *Biochemistry*. 47: 6178-6188.
2. Wagner, G., and K. Wüthrich. 1982. Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution: Studies with two-dimensional nuclear magnetic resonance. *Journal of Molecular Biology*. 160: 343-361.

3. Marcsisin, S.R., and J.R. Engen. 2010. Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Analytical and Bioanalytical Chemistry*. 397: 967-972.
4. Gemmecker, G., W. Jahnke, and H. Kessler. 1993. Measurement of fast proton exchange rates in isotopically labeled compounds. *J. Am. Chem. Soc.* 115: 11620-11621.
5. Hwang, T.-L., S. Mori, A.J. Shaka, and P.C.M. van Zijl. 1997. Application of Phase-Modulated CLEAN Chemical EXchange Spectroscopy (CLEANEX-PM) to Detect Water-Protein Proton Exchange and Intermolecular NOEs. *J. Am. Chem. Soc.* 119: 6203-6204.
6. Hwang, T.-L., P.C.M. van Zijl, and S. Mori. 1998. Accurate Quantitation of Water-amide Proton Exchange Rates Using the Phase-Modulated CLEAN Chemical EXchange (CLEANEX-PM) Approach with a Fast-HSQC (FHSQC) Detection Scheme. *Journal of Biomolecular NMR*. 11: 221-226.
7. Jeener, J., B.H. Meier, P. Bachmann, and R.R. Ernst. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics*. 71: 4546.
8. Mori, S., M.O. Johnson, J.M. Berg, and P.C.M. van Zijl. 1994. Water Exchange Filter (WEX Filter) for Nuclear Magnetic Resonance Studies of Macromolecules. *J. Am. Chem. Soc.* 116: 11982-11984.

9. Linderstrøm-Lang, K.U., and J.A. Schellman. 1959. Protein structure and enzyme activity. In: *The Enzymes*. NY: Academic Press. pp. 443-510.
10. Bai, Y., J.S. Milne, L. Mayne, and S.W. Englander. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins*. 17: 75-86.
11. Connelly, G.P., Y. Bai, M.F. Jeng, and S.W. Englander. 1993. Isotope effects in peptide group hydrogen exchange. *Proteins*. 17: 87-92.
12. Fan, J.-S., J. Lim, B. Yu, and D. Yang. 2011. Measurement of amide hydrogen exchange rates with the use of radiation damping. *Journal of Biomolecular NMR*. 51: 151-162.

## CHAPTER 4

### Comparing Hydrogen Exchange Models to Measured Experiments

#### 1. Introduction

The main chain amide hydrogens of protein molecules engage in continual exchange with the hydrogens of solvent water. Hydrogen exchange (HX) chemistry is well understood (1), and the HX rates of exposed amides in structureless polypeptides -- in any sequence, ambient condition, and isotope combination -- can be accurately predicted (2, 3). Amide hydrogens in structured proteins can exchange far more slowly, over a great range of rates, that depend on and therefore can provide information about biophysical properties (structure, stability, dynamics, energetics) and functional properties (interactions, structure change, folding) resolved to the individual amino acid level. To most effectively exploit this rich source of information, it will be necessary to understand the determinants of HX behavior. The common presumption that the determinants are well understood is far from true. These uncertainties limit the interpretive power of the many structural and functional HX studies that are now being reported.

In formative work before the first protein structure had been solved, Linderstrøm-Lang explored protein HX with the intention of looking for Pauling's H-bonded helices and sheets. Lang took the view that the slowly exchanging hydrogens found by his group did indeed represent Pauling's H-bonded structures. He proposed a simple

phenomenological model, namely that the exchange process requires structural protection to be relieved by some dynamic structural event that exposes the hydrogen to exchange (4–6). The picture is commonly represented as in Scheme 4.1 where  $k_{op}$  and  $k_{cl}$  are structural opening and reclosing rates, and  $k_{ch}$  is the chemical rate expected for freely exposed hydrogens (2).



It should be noted that this Scheme 4.1 is wholly kinetic in nature, providing no basis for distinguishing the closed and open states structurally. Over the years, the search for structural determinants of HX slowing has elicited a number of proposals. In general, prior workers have attempted top down strategies. Some broad structural determinant of HX rate is proposed and then an HX data base is tested for correlation with that factor. In order to examine these uncertainties we obtained a large HX data set for most of the amide hydrogens of the staphylococcal nuclease protein (SNase; Fig. 4.1).

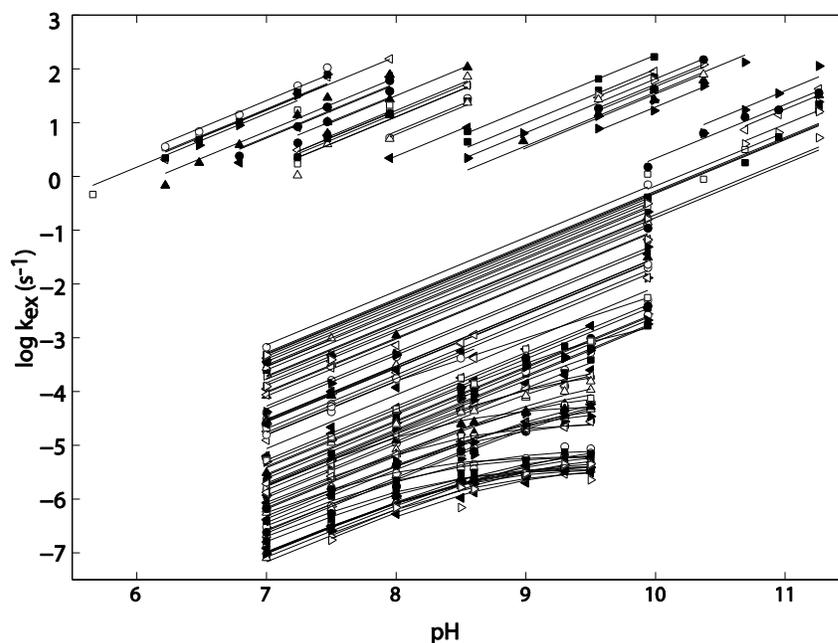
HX measurements were collected for the staphylococcal nuclease (SNase) mutant P117G/H124L. Assignments and  $^2\text{H}$ - $^1\text{H}$  experiments for this mutant were previously published (7). We have extended this analysis by measuring rapid  $^1\text{H}$ - $^1\text{H}$  exchange using the Cleanex-PM method (8). Combining these methods provided a nearly complete data set for the backbone amides in SNase (109 out of 143 backbone amide hydrogens and a Trp indole ring NH). In addition to the raw rates, information regarding the pH dependence (7) and denaturant dependence (9) reveal the nature of the opening reactions that allow HX to occur. We use these data to test models previously proposed in attempts

to understand protein HX rates based on known structure and calculated properties. Intuitively-based penetration models suppose that protein surface (“solvent accessible”) hydrogens will exchange rapidly and that the slower exchange of more buried hydrogens depends on the entry of water or the HX catalysts, hydroxide and hydronium ion, into the protein matrix (reviewed in (6). Electrostatic field has been suggested to play a major role in modulating the HX rate of amides exposed to solvent at the protein surface (10–13). For more buried hydrogens, algorithms based on local interaction density (14, 15) and on segmental unfolding reactions (16, 17) have been developed. We find these models do not adequately explain SNase HX data. More realistic determinants of HX rates are proposed in Chapter 5 on the basis local context.

## **2. Materials and Methods**

### **2.1 HX Rates**

$^1\text{H}$ - $^1\text{H}$  exchange measurements were collected by Dr. W. Lim as described in Chapter 3.  $^2\text{H}$ - $^1\text{H}$  measurements were collected by Dr. S. Bédard (7).



**Figure 4.1** Log  $k_{\text{ex}}$  vs pH. Lines are drawn with unit slope except for the slowest residues which were fit to eq. 4.1.

Measured HX rates as a function of pH are displayed in Fig. 4.1. The general HX behavior can be considered in terms of the Linderstrøm-Lang conformational kinetic model described in Scheme 1 (18). Once transient opening occurs, a kinetic competition ensues between exchange and reclosing. When structure is stable ( $k_{\text{cl}} > k_{\text{op}}$ ) and reclosing is fast ( $k_{\text{cl}} > k_{\text{ch}}$ ), the measured exchange rate,  $k_{\text{ex}}$ , depends on the fraction of time open (exchange competent), equal to the equilibrium constant  $K_{\text{op}}$ , as in Eq. 4.1. This is the so-called EX2, bimolecular exchange, case since the chemical exchange rate,  $k_{\text{ch}}$ , is proportional to the concentration of solvent HX catalyst, namely hydroxide ion over the pH range studied here. This dependence generates the unit slope of the rate – pH behavior in Fig. 4.1.

$$k_{\text{ex}} = K_{\text{op}} / (K_{\text{op}} + 1) k_{\text{ch}} \sim K_{\text{op}} k_{\text{ch}} \quad \text{Eq. 4.1}$$

With increasing pH,  $k_{ch}$  increases and may exceed  $k_{cl}$ . In this so-called EX1 case,  $k_{ex}$  reaches a limiting rate equal to  $k_{op}$ . The slowest exchanging SNase hydrogens in Fig. 4.1 exhibit an EX2 to EX1 transition, which will occur only for protected hydrogens when structural reclosing is slow relative to  $k_{ch}$ .

For EX2 exchange, hydrogen exchange rates are reported here in terms of the HX protection factor ( $P_f$ ) as in Eq. 4.2.

$$P_f = (K_{op} + 1) / K_{op} = k_{ch} / k_{ex} \quad \text{Eq. 4.2}$$

## 2.2 Structure-Based Calculations

Calculations were based on the SNase crystal structure 1SNO. Pro117 was mutated to glycine by deleting the extra atoms. Crystallographic waters were removed and hydrogens were added using CHARMM (19). Solvent accessible surface area (SASA) was calculated using Surfsv (20) with a 1.4 Å rolling ball probe radius and CHARMM22 (19) atom radii.

Electrostatic Poisson-Boltzmann continuum dielectric calculations were performed with Qnifft (21, 22). The  $\Delta pK_a$  for each amide was calculated by subtracting the energy of the protein with that amide deprotonated from the energy of the reference state. We used the same parameters previously described by LeMaster (12, 13), namely CHARMM22 atomic radii and charge distribution, an internal dielectric of 3 and an external dielectric of 78.5. The probe size was set to 1.4 Å with a grid size of 193 and scaling was set to 3.0. Temperature was set to 298 K with a monovalent salt

concentration of 150 mM. Other crystal structures used for electrostatic calculations were 1BQ8 (rubredoxin), 2PPN (FKBP12) 1LW6 (CI2) and 1UBQ (ubiquitin).

Structure-based predictions for individual residue HX rates in SNase P1175/H124L were provided by M. Vendruscolo and by T. Liu, based on their previously published methods (15, 23, 24).

### **3. Results and Discussion**

#### **3.1 Solvent Accessibility**

It is often stated that HX rates measure solvent accessibility, which leads to the expectation that protein surface hydrogens will exchange at rates close to those calibrated for amides in unstructured polypeptides. SNase observations are contrary to this view. Measured HX protection factors ( $P_f$ ; Eq. 4.2) versus distance to the protein surface are displayed in Fig. 4.2. Some hydrogens on unstructured SNase segments at the N- and C-termini do exchange rapidly, at their expected unprotected rates with protection factor  $\sim 1$  ( $\log P_f \sim 0$ ). However, other near-surface hydrogens exchange almost as slowly as the most deeply buried ones. Most of these are explained by protection due to hydrogen bonding on the solvent-exposed surfaces of regular secondary structures or loops. Other apparently unprotected hydrogens, but on structured segments, exchange more slowly than expected by factors between 2 and 40-fold. These are considered in Fig. 4.6 and in detail in Chapter 5.

In summary, whereas fast exchanging hydrogens that approximate the expected free peptide rate are placed only at the protein surface, many other hydrogens at the protein surface can exchange far more slowly.

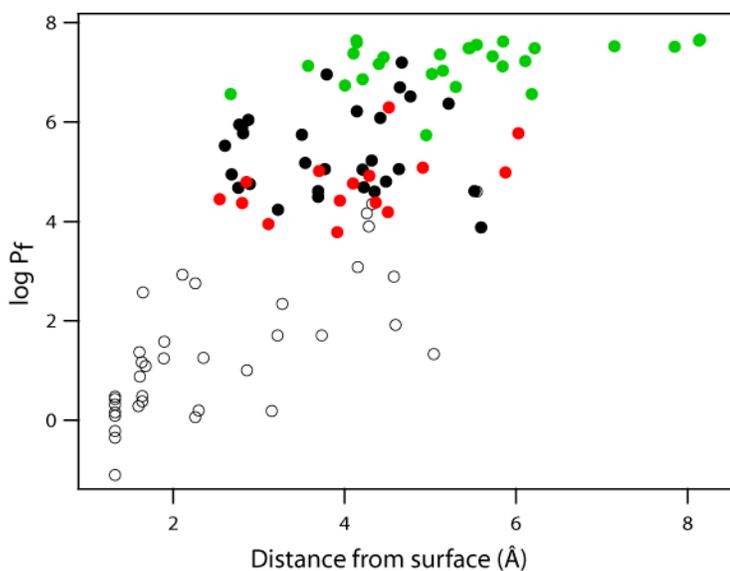
### 3.2 Solvent Penetration

A hypothesis related to the accessibility-dependent view is that slow exchange represents buried hydrogens and the HX process involves the penetration of water and/or hydroxide catalyst into the protein so that the rate-determining proton transfer event occurs within the protein matrix. Fig. 4.2 shows that many near-surface hydrogens, where the concept of penetration has little meaning, can exchange as slowly as buried ones. Also, HX protection has no significant dependence on depth of burial.

Further, many of the slowest hydrogens are known to exchange by way of unfolding reactions, both global and partial, rather than being reached *in situ* by incoming solvent species. This is indicated by the dependence of their exchange rate on GdmCl concentration (9). HX that occurs by way of a large unfolding reaction can be recognized by the observation of a sharp increase in HX rate with low concentrations of added denaturant (25). This occurs because denaturant promotes unfolding reactions, even under fully native conditions far below the melting concentration. In agreement, the unfolding reaction indicated for the slowest SNase hydrogens has a computed free energy equal to global unfolding ( $\Delta G_{op} = -RT \ln K_{op} = -RT \ln k_{ex}/k_{ch} = 10 \text{ kcal/mol}$ ) (7). Also, the HX rate for many very slow hydrogens is seen (Fig. 4.1) to roll over and approach pH-independent EX1 behavior at high pH (7). This behavior indicates unfolding reactions

that require  $> 1$  msec to reclose. Thus the slowest SNase hydrogens, which should be prime candidates for exchange by way of a penetration mechanism, instead exchange by way of sizeable unfolding reactions that expose them to solvent.

If a penetration mechanism exists, these results limit the HX rate that it could mediate. It must be slower than the slowest hydrogens measured here, namely at least 8 orders of magnitude slower than the structurally unhindered rate found for fully exposed amides.

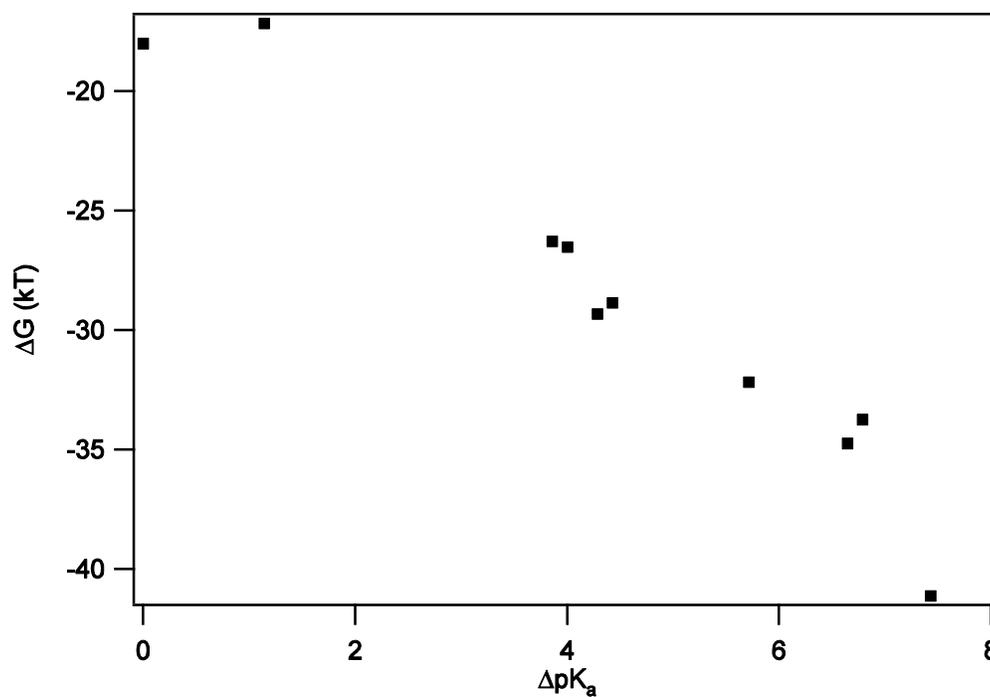


**Figure 4.2** Protection factors as a function of distance from the amide hydrogen to the water-protein interface. Colors indicate the exchange reaction occurs by way of a large unfolding reaction (green), a local fluctuation (red), or that the exchange reaction could not be determined (black). Data collected by  $^1\text{H}$ - $^1\text{H}$  exchange are indicated as open circles. A distance of  $1.32 \text{ \AA}$  is plotted for residues that did not appear in the crystal structure.

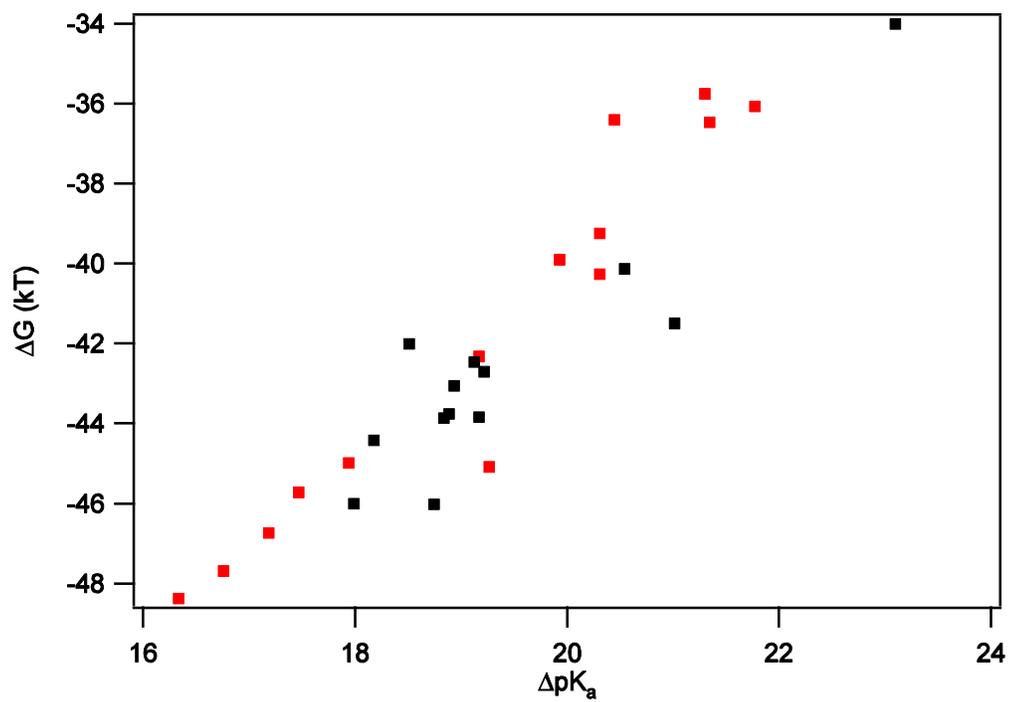
### 3.3 Electrostatic Effects on HX Rates

#### 3.3.1 Validating Electrostatic Calculations

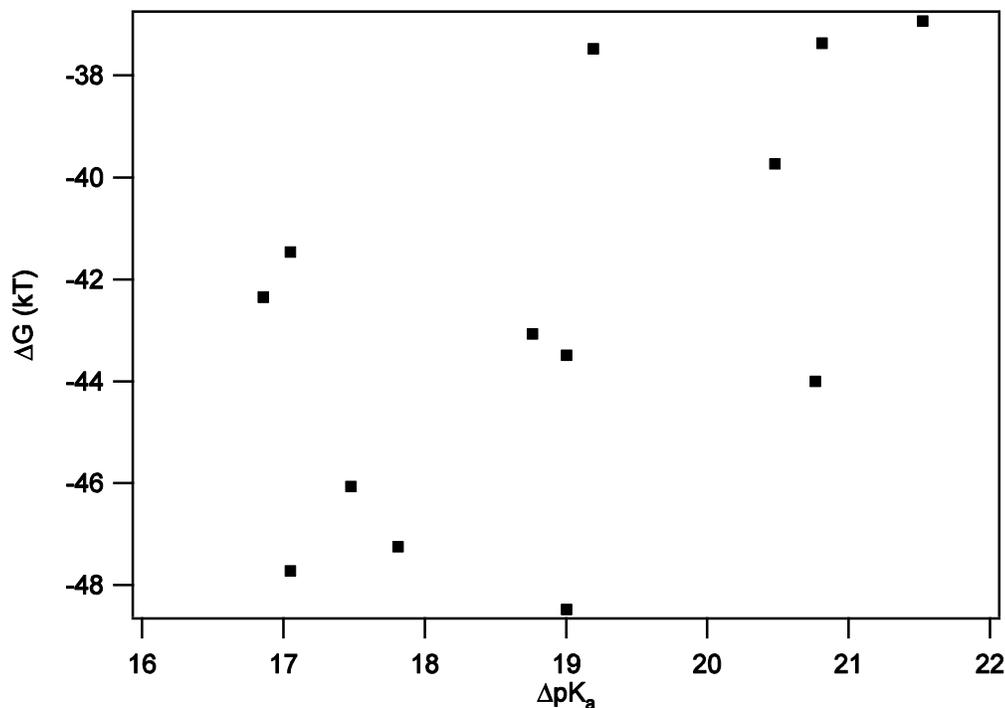
Some unprotected amide hydrogens proximal to water at the protein surface exchange more slowly than freely solvent exposed amides by up to 40-fold. A more extreme result has been described by the LeMaster group who reported the retardation of exchange rates for apparently solvent accessible amides in several small proteins by up to a billion-fold (12). They attributed this behavior to electrostatic effects on the relative acidity of the amide. To test this possibility, they used Poisson-Boltzmann continuum dielectric calculations to evaluate the change in amide  $pK_a$  values due to immediately local formal and partial charges. They plotted the calculated  $\Delta pK_a$  against measured HX rates, and fit a number of factors to improve the correlation. Only amides that have a solvent accessible surface area (SASA)  $\geq 0.5 \text{ \AA}^2$  were considered. It was assumed that this would ensure an unhindered rate of encounter with hydroxide catalyst, so that only the proton transfer equilibrium within the encounter complex, dependent on  $pK_a$ , would come into play.



**Figure 4.3**  $\Delta G$  of deprotonation calculated by Qniff compared to previously published  $\Delta pK_a$  calculations for rubredoxin (12). Published  $\Delta pK_a$  values were adjusted so that the slowest exchanging residue, Val38, was set to 0. The sign for published  $\Delta pK_a$  has presumably been reversed.  $\Delta G$  is as reported by Qniff. For both axes the offset should be considered arbitrary.



**Figure 4.4**  $\Delta G$  of deprotonation calculated by Qniff compared to previously published  $\Delta pK_a$  calculations for CI2 (black) and FKBP12 (red; (26)).  $\Delta pK_a$  scale was set relative to N-methylacetamide.



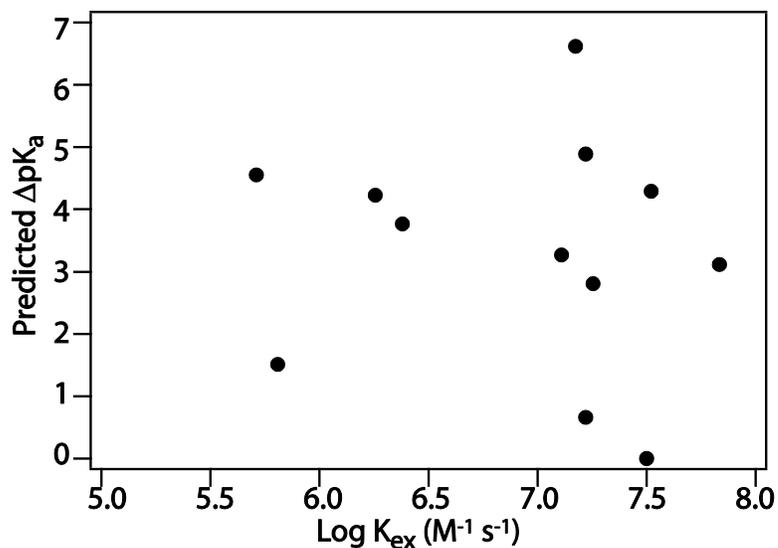
**Figure 4.5**  $\Delta G$  of deprotonation calculated by Qniff compared to previously published  $\Delta pK_a$  calculations for ubiquitin (13, 26).  $\Delta pK_a$  scale was set relative to N-methylacetamide. Published values were based on an extended NMR ensemble while our calculations were based on the crystal structure 1UBQ.

We repeated electrostatic calculations for the proteins studied by the LeMaster group using their favored parameters. Our calculations for rubredoxin, CI2 and FKBP12, essentially duplicate their results (Fig. 4.3 and 4.4). The small deviations likely arise from differences in how hydrogens were added to the crystal structures. Much larger deviations were seen for ubiquitin calculations (Fig. 4.5). These differences likely arise because our calculations were performed on a crystal structure while LeMaster and coworkers calculated  $pK_a$  values for ubiquitin using NMR-restrained ensembles (13). Not surprisingly, the ensemble based calculations match the measured HX rates better than

those calculated from a crystal structure. It is not clear why these calculations require the use of an ensemble for ubiquitin, but not for the other proteins studied by the LeMaster group.

### 3.3.2 Applying Electrostatic Calculations to SNase

The same calculations were applied to the slow amide protons on structured segments of SNase with computed amide H SASA  $> 0.5 \text{ \AA}^2$ . These results are displayed in Fig. 4.6. The measured HX rates show no correlation with calculated electrostatic field. It is notable that the calculated pKa range corresponds to a  $k_{\text{ex}}$  range that is seven orders of magnitude wide, as before, but the observed range of HX rates spans less than three orders of magnitude.



**Figure 4.6**  $\Delta pK_a$  from electrostatic calculations vs measured HX rates for amides with  $>0.5 \text{ \AA}^2$  ASA.

It is noteworthy that rate versus field correlations found in previous studies with other proteins (12, 26) depend heavily on a few very slow hydrogens that anchor the correlation curves. Is it possible that these hydrogens are structurally protected in some non-obvious way? In chapter 5 it is demonstrated that structurally bound water molecules protect a number of SNase amides from HX. When solvent is treated as a continuum, as in previous studies, the usual rolling ball SASA analysis will fail to distinguish near-surface structurally-bound waters from free solvent. Prior work is considered from this point of view.

Analysis shows that some slowly exchanging hydrogens considered to be solvent exposed in previous studies are H-bonded to crystallographically defined water molecules held in place by specific protein interactions. In CI2 (26), the slowest supposedly unprotected hydrogen (Arg62) is H-bonded to a crystallographically defined water molecule that also H-bonds to a main chain carbonyl between two beta strands. Other slow amides (Val31, Ile44, Asp45) are H-bonded to waters that are also H-bonded to two main chain or side chain carbonyls. These waters could not be simply replaced with catalyzing hydroxide (27) without some structural adaptation, and therefore must offer protection against HX. Five of the remaining ten amides in CI2 that contribute to the rate - field correlation are in a loop that interacts with a neighboring protein in the crystal. This makes uncertain their local environment in the monomeric solution condition of the HX experiments and therefore the calculated field.

In rubredoxin (12), the very slowly exchanging Lys46 amide hydrogen is H-bonded to a defined water molecule that is held also by H-bonding to main chain carbonyls. The Asp14 amide is H-bonded to a protecting water that is also H-bonded to a

main chain carbonyl and to a tyrosine side chain OH. The slowest hydrogen, on Val38, is H-bonded to a defined water molecule that is involved in a network of water molecules held in a cleft between Val38 and an external loop. In other cases where apparently exposed but slowly exchanging hydrogens help to anchor the field/rate correlation curve, the slow amides are also H-bonded to defined water molecules held in place by H-bonding to protein groups.

The different water molecules in this list are often held in the protein structure by two other H-bonds, but in some cases the water binds to only one other protein group perhaps making the question of stable protection ambiguous. However, recent work suggests that hydroxide does not act as an H-bond donor (27), which suggests that hydroxide is unlikely to replace any water molecule that donates an H-bond to the protein. The protein - water interactions for the hydrogens that anchor the previously described rate-field correlation curves will naturally produce extreme values. They make exchange slow by blocking attack by hydroxide HX catalyst, and they ensure that the Poisson-Boltzmann calculation will compute a strongly unfavorable deprotonation environment. When the slowly exchanging hydrogens identified as structurally protected in this way are removed from the field/HX correlation curves only a much smaller range of HX rates due to clearly unprotected solvent exposed hydrogens remains and no correlation of HX rates with electrostatic field is apparent.

Previous workers have found that exchange rates of freely exposed amides can be affected by and in some cases explained by local electrostatic effects (11, 28, 29). The question considered here is whether such effects account for the surprisingly large HX retardation seen for some protein surface amides. Considerations just described leave

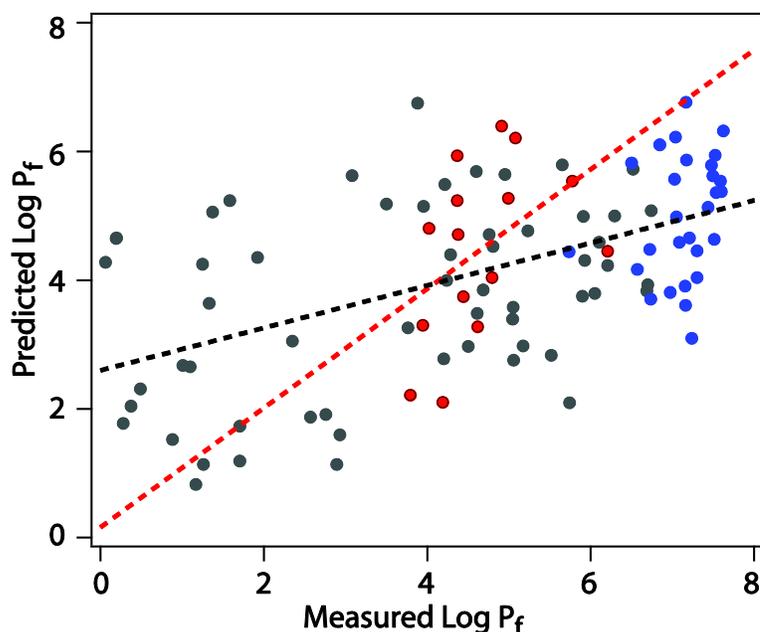
uncertainty concerning the reality of the correlation inferred before. Some re-examination of factors in the analysis may be useful. This especially includes explicitly bound water molecules, but also the use of a  $0.5 \text{ \AA}^2$  filter in a continuum water calculation to identify freely accessible hydrogens, and the use of an internal dielectric model for the protein with  $\epsilon = 3$  since the surprisingly large dynamic range calculated for  $\text{pK}_a$  varies linearly with  $1/\epsilon$  (12).

### 3.4 Packing Density

Vendruscolo and coworkers investigated the use of HX data to calculate the dynamic protein ensemble by including an additional potential energy term in their molecular dynamics calculation that biases the simulation to match a subset of known HX data (15, 23, 30). They hypothesized that HX rates are connected to structure by Eq. 4.3.

$$\ln P_f = B_c N_c + B_h N_h \quad \text{Eq. 4.3}$$

For each amino acid amide,  $P_f$  is the HX protection factor (Eq. 4.2),  $N_c$  and  $N_h$  are, respectively, the number of contacts and the number of H-bonds. The B coefficients were optimized by comparing with known HX rates for a number of reference proteins (15, 30). Fig. 4.7 shows that the HX results for SNase are not well correlated with local interactions in the structure as formalized in Eq. 4.3.



**Figure 4.7** Measured log  $P_f$  vs log  $P_f$  predicted based on local packing density. Colors indicate cases where exchange is known to occur by a large unfolding reaction (blue), or local fluctuation (red), or if the unfolding reaction is unknown (black). The best fit line (black) to the entire dataset has an  $r^2$  of 0.26. Predictions were provided by M. Vendruscolo.

One reason that these predictions poorly correlate with measurements is that these calculations only take into account local structure while HX is known to occur through multiple mechanisms (31). Best and Vendruscolo pointed to this factor as the explanation for why their calculations grossly underestimate  $P_f$  values in CI2 for residues known to exchange by global unfolding (23). However, this explanation is not consistent with a model in which residues can exchange by multiple mechanisms with the fastest pathway dominating the measured rate. For residues that exchange by global unfolding one would expect the  $P_f$  values associated with their local fluctuation pathways to be even higher than the measured  $P_f$  rather than the other way around.

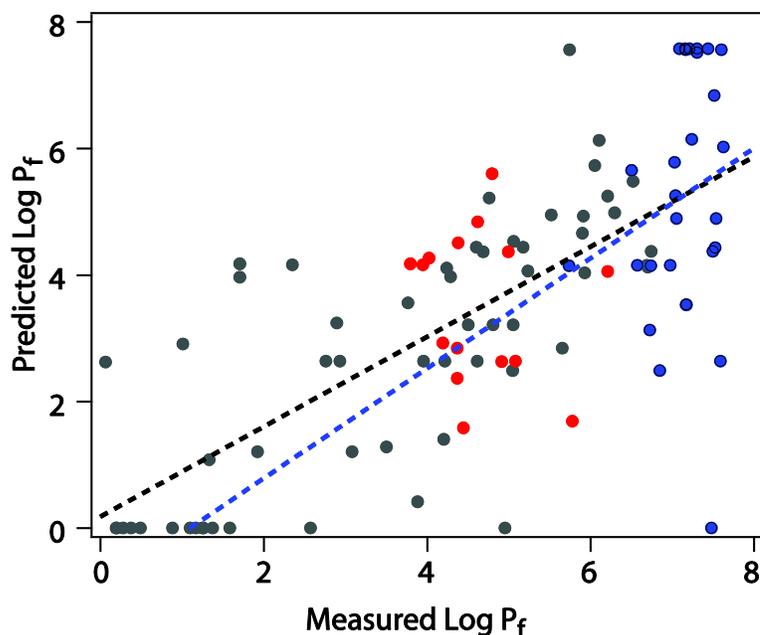
Another reason for the less than ideal correlation is that coefficients were parameterized using HX data for which the exchange mechanism was unknown. This could explain why predictions do not fare better for residues known to exchange by local fluctuations than they do for the entire data set (Fig. 4.7).

Perhaps the most likely explanation is that Eq. 4.3 is not representative of the factors underlying local fluctuations. Based on Scheme 4.1 one can think of each amide existing in equilibrium between exchange competent conformations and protected conformations. If an equation is to accurately predict HX rates then it must somehow relate to the energetics underlying that equilibrium. Eq. 4.3 assumes that each contact made by the residue is equally valuable for protecting it from exchange except for the H-bond, which has an extra weighting factor. This assumption breaks down for any structured residue whose backbone amide does not engage in an intramolecular H-bond (i.e. half of the residues on the edge of a  $\beta$ -sheet). In this case Eq. 4.3 would predict protection due to local contacts while Scheme 4.1 would predict no protection. Similarly, Eq. 4.3 takes into account the environment of the H-bond donor, but not the acceptor. Based on Scheme 4.1 one expects  $P_f$  to be independent of whether it is the displacement of the amide or the H-bond acceptor that exposes the amide to HX catalyst. Analysis described in Chapter 5 indicates that some residual slowing does occur when only the H-bond acceptor is displaced, but that this slowing is limited to  $P_f < 40$ .

### 3.5 COREX

Hilser and coworkers have developed the COREX algorithm for predicting the stability of each residue in a protein based on the native structure and the measured global stability (32, 33). This algorithm generates an ensemble of all possible states in which one or more regions of consecutive residues are unfolded. Each state is assigned an energy based on known thermodynamic parameters regarding the exposed surface area of the unfolded segments and the remaining protein. For standard COREX calculations a residue-specific stability is determined by summing the energy of the states in which the residue is in the native structure and dividing by the energy summation for all states in which the residue is unfolded. Recently, an additional step has been taken to relate COREX calculations to HX protection factors in what is now referred to as H-COREX (24). This extended algorithm functions similarly to the original COREX algorithm, but now accounts for the states in which a residue remains structured but is HX competent presumably due to unfolding of residues acting as its H-bond acceptor. This algorithm calculates  $P_f$  by dividing the number of states in which the residue is folded and protected by the number of states in which it is unfolded plus the number of states in which it is folded yet exposed (equivalent to  $1/K_{op}$ ).

Fig. 4.8 shows that the SNase HX data do not match the H-COREX predictions. One expects discrepancies to arise if the ensemble predicted by COREX does not match reality or if the criteria used to define the “structured yet exposed” states do not correspond to the exchange competent states.



**Figure 4.8** Log  $P_f$  predicted by H-COREX vs measured log  $P_f$ . Colors indicate that exchange takes place by a large unfolding reaction (blue), a local fluctuation (red) or if the unfolding reaction is unknown (black). The best fit line (black) to the full dataset has an  $r^2$  of 0.52. Predictions were provided by T. Liu.

One limitation of COREX calculations is that a minimum window size must be used when generating the ensemble (typically six to eight residues). This prevents COREX from describing unfoldings that are smaller than the minimum window size. Many residues in SNase exhibit no measurable change in  $P_f$  in low concentrations of denaturant indicating that very little additional surface area is exposed when these residues access their exchange competent states (25). However, H-COREX does not fare any worse when predicting  $P_f$  for these residues than it does when predicting  $P_f$  for residues that exhibit significant denaturant dependence (Fig. 4.8). It should be noted that Hilser and coworkers expect COREX to be just as predictive for residues without denaturant dependence as it is for all other residues (16). The suggestion is that the fully folded state seen in crystal structures is rarely populated and that this state actually represents the average structure. They postulate that any given protein molecule is more

likely to have a small section of structure unfolded than it is for every residue to be in the folded state. The argument follows that these residues do not exhibit denaturant dependence because the surface area that is exposed when they exchange is approximately equal to the exposed surface area of the most populated states. However, it seems unlikely that the ensemble is predominated by partial unfoldings similar to those that expose these residues as their  $P_f$  values indicate that any given amide is exposed less than 0.1% of the time. A detailed structural analysis in the next chapter will reveal cases where the motions underlying exchange must be smaller than those assumed by COREX.

The ensemble generated by COREX will also deviate from reality when partial unfolding leads to non-native interactions or when the unfolded regions do not behave as random coils. Non-native interactions have been attributed to high protection factors in SNase (7) as well as other proteins. These interactions occur when an amide finds an alternative H-bond acceptor while its neighboring residues are unstructured. Similarly, local environment may prevent an unstructured segment from behaving as a random coil. This is particularly likely for residues near the folded-unfolded interface and for all residues in very short segments. For these residues, steric considerations will prevent the backbone from randomly sampling Ramachandran space, which could limit amide exposure or alter the intrinsic exchange rate (34).

Even if the COREX algorithm perfectly described the protein ensemble there is no guarantee that this ensemble is accurately translated into  $P_f$ . Any attempt to determine  $P_f$  from an ensemble will require an accurate structural description of the HX competent state. The phenomenological equation described in Scheme 1 is useful for interpreting

HX data, but it is silent regarding how to categorize structure as “open” or “closed.” The use of ASA calculations for this purpose may be an oversimplification given that protein-solvent interactions cannot always be described by a continuum water model (35). The role of water molecules as protecting H-bond acceptors is explored in Chapter 5.

#### 4. Conclusions

The evidence against exchange *in situ* for both surface and buried hydrogens (Fig. 4.2) favors a dominant role for dynamic structural perturbations in determining protein HX behavior, as in the general Linderstrøm-Lang model (Scheme 4.1). We find that even surface hydrogens often exchange slowly because they are protected by local structure, and they require dynamic perturbations that expose the hydrogen to attack by HX catalyst. Other information makes it clear that buried hydrogens require even more dramatic structural excursions to bring them into an exchange competent state. Up to 40-fold slowing was observed for constitutively exposed residues, but was not explained by standard electrostatic calculations (Fig. 4.6). Predictions based on local contacts (Fig. 4.7) or on unfolding reactions (Fig. 4.8) each capture the general trend in HX rates, but do not predict Pf with a high degree of accuracy ( $r^2 = 0.26$  and  $0.52$  respectively).

Each of the methods tested here attempt to account for one facet of HX. A truly predictive method would need to be based on a model that accounts for all of the determinants of HX rates. This work has inspired a re-examination of the factors underlying exchange which is discussed in the following chapter.

## 5. Bibliography

1. Eigen, M. 1964. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: Elementary Processes. *Angew. Chem. Int. Ed. Engl.* 3: 1-19.
2. Bai, Y., J.S. Milne, L. Mayne, and S.W. Englander. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins.* 17: 75-86.
3. Connelly, G.P., Y. Bai, M.F. Jeng, and S.W. Englander. 1993. Isotope effects in peptide group hydrogen exchange. *Proteins.* 17: 87-92.
4. Linderstrøm-Lang, K.U., and J.A. Schellman. 1959. Protein structure and enzyme activity. In: *The Enzymes*. NY: Academic Press. pp. 443-510.
5. Hvidt, A., and S.O. Nielsen. 1966. Hydrogen exchange in proteins. *Advances in Protein Chemistry.* 21: 287-386.
6. Englander, S.W., and N.R. Kallenbach. 1983. Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids. *Quarterly Reviews of Biophysics.* 16: 521-655.
7. Bédard, S., L.C. Mayne, R.W. Peterson, A.J. Wand, and S.W. Englander. 2008. The Foldon Substructure of Staphylococcal Nuclease. *Journal of Molecular Biology.* 376: 1142-1154.
8. Skinner, J.J., W.K. Lim, S. Bédard, B.E. Black, and S.W. Englander. 2011. Protein hydrogen exchange: testing current models. *Biophys J.* .

9. Wrabl, J.O. 1999. Investigations of denatured state structure and m-value effects in staphylococcal nuclease. .
10. Matthew, J.B., and F.M. Richards. 1983. The pH dependence of hydrogen exchange in proteins. *Journal of Biological Chemistry*. 258: 3039 -3044.
11. Fogolari, F., G. Esposito, P. Viglino, J.M. Briggs, and J.A. McCammon. 1998. pKa Shift Effects on Backbone Amide Base-Catalyzed Hydrogen Exchange Rates in Peptides. *J. Am. Chem. Soc.* 120: 3735-3738.
12. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2008. A Billion-fold Range in Acidity for the Solvent-Exposed Amides of *Pyrococcus furiosus* Rubredoxin. *Biochemistry*. 47: 6178-6188.
13. LeMaster, D.M., J.S. Anderson, and G. Hernández. 2009. Peptide Conformer Acidity Analysis of Protein Flexibility Monitored by Hydrogen Exchange. *Biochemistry*. 48: 9256-9265.
14. Bahar, I., A. Wallqvist, D.G. Covell, and R.L. Jernigan. 1998. Correlation between Native-State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model†. *Biochemistry*. 37: 1067-1075.
15. Vendruscolo, M., E. Paci, C.M. Dobson, and M. Karplus. 2003. Rare Fluctuations of Native Proteins Sampled by Equilibrium Hydrogen Exchange. *J. Am. Chem. Soc.* 125: 15686-15687.

16. Wooll, J.O., J.O. Wrabl, and V.J. Hilser. 2000. Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *Journal of Molecular Biology*. 301: 247-256.
17. Hilser, V.J., B. García-Moreno E., T.G. Oas, G. Kapp, and S.T. Whitten. 2006. A Statistical Thermodynamic Model of the Protein Ensemble. *Chem. Rev.* 106: 1545-1558.
18. Englander, S.W., L. Mayne, Y. Bai, and T.R. Sosnick. 1997. Hydrogen exchange: the modern legacy of Linderstrøm-Lang. *Protein Sci.* 6: 1101-1109.
19. MacKerell, A., B. Brooks, C.L. Brooks, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998. CHARMM: the energy function and its parameterization with an overview of the program. In: Schleyer P, editor. *The encyclopedia of computational chemistry*. Chichester: Wiley. pp. 271-277.
20. Nicholls, A., K.A. Sharp, and B. Honig. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Structure, Function, and Bioinformatics*. 11: 281-296.
21. Gilson, M.K., K.A. Sharp, and B.H. Honig. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry*. 9: 327-335.
22. Sharp, K.A., R.A. Friedman, V. Misra, J. Hecht, and B. Honig. 1995. Salt effects on polyelectrolyte–ligand binding: Comparison of Poisson–Boltzmann, and limiting law/counterion binding models. *Biopolymers*. 36: 245-262.

23. Best, R.B., and M. Vendruscolo. 2006. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure*. 14: 97-106.
24. Liu, T., D. Pantazatos, S. Li, Y. Hamuro, V.J. Hilser, and V.L. Woods. 2011. Quantitative Assessment of Protein Structural Models by Comparison of H/D Exchange MS Data with Exchange Behavior Accurately Predicted by DXCOREX. *Journal of The American Society for Mass Spectrometry*. .
25. Bai, Y., T. Sosnick, L. Mayne, and S. Englander. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science*. 269: 192 -197.
26. Hernández, G., J.S. Anderson, and D.M. LeMaster. 2009. Polarization and polarizability assessed by protein amide acidity. *Biochemistry*. 48: 6482-6494.
27. Marx, D., A. Chandra, and M.E. Tuckerman. 2010. Aqueous basic solutions: hydroxide solvation, structural diffusion, and comparison to the hydrated proton. *Chem. Rev.* 110: 2174-2216.
28. Avbelj, F., and R.L. Baldwin. 2004. Origin of the neighboring residue effect on peptide backbone conformation. *Proceedings of the National Academy of Sciences of the United States of America*. 101: 10967 -10972.
29. Avbelj, F., and R.L. Baldwin. 2009. Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added. *Proceedings of the National Academy of Sciences*. 106: 3137 -3141.

30. Tartaglia, G.G., A. Cavalli, and M. Vendruscolo. 2007. Prediction of Local Structural Stabilities of Proteins from Their Amino Acid Sequences. *Structure*. 15: 139-143.
31. Skinner, J.J., W.K. Lim, S. Bédard, B.E. Black, and S.W. Englander. 2011. Protein dynamics: The view from hydrogen exchange. *Proc Natl Acad Sci US*. .
32. Hilser, V.J., and E. Freire. 1996. Structure-based Calculation of the Equilibrium Folding Pathway of Proteins. Correlation with Hydrogen Exchange Protection Factors. *Journal of Molecular Biology*. 262: 756-772.
33. Hilser, V.J. 2001. Modeling the Native State Ensemble. In: *Protein Structure, Stability, and Folding*. New Jersey: Humana Press. pp. 093-116.
34. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2009. Backbone conformational dependence of peptide acidity. *Biophysical Chemistry*. 141: 124-130.
35. Ho, J., and M.L. Coote. 2009. A universal approach for continuum solvent pK<sub>a</sub> calculations: are we there yet? *Theoretical Chemistry Accounts*. 125: 3-21.

## CHAPTER 5

### Interpreting Protein Dynamics from Hydrogen Exchange Data

#### 1. Introduction

The naturally occurring exchange of protein and water hydrogens depends on and therefore can provide detailed information about the biophysical properties and functional behavior of protein molecules (1). This capability is widely exploited in current protein studies (2–5). It is often conceived that the determinants of protein hydrogen exchange (HX) rates and behavior are well understood, but this is not true. The lack of detailed understanding diminishes the interpretive power of the many studies that are now being done.

In searching for the structural factors that determine HX slowing, prior workers have attempted top down strategies (described in (6)). One hypothesizes some broad general determinant and then tests an HX database for correlation with that factor. To explain unexpectedly slow exchange at the protein surface, factors considered have included relative solvent exposure (7) and electrostatic field (8). For more buried hydrogens, various hypotheses have considered solvent penetration into the protein (9), degree of burial, H-bonding (10, 11), local packing density (12), and transient unfolding reactions (13). In the previous chapter these different models were compared to a large HX data set that covers the entire HX range of the amide hydrogens of the well-studied protein, staphylococcal nuclease (SNase; (6)). Predictive success was not impressive suggesting that multiple mechanisms are involved.

The availability of detailed structural information and extensive HX data now makes it possible to pursue a more systematic examination. We take a bottom up approach and attempt to discern HX mechanism at an amino acid level by examining the HX behavior of many individually resolved SNase residues and their neighbor relationships (14). The  $\alpha/\beta$  protein SNase is sufficiently large and varied that it provides a good sampling of the various possible HX mechanisms. The detailed HX information is supplemented by knowledge of the sensitivity of the exchange of SNase hydrogens to denaturant (15) and pH (16) which can be used to distinguish the relative size of structural excursions (17, 18). All of this information together provides a resource for considering in some detail the static and dynamic factors that determine HX rates. The comparison of amino acid-resolved HX results and patterns in detailed structural context produces a closer view of the various modes by which dynamic local and larger protein motions can release structural protection, on scales ranging from a crankshaft motion about a pair of alpha carbons to the cooperative unfolding of the entire protein.

## **2. Materials and Methods**

$^1\text{H}$ - $^1\text{H}$  exchange measurements were collected by Dr. W. K. Lim. HX rates were determined from these data as described in Chapter 3.  $^2\text{H}$ - $^1\text{H}$  measurements were collected by Dr. S. Bédard (16). Denaturant dependence was reported by Dr. J. Wrabl (15). Details regarding experimental conditions are described in Chapter 4.

The SNase crystal structure 1SNO was used as the starting structure for analysis. The structure was mutated from H124L to P117G/H124L by deleting the extra atoms in

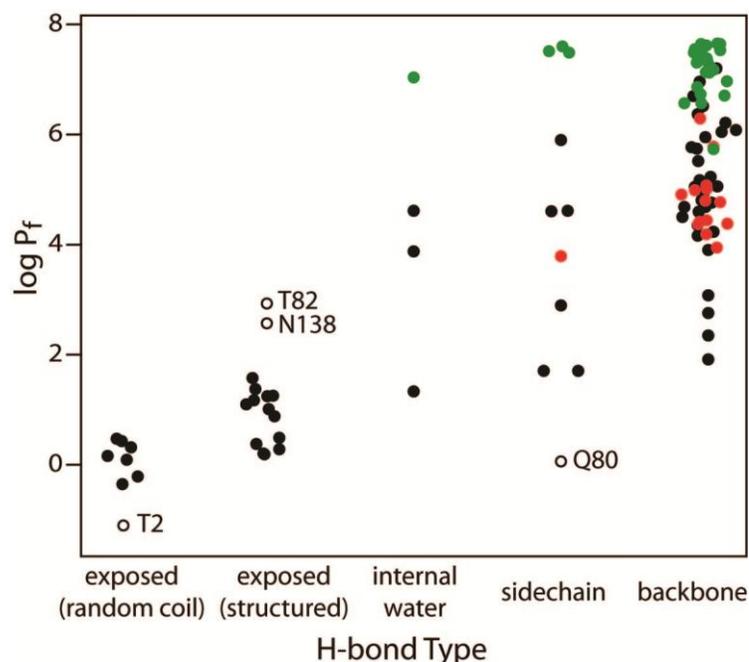
residue 117. Hydrogens were then added using CHARMM(19). The NMR structure set 2KQ3 was used for comparison. Figures 5.2-5.8, 5.10 and 5.11 were created using PyMol. Crystal lattice analysis was done using Coot (20).

An inventory of all possible hydrogen bonds was created using the broad criteria of maximum H-O distance of 3.5Å, minimum NHO angle of 100° and minimum HOC angle of 90°. This list was then reduced by inspecting the structure manually.

### **3. Results and Discussion**

#### **3.1 Hydrogen Bond Acceptor Types**

The secondary structural elements of SNase include a distorted 5-stranded  $\beta$ -barrel, three major  $\alpha$ -helices, and connecting loops. Examination shows that almost all SNase amide hydrogens are involved in H-bonding, either to other protein groups or to water molecules. This seems to be true of proteins in general because the incorporation of a non-H-bonded amide incurs a cost of several kcals in stabilization free energy (21). Fig. 5.1 presents an H-bonding inventory of SNase amides according to their H-bond acceptor, most often main chain or side chain groups of the protein. Some amides form H-bonds to a defined water molecule, which can be internal or can interact with bulk solvent. For a few surface amides, classified as exposed, no H-bonding partner or interacting water is well defined, presumably due to translational freedom or lattice contacts.



**Figure 5.1** Hydrogen bond inventory for SNase. Colors indicate amide hydrogens that exchange by way of a large unfolding reaction (green; high denaturant dependence and/or EX1 behavior), or a small local fluctuation (red; observed denaturant dependence about zero), or whether this information is not available (black). Amides that H-bond to solvent are categorized by whether they reside on unstructured or well-structured segments. Open circles represent ambiguous cases. Among unstructured amides, the HX rate for Thr2, obtained only at pH 11.3, is unreliable. Contrary to x-ray analysis (Fig. 5.7), NMR shows Asn138 H-bonded to the  $\alpha$ 2 helix (Fig. 5.10). HX rates and the presence of an interfering lattice contact suggest that the Gln80 side chain in fact protects Thr82 rather than Gln80.

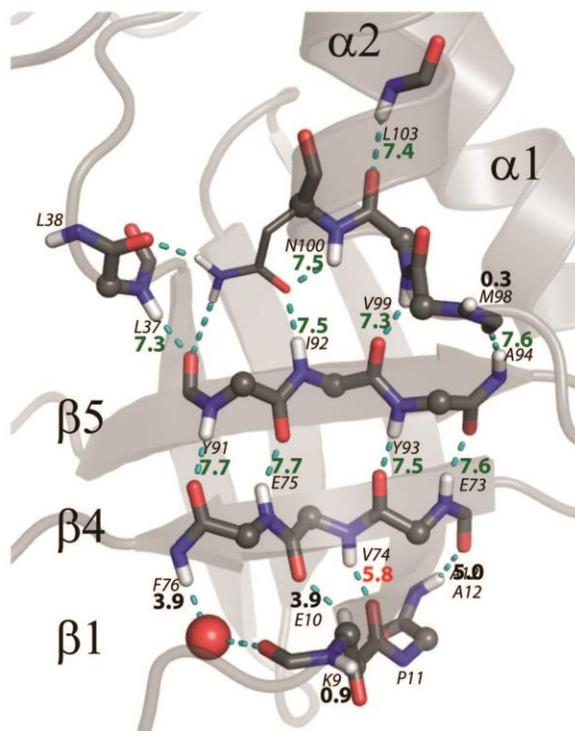
Eight measured amides are in apparently unstructured segments near the protein termini that are not defined in the x-ray structure but do produce sharp NMR lines, indicating high structural mobility. They exchange at close to their expected freely exposed amide rate (22) with HX protection factor  $\sim 1$  ( $\log P_f \sim 0$ ). However, surface amides on structured segments, although H-bonded to solvent water, are slowed on average by  $\sim 10$  fold ( $P_f$  range from 1.5 and 38). Amides protected by H-bonding to backbone carbonyls, side chains, or structurally incorporated water molecules all span essentially the whole HX rate range indicating that HX slowing is not determined by the H-bond type.

We find that hydrogens protected by H-bonding to other protein groups or internal waters cannot exchange at all, even when they are in direct contact with solvent (see below). One wants to understand the structural distortion reactions that allow exchange. Figs. 5.2 to 5.5 each show a region of the SNase structure, with H-bond donors and acceptors, and the measured  $\log P_f$  of individual amide hydrogens. In many of these cases, examination of the pattern of exchange of neighboring residues indicates the deprotection motions.

## **3.2 Structural Context**

### **3.2.1 Global unfolding**

Fig. 5.2 shows one face of the 5-stranded SNase  $\beta$ -barrel. Protecting H-bond acceptors include main chain carbonyls, a side chain carbonyl, and a crystallographically defined water molecule. Some conclusions about mechanism can be drawn.



**Figure 5.2** HX protection pattern indicating global unfolding on one face of the SNase  $\beta$ -barrel. Standard colors and stick diagrams identify main chain atoms, superimposed on secondary structural elements and connecting loops diagrammed as background. For clarity C $\alpha$ s are displayed as small spheres and the Asn100 sidechain is shown as thin sticks. H-bonds are displayed as dashed lines and amide donors are labeled. Log  $P_f$  is shown for each amide for which it could be determined. Color of Log  $P_f$  indicates that HX occurs by way of unfolding (green), local fluctuations (red), or unknown (black). The red sphere is a crystallographically defined water molecule.

The surface of the  $\beta$ -structure is in direct contact with water. For example, the amide NHs of Lys9 and Met98 face into the solvent and exchange at near the free peptide rate. Nevertheless immediately neighboring amides, all at the aqueous surface, exchange more slowly by 10 million-fold. This behavior is shared by amides in varied structures, in  $\beta$ -sheet and  $\alpha$ -helix, and they are protected by varied H-bonding, between  $\beta$  strands, within a helix, and even with a side chain (Asn100). Clearly, close proximity to solvent

does not ensure exchange. These results limit the rate of exchange by direct transfer to solvent HX catalyst (hydroxide) to  $< 10^{-8} \text{ M}^{-1} \text{ s}^{-1}$ .

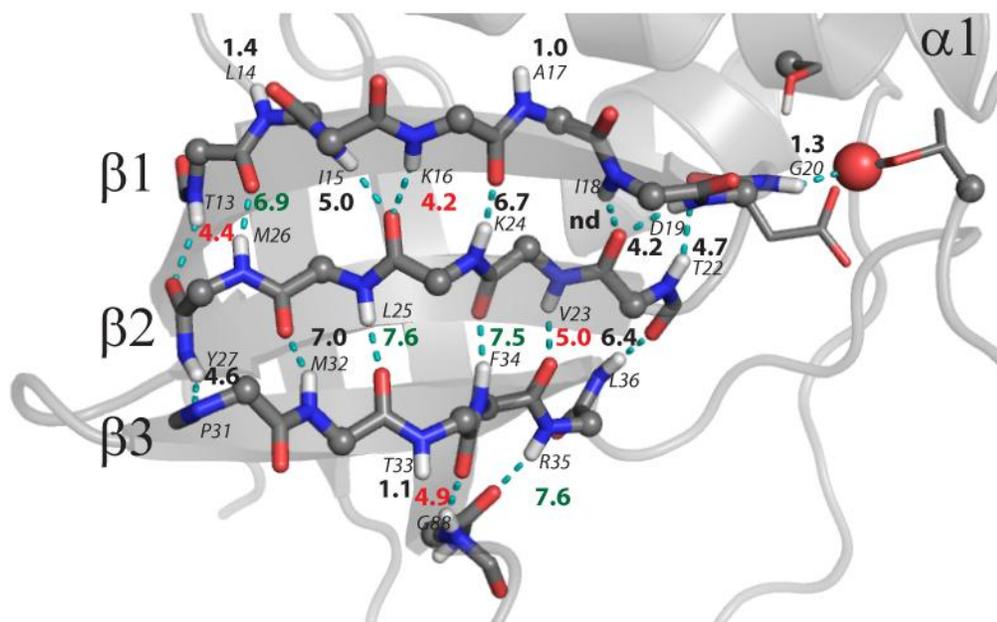
Exchange of the very slow hydrogens exhibits a sharp dependence on denaturant (15), suggesting a large scale unfolding. Many exchange with the same large protection factor ( $\log P_f$  between 7.3 and 7.7), corresponding to an unfolding free energy of  $\sim 10$  kcal/mol, equal to the protein stability (16). Also, they exhibit EX1 HX at pH 8.5 and above (16), again indicating a large unfolding – slow refolding HX mechanism. The measured EX1 HX rate indicates a reclosing rate ( $\sim 10^2 \text{ s}^{-1}$ ) (16), close to the rate for the first folding phase in refolding from the denatured state (23). These results demonstrate that the exchange of the very slow hydrogens is mediated by the reversible global unfolding/refolding reaction under native conditions.

Despite being involved in similar structure, nearby amides H-bonded between  $\beta 4$  and  $\beta 1$  exchange with varied faster rates ( $\log P_f = 3.9$  to 5.8), suggesting a number of smaller fluctuations. The fluctuations that expose the amide NHs of Val74 and Phe76 involve their H-bond acceptors and not the amides themselves. This follows from the fact that peptide groups are rigidly planar. Displacement of the Val74 and Phe76 amide NHs would equally displace their carbonyls and deprotect the  $\beta 3/\beta 4$  interface amides, but this does not occur. The  $\beta 4$  main chain remains in place. Exchange seems to involve a fraying motion at the N-terminus of  $\beta 1$ . In agreement, Val74 has zero dependence on denaturant, indicating exchange by a local fluctuation. Similar motions seem possible for the slower hydrogens in principle, but the data require that such modes can only provide even slower exchange than is observed.

Note the contrast with the side chain of Asn100 which provides strong protection, presumably because it is rigidly held in place, packed between a sheet and a helix which are themselves well packed. Similarly, a water molecule provides HX protection for Phe76 (see also Fig. 5.8). The degree of protection should not depend on the frequency of water dissociation but on the preferential equilibrium binding of H<sub>2</sub>O versus OH<sup>-</sup> (like K<sub>op</sub> in Scheme 1). In this particular case, deprotection occurs not when the water leaves but when  $\beta$ 1 frays, as discussed in section 3.2.4.

### 3.2.2 Heterogeneous Unfolding in a Beta Sheet

Fig. 5.3 shows the opposite face of the  $\beta$ -barrel, starting farther along the  $\beta$ 1 strand as it wraps around the barrel. As before, hydrogens are well protected by H-bonding even though most of this surface is in direct contact with solvent, but here the HX pattern is more varied. Some of the hydrogens exchange very slowly, with denaturant dependence (15) and EX1 behavior (16) indicative of a concerted unfolding reaction (green), as before. Other immediately neighboring residues undoubtedly experience the same large unfolding but they show a diversity of faster rates. Some individual residues on  $\beta$ 1 and  $\beta$ 2 are much faster than their neighbors, pointing to exchange by way of local fluctuations in which the major strands remain in place. In agreement, the denaturant dependence measured for some of these residues is negligible (red).



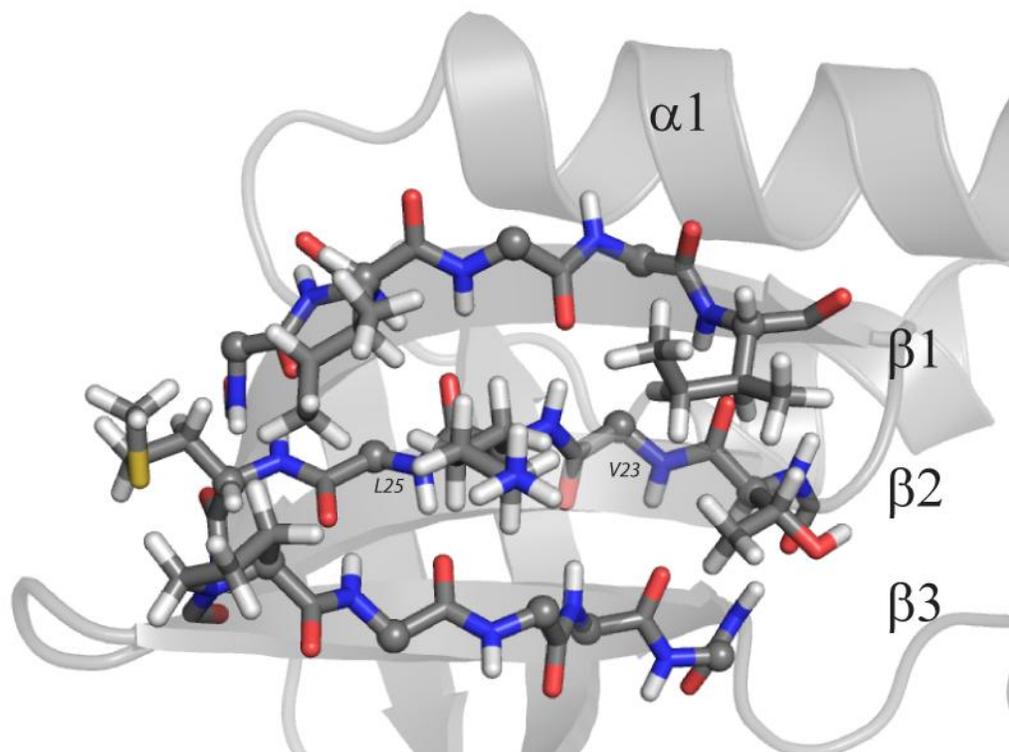
**Figure 5.3** Heterogeneous exchange in the SNase  $\beta$ -barrel. Identification scheme as described in Fig. 5.2.

Several amides -- Leu14 and Ala17 ( $\beta$ 1), and Thr33 ( $\beta$ 3) -- are fully exposed to bulk water. Gly20 ( $\beta$ 1) is bound to a water that is held by two side chains which, however, seem freely mobile. Protection factors are low, close to 10 in each case, indicating that solvent hydroxide has easy access to these amides but  $P_f$  values are larger than for hydrogens in unstructured segments, as in Fig. 5.1.

For Val23 the detailed motions that permit HX to occur can be inferred by the local context. The motion that breaks the protecting Val23 H-bond cannot involve most of its immediately neighboring residues as evidenced by their significantly higher  $P_f$  values. Given the high  $P_f$  of Arg35 and the rigid planarity of the peptide bond, the motion that exposes Val23 cannot involve the displacement of its H-bond acceptor. Similarly, the  $P_f$  values for Lys24, Phe34, and Leu36 indicate the Val23's HX competent state does not disturb the H-bonds that protect its immediate neighbors in sequence. This strongly

suggests a crankshaft motion of the Thr22/Val23 peptide group about its two neighboring  $\alpha$ -carbons, which would expose the Val23 amide to solvent and would insert the Thr22 carbonyl into the hydrophobic center of the  $\beta$ -barrel. The non-ideal interactions associated with this state help to explain why it is only populated  $10^{-5}$  of the time.

Why are local fluctuations similar to Val23 not seen throughout the  $\beta$ -barrel? The only remarkable characteristic about the Thr22/Val23 peptide group is that the carbonyl accepts two H-bonds. An identical situation can be seen for the Lys24/Leu25 peptide group, but Leu25 is much more protected and exchanges by way of a large unfolding reaction. These main chain similarities suggest that the difference depends upon their sidechains. Inspection reveals a pocket of large hydrophobic sidechains between Lys25 and solvent is a pocket of long hydrophobic sidechains with the charged end of the Lys25 sidechain acting as a polar cap (Fig. 5.4). In contrast, the surface near Val23 is relatively unencumbered by sidechains. A possible implication is that, even if the Lys24/Leu25 peptide group experienced a similar motion, the Leu25 amide would remain protected from catalyst.

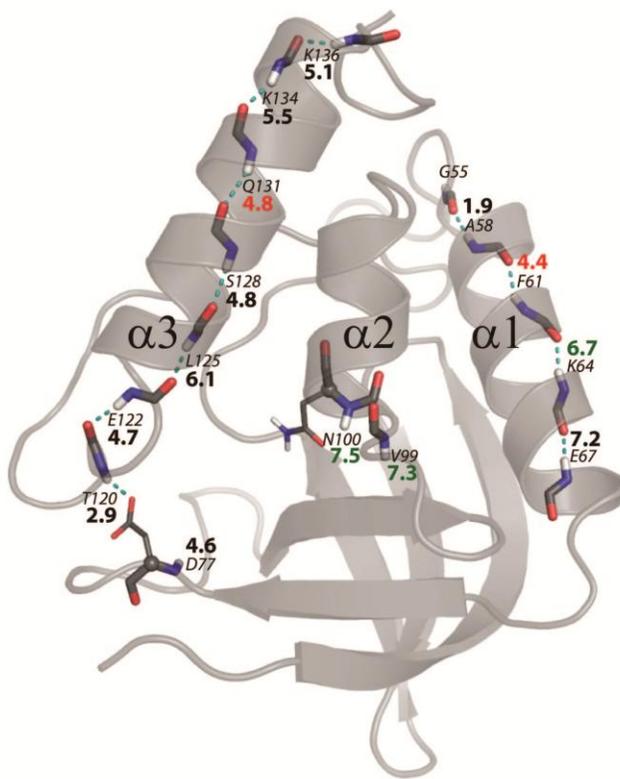


**Figure 5.4** Sidechains between residues Val23 and Leu25 and solvent.

### 3.2.3 Varied Exchange in Three $\alpha$ Helices

The three  $\alpha$ -helices of SNase each exhibit different HX patterns. Fig. 5.5 displays each of these helices in the context of the entire protein. In  $\alpha 1$ ,  $\log P_f$  values seen for the H-bonded amides from residue 58 to 69 are 1.9, 3.1, nd, 4.4L, 5.7U, 6.7, 6.7U, 6.6U, 7.0U, 7.2, 2.4, 4.0L, 1.7, with known local fluctuations (L) and known unfolding (U) noted (15, 16). H-bonded amides through the middle of the helical length, on both the aqueous and inner surfaces, show about the same  $P_f$  and a large denaturant dependence, indicating a cooperative unfolding reaction. The more N-terminal helical residues presumably participate in the same cooperative unfolding but they exchange through

some more facile opening reaction which produces a fraying progression. Interestingly, denaturant dependence is low, apparently due to small surface exposure in the fray. No fraying pattern appears at the  $\alpha 1$  C-terminus; it is stabilized by a capping interaction (see Fig. 5.6).



**Figure 5.5** HX behavior in the SNase  $\alpha$ -helices. The entire SNase molecule is shown. Identification scheme as described in Fig. 5.2

In helix  $\alpha 2$ , residues 102 to 109 exchange by way of a large subglobal unfolding reaction, with high denaturant dependence but with measured  $\Delta G_{op}$  and calculated EX1  $k_{op}$  rates smaller than for the global unfolding. Interestingly, no local fluctuational exchange is seen, unlike the other helices, perhaps because the  $\alpha 2$  helix is packed between the other two helices. The N-terminal residues 99 and 100 intersect  $\beta 5$  (see also

Fig. 5.2) and appear to exchange only with the global unfolding (16), as for  $\beta 5$  and its neighbors. One expects  $\alpha 2$  residues 99 and 100 will participate in the same cooperative subglobal unfolding as the rest of  $\alpha 2$ , but they instead appear to remain protected through the helix unfolding and exchange only when global unfolding occurs.

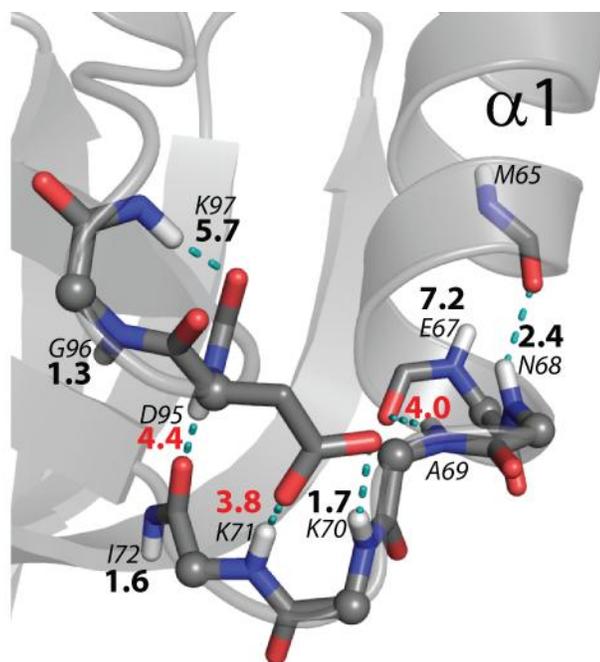
In contrast, the H-bonded amides in helix  $\alpha 3$  show disparate rates, and do not exchange by a cooperative helix unfolding. Log  $P_f$  observed for residues Glu122 to Leu137 are 4.7, 0.2, 1.3, 6.1, 6.3L, 4.2, 4.8L, 6.1, 4.8, 6.5, 6.2, 5.5, 5.0L, 5.0, 5.2. The  $\alpha 3$  residues show a mildly oscillating HX rate pattern that rises and falls with a rough helical periodicity apparently as the helical surface goes from buried (with  $\alpha 2$ ) to solvent exposed and back again. The variability in  $P_f$  values is an indication that local fluctuations predominate. Arg126 is one of the most highly protected amides in SNase ( $\log P_f = 6.3$ ), yet its lack of denaturant dependence demonstrates it exchanges by way of a local fluctuation. The suggestion is that  $\alpha 3$  unfolds as a cooperative unit less than  $10^{-6.3}$  of the time. In this case local fluctuations provide much faster HX pathways.

A general principle for  $\alpha$ -helices seems to be that the presence or absence of local fluctuations is determined by tertiary interactions. It should be noted that the presence of local fluctuations does not prevent helices from cooperative unfolding, but rather interferes with measuring this behavior by HX. The cooperative unfolding behavior seen for  $\alpha 1$  and  $\alpha 2$  is likely common to all helices including  $\alpha 3$ , though these reactions may occur infrequently for some helices due to their high stability.

### 3.2.4 Loops and Water

In SNase several loops serve to connect pieces of secondary structure and some loops are marginally stable. The backbone  $\phi$  and  $\psi$  angles in loops occupy the same regions of Ramachandran space found in helices and sheets. The difference is that in loops these angles do not appear in regularly repeating patterns. Yet the H-bond requirements for backbone amides and carbonyls remain the same. For some residues these requirements can be fulfilled by other backbone atoms, but often sidechains and even water molecules are required. For a protein the size of SNase this can result in one or more water molecules being encapsulated within the protein. Several examples of HX behavior in loops, including those involving internal water molecules, are discussed in this section.

Fig. 5.6 displays HX behavior in the short  $\alpha 1/\beta 4$  loop. Three neighboring amides, two within the loop (Ala69 and Lys71) and one from an adjoining loop (Asp95), with varied H-bonding (backbone, side chain), have similar protection factors ( $\log P_f = 4.0, 3.8, 4.4$ ) suggesting that they are exposed together by a concerted loop unfolding. The concerted unfolding of  $\Omega$ -loops has been noted before (24) and may reflect a common dynamic loop mode. The loop residues exhibit low denaturant dependence indicating that HX occurs without exposing much additional surface area.

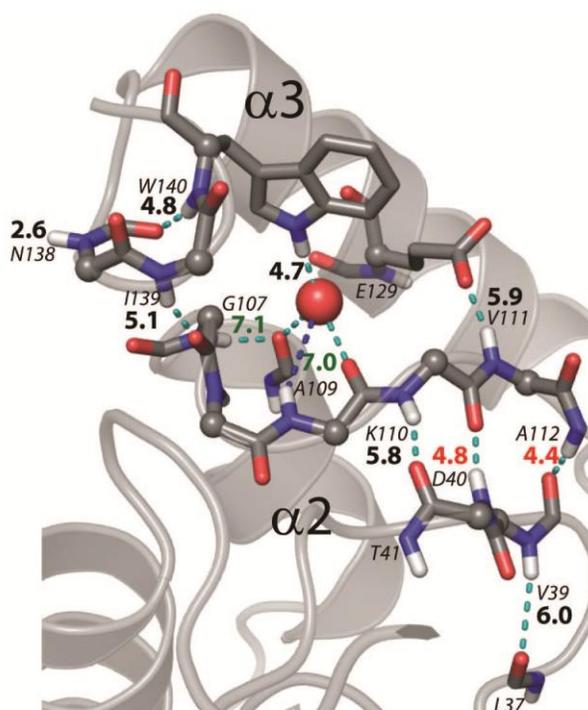


**Figure 5.6** Cooperative motion in the  $\alpha 1/\beta 4$  loop. Identification scheme as described in Fig. 5.2

The amide NH of the intervening loop residue, Lys70, exchanges much faster ( $\log P_f = 1.7$ ), indicating a local motion in which the Ala69/Lys70 peptide group is separately exposed, apparently by a crankshaft rotation about the two neighboring alpha carbons or possibly by reorientation of the Asp95 sidechain. Other alternatives are ruled out by the slow exchange of the Lys71 and Lys97 amide NHs, and the rigidly connected Ala94/Asp95 peptide unit. These results also show that the Asp95 NH is exposed to exchange by the movement of its acceptor in the loop unfolding.

Fig. 5.7 shows residue interactions in the  $\alpha 2/\alpha 3$  loop, the  $\alpha 1/\beta 3$  loop, and the near C-terminal region (residues 141 to 149 are unstructured). A defined water molecule is held in place by interactions with  $\alpha 3$  and the  $\alpha 2/\alpha 3$  loop. It is held in place by water donating H-bonds to two main chain carbonyls and accepting two others from the indole

ring NH of Trp140 and the amide NH of Ala109 (C-termini of  $\alpha 3$  and  $\alpha 2$ , respectively). The Ala109 H-bond to the water is non-ideal, 4.4 Å N to O, but there is no other available acceptor. The local context indicates that neither of these hydrogens exchange by way of water diffusing in and out of the protein. Ala109 is highly protected ( $\log P_f = 7.0$ ) and ultimately exchanges as part of a large unfolding unit, indicated by common equilibrium and kinetic parameters for the entire  $\alpha 2$  helix. The indole ring NH is less protected and may exchange through the same local motion that exposes its similarly protected backbone ( $\log P_f$  of 4.7 and 4.8).



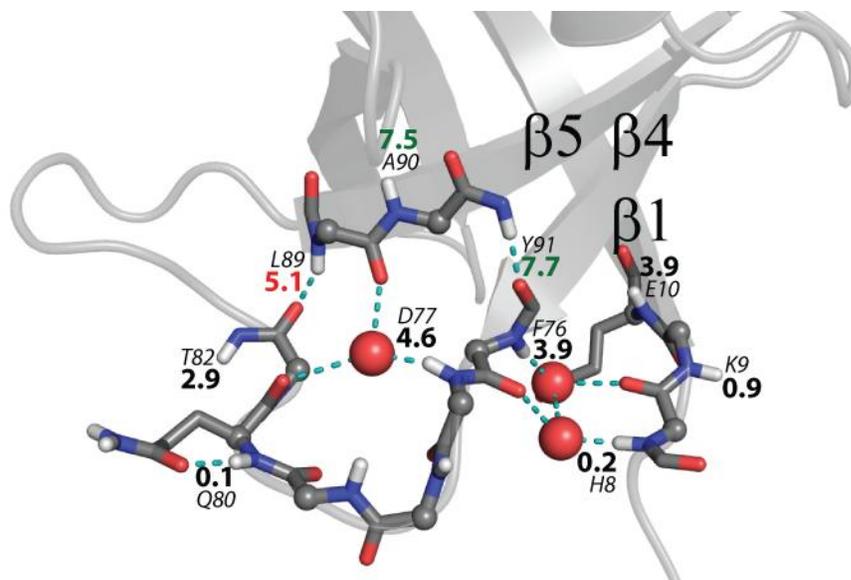
**Figure 5.7** An internal water molecule and several loops. Identification scheme as described in Fig. 5.2. The high  $\log P_f$  associated with Asn138 is explained by the NMR structure (Fig. 5.10).

The HX pattern in the  $\alpha 1/\beta 3$  loop is suggestive of cooperative motions. Residues Asp40 and Ala112 are similarly protected ( $\log P_f = 4.8$  and 4.4). They are surrounded by

another set of residues (Val39, Lys110, Val111) with similar protection ( $\log P_f = 6.0, 5.8, 5.9$ ). If this pattern is the result of cooperative motion then it is not the type typically associated with unfolding. The motion suggested is one in which the two strands move apart while surrounding interactions remain intact. Alternatively, this pattern may arise by chance from unrelated, independent motions.

Asn138 is unprotected in the crystal structure, but has a significant protection factor ( $\log P_f 2.6$ ). This is explained by differences in the NMR structure, which is discussed below (Fig. 5.10).

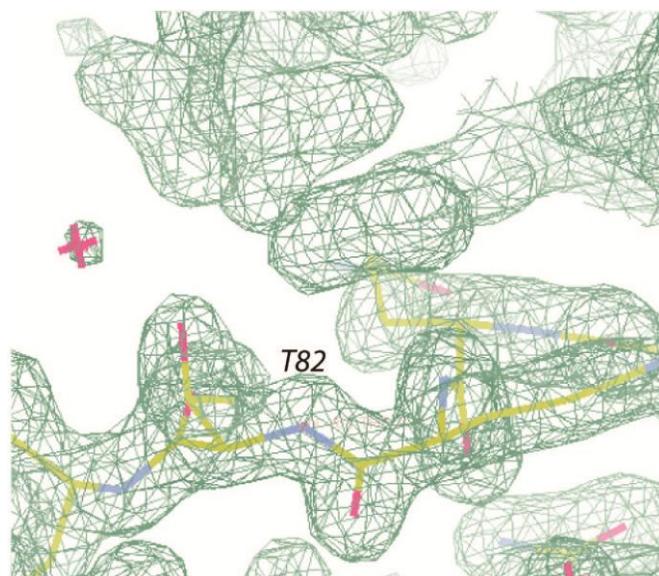
Fig. 5.8 shows the long  $\beta_4/\beta_5$  loop and its interaction with the near N-terminal segment. A number of interactions are mediated by water molecules. His8, proximal to the unstructured N-terminal segment (residues 1 to 5), H-bonds to a defined water, which is in direct contact with bulk, and protection is low ( $P_f < 2$ ). The Phe76 water acceptor is strongly protecting even though it is only one water removed from bulk. The Asp77 water acceptor is cut off from bulk water and offers strong protection. Both donate two H-bonds to the protein. Thus their replacement by hydroxide catalyst is disfavored. It would not fulfill the local H-bonding requirements and would require some structural reorganization.



**Figure 5.8** Loops with internal waters. Identification scheme as described in Fig. 5.2

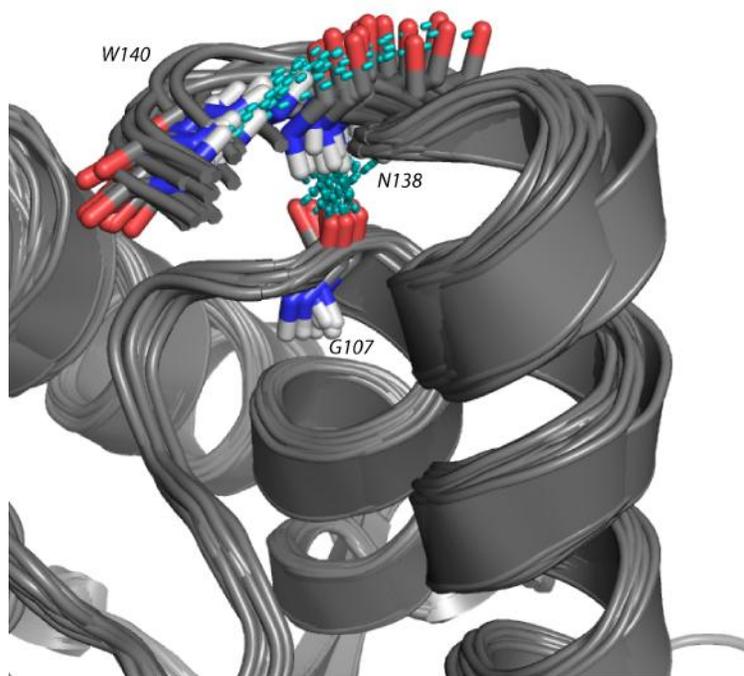
### 3.2.5 Resolving Structural Ambiguities

HX rates may help to resolve structural ambiguities. In the SNase x-ray structure, the Gln80 amide appears to be protected by its own side chain but solution HX finds it to be the least protected of all side chain-protected amides ( $P_f \sim 1$ ; Fig. 5.1). The x-ray structure provides no H-bond acceptor for the neighboring Thr82, but solution HX data shows that it is the most protected ( $P_f \sim 1000$ ) among exposed amides. Examination of the X-ray data reveals that Thr82 interacts with the side chain of a neighboring protein in the crystal lattice (Fig. 5.9). Both anomalies could be reconciled by the suggestion that the Gln80 side chain actually protects Thr82 in solution. The NMR structure provides no NOE information in this case.



**Figure 5.9** Electron density map surrounding Thr82 shows that this residue is involved in a lattice contact with a sidechain from a neighboring molecule.

Another ambiguous case involves Asn138, which appears to be H-bonded to external solvent in the crystal structure (Fig. 5.7). The measured  $\log P_f$  of 2.6 is large compared to similar cases (Fig. 5.1) and seems more consistent with an amide that is engaged in an intra-molecular H-bond. Examination of the NMR structures (2KQ3) reveal Asn138 to be oriented such that it H-bonds to the Gln106 carbonyl in all 20 structures (Fig. 5.10).



**Figure 5.10** The protection factor of Asn138 ( $\log P_f = 2.6$ ) is consistent with the SNase NMR structure (PDB 2KQ3). This structure reveals that Asn138 is H-bonded to the carbonyl of Gln106. This reorientation relative to the crystal structure (Fig. 5.7) still permits Trp140 to H-bond to the carbonyl of Leu137, which is linked to Asn138 through the peptide bond.

NMR models proved useful for resolving apparent conflicts between the crystal structure and HX data. However, it should be noted that the NMR structures were not without limitations of their own. This is especially true when water molecules are incorporated into the protein structure. Waters were not modeled into the NMR structures. Without these waters the force fields used to refine NMR models distort the structures in order to satisfy H-bond requirements. In the case of SNase neither X-ray nor NMR structures are completely consistent with H-bond requirements and HX measurements. Where discrepancies between X-ray and NMR structures exist, HX data can be used to determine which model is more consistent with the aqueous protein.

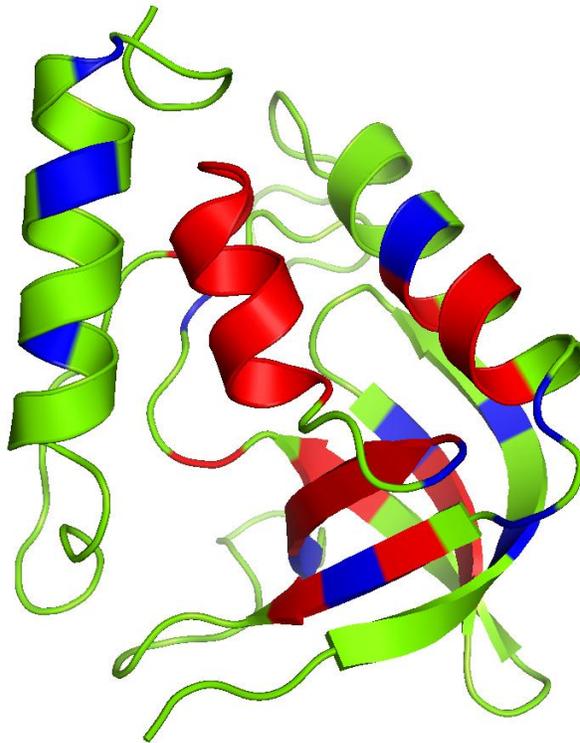
### 3.2.6 Local Fluctuations

Residues with HX rates that are insensitive to low concentrations of denaturant have been said to exchange by way of local fluctuations (18). The lack of denaturant dependence implies that little additional surface area is exposed during the exchange reaction. Therefore it has long been assumed that such residues exchange by way of very local, independent motions likely limited to a single residue. Examination of HX in SNase provides the most detailed analysis of local fluctuations to date (14).

Several examples provide insight into the size of local fluctuations. Val23 (Fig. 5.3) demonstrates that local fluctuations can arise from the independent motion of a single residue, as has been suggested (25). Apparently cooperative motions are observed in other cases. Sequentially increasing  $P_f$  values were observed in the N-terminal region of  $\alpha 1$  (Fig. 5.5), implying HX occurs by helical fraying. As the  $P_f$  values increase HX rates change from being insensitive to low denaturant concentrations to being denaturant dependent, suggesting that a fray must be of sufficient size before denaturant dependence is detectable. Similarly, the  $\alpha 1/\beta 4$  loop appears to exchange cooperatively, even though denaturant dependence is not detectable (Fig. 5.6). From these observations we can deduce that denaturant dependence can be undetectable for cooperative unfolding reactions as long as they are sufficiently small (approx. 4 residues).

Some general conclusions can be drawn from the overall pattern of local fluctuations in SNase (Fig. 5.11). Local fluctuations are seen throughout SNase, but are found predominantly near the edges of secondary structure. This observation is consistent with a model in which the measured HX rate arises from a competition between local

fluctuations and larger unfolding reactions. Interactions that restrict local motions will bias a residue toward HX by way of unfolding. Restrictive contacts are more common for residues surrounded by stable secondary structure. It should be noted that proximity to the edge of secondary structure does not necessarily correspond to nearness to the protein surface, which does not strongly correlate with  $P_f$  (Fig. 4.2). This point is best illustrated by fraying in  $\alpha 1$  where  $P_f$  increases with the number of residues from the N-terminus while distance to surface oscillates with residue number.



**Figure 5.11** Exchange mechanisms displayed on the structure of SNase. Colors indicate residues known to exchange by large unfolding (red), by local fluctuations (blue), and residues for which the HX mechanism is unknown (green).

Algorithms used to predict HX rates were discussed in Chapter 4. The H-COREX algorithm (26) calculates the probability of unfolding for each residue based on the

energy of all states in which a fragment containing that residue is unfolded relative to the energy of all states in which that residue is folded. This information is then translated into  $P_f$  by considering a residue to be HX competent if either the residue or its H-bond acceptor is unfolded.

It has been argued that the COREX-based method should account for all HX rates including those arising from local fluctuations (27). The suggestion is that what is seen as local fluctuations actually arise from the combination of many larger unfolding of various sizes. The lack of observed denaturant dependence is explained using the idea that the fully structured protein actually makes up a very small percentage of protein population. The argument is that the population is dominated by proteins that are partially unfolded, but, since experiments typically measure the average structure, these partial unfoldings are masked.

However, this argument fails to account for the observation that many residues that exchange by local fluctuations have significantly lower  $P_f$  values than their immediate neighbors. The COREX algorithm assumes that all unfolding reactions include multiple consecutive residues. Given this restriction, a residue will have the lowest  $P_f$  value relative to its two immediate neighbors when this residue is the only residue that is contained in two equally probable unfolding reactions, one including its upstream neighbors and the other including its downstream neighbors. In this case the central residue would have a  $P_f$  that was half that of its immediate neighbors. Yet in SNase many cases can be seen in which a single residue or a short segment of consecutive residues have  $P_f$  values at least two orders of magnitude lower than immediate neighbors (Table 5.1).

**Table 5.1** SNase protection factors. Abbreviations for acceptor types are unstructured (un), structured but exposed to solvent (ex), backbone carbonyl (bb), sidechain (sc) and internal water (wa).

Residue	Measured log P <sub>f</sub>	Exchange reaction	Acceptr	Residue	Measured log P <sub>f</sub>	Exchange reaction	Acceptr
Thr2	-1.1		un	Arg87	4.61		sc
Lys5	0.43		un	Gly88	4.91	local	bb
Leu7	0.38		ex	Leu89	5.08	local	bb
His8	0.19		ex	Ala90	7.49	denat EX1	bb
Lys9	0.88		ex	Tyr91	7.66	denat EX1	bb
Glu10	3.90		bb	Ile92	7.51	EX1	sc
Ala12	4.95		bb	Tyr93	7.53	EX1	bb
Thr13	4.37	local	bb	Ala94	7.64	denat EX1	bb
Leu14	1.37		ex	Asp95	4.42	local	bb
Ile15	5.04		bb	Gly96	1.26		ex
Lys16	4.19	local	bb	Lys97	5.74		bb
Ala17	1.01		ex	Met98	0.28		ex
Asp19	4.16		bb	Val99	7.32	denat EX1	bb
Gly20	1.33		wa	Asn100	7.48	denat EX1	sc
Asp21	4.60		sc	Glu101	7.60	EX1	sc
Thr22	4.68		bb	Ala102	7.13	EX1	bb
Val23	4.99	local	bb	Leu103	7.36	EX1	bb
Lys24	6.71	denat	bb	Val104	7.23	EX1	bb
Leu25	7.55	denat EX1	bb	Arg105	7.17	EX1	bb
Met26	6.86	denat	bb	Gln106	7.38	denat EX1	bb
Tyr27	4.60		bb	Gly107	7.12	denat EX1	bb
Lys28	0.49		ex	Leu108	6.57	denat EX1	bb
Gly29	1.17		ex	Ala109	7.04	EX1	wa
Gln30	5.17		bb	Lys110	5.77		bb
Met32	6.96		bb	Val111	5.90		sc
Thr33	1.09		ex	Ala112	4.38	local	bb
Phe34	7.48	denat	bb	Asn119	1.70		sc
Arg35	7.62	EX1	bb	Thr120	2.89		sc
Leu36	6.37		bb	Glu122	4.68		bb
Leu37	7.30	denat EX1	bb	Gln123	0.20		ex
Val39	5.95		bb	Leu124	1.25		ex
Asp40	4.77	local	bb	Leu125	6.08		bb
Ala58	1.92		bb	Arg126	6.29	local	bb
Ser59	3.08		bb	Lys127	4.24		bb
Phe61	4.44	local	bb	Ser128	4.76		bb
Thr62	5.73	denat	bb	Ala130	6.05		bb
Lys63	6.70		bb	Gln131	4.79	local	bb
Lys64	6.73	denat	bb	Ala132	6.52		bb
Met65	6.57	denat	bb	Lys133	6.21		bb
Val66	6.97	denat	bb	Lys134	5.52		bb
Glu67	7.20		bb	Glu135	5.02	local	bb
Asn68	2.35		bb	Lys136	5.06		bb
Ala69	3.95	local	bb	Leu137	5.23		bb
Lys70	1.70		sc	Asn138	2.57		ex

Residue	Measured log P <sub>f</sub>	Exchange reaction	Acceptr	Residue	Measured log P <sub>f</sub>	Exchange reaction	Acceptr
Lys71	3.79	local	sc	Ile139	5.05		bb
Ile72	1.58		ex	Trp140	4.80		bb
Glu73	7.55	denat EX1	bb	Ser141	4.50		bb
Val74	5.78	local	bb	Asn144	0.47		un
Glu75	7.65	EX1	bb	Ala145	0.09		un
Phe76	3.88		wa	Asp146	0.16		un
Asp77	4.61		wa	Ser147	-0.22		un
Gln80	0.06		sc	Gly148	0.32		un
Thr82	2.93		ex	Gln149	-0.35		un
Asp83	4.35		bb	Trp140 sc	4.68		wa
Gly86	2.75		bb				

### 3.3 Modeling HX behavior

In the previous chapter several models were used to predict SNase HX P<sub>f</sub> values. None of these predictions matched measured HX rates with high accuracy. The observations considered above suggest several factors that must be accounted for in an accurate HX model.

#### 3.3.1 Accuracy of k<sub>ch</sub>

HX rates were originally understood based on reaction Scheme 5.1 where k<sub>op</sub> and k<sub>cl</sub> are structural opening and reclosing rates, and k<sub>ch</sub> is the chemical rate expected for freely exposed hydrogens (10).



Whenever k<sub>cl</sub> is fast relative to k<sub>ch</sub> (EX2), P<sub>f</sub> can be described by eq. 5.1

$$P_f = k_{cl}/k_{op} + 1 = k_{ch} / k_{ex} \quad \text{Eq. 5.1}$$

This phenomenological description of HX rates is largely consistent with observations for SNase. However, this model fails to explain the relatively modest slowing observed for amides that are exposed to solvent in static regions of the protein (Fig. 5.1). For these residues the concept of opening and closing rates has little meaning so based on Scheme 5.1 one expects  $k_{ex}$  to equal  $k_{ch}$ . The fact that these rates don't match within experimental error suggests that  $k_{ch}$  rates for structured segments do not match  $k_{ch}$  rates calculated for a random coil (22). It may be that the amide  $pK_a$  is dependent upon backbone conformation (28) and sidechain (29) conformation, even though electrostatic calculations do not correlate with observed rates in SNase (Fig. 4.5; (6)).

The implications of incorrectly calculating  $k_{ch}$  extend beyond amides that are constitutively exposed to HX catalyst. Any time the actual  $k_{ch}$  does not match the predicted  $k_{ch}$  there is a danger of attributing the slowing to the equilibrium factor  $K_{op}$ . This could occur when an amide remains structure but is exposed to catalyst due to the displacement of its H-bond acceptor. Similarly, when residues in a short segment unfold they may be restricted to a subset of Ramachandran space and therefore not exchange as they would in a random coil. This uncertainty may account for the spread in  $P_f$  values observed for many cooperative unfolding units. Fortunately the differences in  $K_{op}$  between these unfolding units are often larger than the range of  $k_{ch}$  uncertainty (40-fold in SNase).

### 3.3.2 The HX Competent State

Scheme 5.1 provides a general framework for interpreting HX rates, but it is silent regarding the nature of the open and closed states. Figure 5.1 demonstrates that  $P_f$  is independent of H-bond acceptor type, which means that an amide cannot be considered to be in the open state on the basis that it is accessible by water. Observed pH dependence (Fig. 4.1) indicates that, at physiological pH, the open state can be considered any conformation in which the amide is accessible to hydroxide.

The subtle differences between water and hydroxide hydrogen bonding patterns have significant implications at the protein surface. Evidence suggests that hydroxide accepts four hydrogen bonds in a planar arrangement and cannot offer its one proton as an H-bond donor (30). Thus any structure that cannot accommodate this arrangement will preferentially exclude hydroxide.

### 3.3.3 Multiple Exchange Pathways

It is clear from the data presented here that HX rates result from structural distortions of various sizes ranging from a single peptide group (Fig. 5.5) to secondary structure units (Fig. 5.3) to global unfolding of the entire protein (Fig. 5.2). The measured HX rate will be dominated by the most favorable opening event. In principle, all distortions and their relative populations could be modeled by all atom simulations such as molecular dynamics calculations. In practice, the necessary computational time is prohibitive and the ensembles generated by such calculations do not always match reality. An alternative approach is to determine  $\Delta G$  for each distortion by calculating the

energy of the native structure relative to the hydroxide accessible state. Previous attempts at this approach have focused on either local fluctuations (12, 31) or larger unfolding reactions (32), but not both.

#### **4. Conclusions**

The present chapter describes a systematic examination of detailed highly resolved HX data in light of the detailed structure of the sizeable  $\alpha/\beta$  SNase protein. The results emphasize the importance of H-bonding for providing structural protection and of transient structural dynamics in separating the protecting H-bond by displacing the donor itself or the acceptor and exposing the hydrogen to solvent catalyst.

Some detailed insights emerge. A new complexity is found for unprotected hydrogens exposed to solvent on the surface of structured regions where  $k_{ch}$  can be slower than the random coil expectation by as much as 40-fold (Fig. 5.1). For structurally protected hydrogens, the degree of HX protection does not simply depend on surface exposure, depth of burial(6), or on the acceptor type which can be a main chain carbonyl, a side chain, or even a bound water. The degree of protection by each kind of acceptor covers the entire range of energetics. It depends on the interaction network that needs to be disrupted to make the hydrogen available for attack by solvent catalyst. Weakly supported surface-exposed side chains provide less protection than well supported buried ones, and the same is true of defined water molecules, but surface-exposed main chain interactions in regular secondary structure can provide strong protection.

Detailed exposure reactions can often be recognized by the HX patterns they produce. In the case of local fluctuation pathways, immediately neighboring hydrogens can experience very different HX rates. Individual hydrogens can exchange by uncooperative local fluctuations even when they are embedded within a large unfolding reaction, allowing individual amides, or small sets of amides in the case of edge fraying, to exchange more rapidly. Exposure can result from displacement of either the amide itself or its acceptor. Local fluctuations are not limited to surface-exposed sites. The connection of local fluctuational HX with a lack of denaturant dependence was supported.

Stable supporting interactions act to suppress local fluctuations and favor HX by unfolding reactions, including global and cooperative subglobal reactions. In these cases, contiguous sets of hydrogens are seen to exchange with the same protection factor (EX2 regime), and even with the same unfolding and refolding rates (EX1 regime), identifying the protein segments that participate in the cooperative unfolding. Subglobal unfolding reactions are seen to involve helices, loops, and parts of  $\beta$ -structure. Large unfolding reactions are not preferentially limited to deeply buried sites, although global unfolding is often taken to identify the protein “core”. The connection of high denaturant sensitivity with large unfolding was supported.

## 5. Bibliography

1. Chamberlain, A.K., and S. Marqusee. 2000. Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. In: Protein folding mechanisms. Academic Press. pp. 283-328.
2. Olofsson, A., M. Lindhagen- Persson, M. Vestling, A.E. Sauer- Eriksson, and A. Öhman. 2009. Quenched hydrogen/deuterium exchange NMR characterization of amyloid-  $\beta$  peptide aggregates formed in the presence of  $\text{Cu}^{2+}$  or  $\text{Zn}^{2+}$ . FEBS Journal. 276: 4051-4060.
3. Kato, H., J. Gruschus, R. Ghirlando, N. Tjandra, and Y. Bai. 2009. Characterization of the N-Terminal Tail Domain of Histone H3 in Condensed Nucleosome Arrays by Hydrogen Exchange and NMR. J. Am. Chem. Soc. 131: 15104-15105.
4. Mayne, L., Z.-Y. Kan, P. Sevugan Chetty, A. Ricciuti, B.T. Walters, and S.W. Englander. 2011. Many Overlapping Peptides for Protein Hydrogen Exchange Experiments by the Fragment Separation-Mass Spectrometry Method. Journal of The American Society for Mass Spectrometry. 22: 1898-1905.
5. Chalmers, M.J., S.A. Busby, B.D. Pascal, G.M. West, and P.R. Griffin. 2011. Differential hydrogen/deuterium exchange mass spectrometry analysis of proteinligand interactions. Expert Review of Proteomics. 8: 43-59.
6. Skinner, J.J., W.K. Lim, S. Bédard, B.E. Black, and S.W. Englander. 2011. Protein hydrogen exchange: testing current models. Biophys J. .

7. Truhlar, S.M.E., C.H. Croy, J.W. Torpey, J.R. Koeppe, and E.A. Komives. 2006. Solvent Accessibility of Protein Surfaces by Amide H/2H Exchange MALDI-TOF Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*. 17: 1490-1497.
8. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2008. A Billion-fold Range in Acidity for the Solvent-Exposed Amides of *Pyrococcus furiosus* Rubredoxin. *Biochemistry*. 47: 6178-6188.
9. Li, R., and C. Woodward. 1999. The Hydrogen Exchange Core and Protein Folding. *PRS*. 8: 1571-1590.
10. Linderstrøm-Lang, K.U., and J.A. Schellman. 1959. Protein structure and enzyme activity. In: *The Enzymes*. NY: Academic Press. pp. 443-510.
11. Milne, J.S., L. Mayne, H. Roder, A.J. Wand, and S.W. Englander. 1998. Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Science*. 7: 739-745.
12. Vendruscolo, M., E. Paci, C.M. Dobson, and M. Karplus. 2003. Rare Fluctuations of Native Proteins Sampled by Equilibrium Hydrogen Exchange. *J. Am. Chem. Soc.* 125: 15686-15687.
13. Hilser, V.J., and E. Freire. 1996. Structure-based Calculation of the Equilibrium Folding Pathway of Proteins. Correlation with Hydrogen Exchange Protection Factors. *Journal of Molecular Biology*. 262: 756-772.

14. Skinner, J.J., W.K. Lim, S. Bédard, B.E. Black, and S.W. Englander. 2011. Protein dynamics: The view from hydrogen exchange. *Proc Natl Acad Sci US*. .
15. Wrabl, J.O. 1999. Investigations of denatured state structure and m-value effects in staphylococcal nuclease. .
16. Bédard, S., L.C. Mayne, R.W. Peterson, A.J. Wand, and S.W. Englander. 2008. The Foldon Substructure of Staphylococcal Nuclease. *Journal of Molecular Biology*. 376: 1142-1154.
17. Englander, S.W., and N.R. Kallenbach. 1983. Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids. *Quarterly Reviews of Biophysics*. 16: 521-655.
18. Bai, Y., T. Sosnick, L. Mayne, and S. Englander. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science*. 269: 192 -197.
19. MacKerell, A., B. Brooks, C.L. Brooks, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998. CHARMM: the energy function and its parameterization with an overview of the program. In: Schleyer P, editor. *The encyclopedia of computational chemistry*. Chichester: Wiley. pp. 271-277.
20. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. *Acta Cryst D*. 60: 2126-2132.
21. Fleming, P.J., and G.D. Rose. 2005. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Science*. 14: 1911-1917.

22. Bai, Y., J.S. Milne, L. Mayne, and S.W. Englander. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins*. 17: 75-86.
23. Maki, K., T. Ikura, T. Hayano, N. Takahashi, and K. Kuwajima. 1999. Effects of proline mutations on the folding of staphylococcal nuclease. *Biochemistry*. 38: 2213-2223.
24. Krishna, M.M.G., Y. Lin, J.N. Rumbley, and S. Walter Englander. 2003. Cooperative Omega Loops in Cytochrome c: Role in Folding and Function. *Journal of Molecular Biology*. 331: 29-36.
25. Maity, H., W.K. Lim, J.N. Rumbley, and S.W. Englander. 2003. Protein hydrogen exchange mechanism: Local fluctuations. *Protein Science*. 12: 153-160.
26. Liu, T., D. Pantazatos, S. Li, Y. Hamuro, V.J. Hilser, and V.L. Woods. 2011. Quantitative Assessment of Protein Structural Models by Comparison of H/D Exchange MS Data with Exchange Behavior Accurately Predicted by DXCOREX. *Journal of The American Society for Mass Spectrometry*. .
27. Wooll, J.O., J.O. Wrabl, and V.J. Hilser. 2000. Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *Journal of Molecular Biology*. 301: 247-256.
28. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2009. Backbone conformational dependence of peptide acidity. *Biophysical Chemistry*. 141: 124-130.

29. Anderson, J.S., G. Hernández, and D.M. LeMaster. 2010. Sidechain conformational dependence of hydrogen exchange in model peptides. *Biophysical Chemistry*. 151: 61-70.
30. Marx, D., A. Chandra, and M.E. Tuckerman. 2010. Aqueous basic solutions: hydroxide solvation, structural diffusion, and comparison to the hydrated proton. *Chem. Rev.* 110: 2174-2216.
31. Best, R.B., and M. Vendruscolo. 2006. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure*. 14: 97-106.
32. Hilser, V.J. 2001. Modeling the Native State Ensemble. In: *Protein Structure, Stability, and Folding*. New Jersey: Humana Press. pp. 093-116.

## CHAPTER 6

### The Mad2 Conformational Transition

#### 1. Introduction

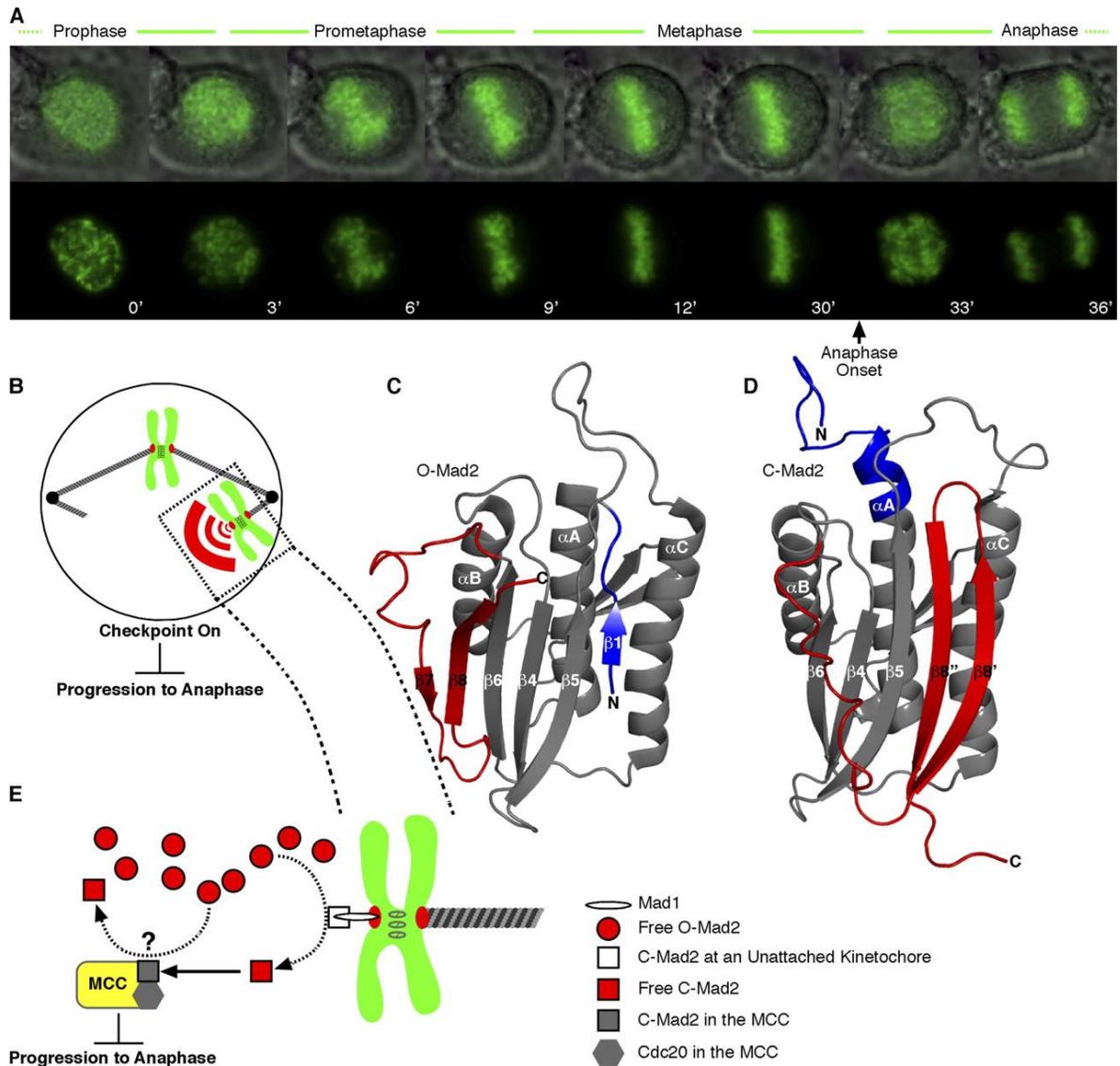
The metamorphic protein Mad2 acts as a molecular switch in the checkpoint mechanism that monitors proper chromosome attachment to spindle microtubules during cell division. The remarkably slow spontaneous rate of Mad2 switching between its checkpoint inactive and active forms is catalyzed onto a physiologically relevant time scale by a self–self interaction between its two forms, culminating in a large pool of active Mad2. Structural, biochemical, and cell biological advances suggest that the catalyzed conversion of Mad2 requires a major structural rearrangement that transits through a partially unfolded intermediate.

#### 1.1 Mad2 Function

The essential goal of mitosis is the equal distribution of sister chromatids into genetically identical daughter cells (1, 2). Chromosome segregation is directed by the centromere, a locus epigenetically defined by a specialized chromatin domain marked by nucleosomes in which the histone variant CENP-A (centromere protein A) replaces H3 (3). The kinetochore, an enormous protein assembly consisting of >80 known proteins, assembles upon the centromere of each chromatid and connects to microtubule-based fibers that extend from opposite poles of the mitotic spindle. Accurate kinetochore attachment to the spindle is monitored by a diffusible checkpoint signal termed the

mitotic checkpoint (also referred to as the spindle assembly checkpoint; (4, 5). The checkpoint inhibits mitosis, halting progression to anaphase until all chromosomes are aligned on the metaphase plate and every kinetochore is properly attached to the spindle (Fig. 6.1A).

On and off switching of the mitotic checkpoint must be fast and definitive because either a weak checkpoint or an asynchronous metaphase to anaphase transition leads to irreversible missegregation of one or more chromosomes. The checkpoint must be active upon entry into mitosis and sufficiently robust so that checkpoint activation is maintained if even a single kinetochore remains unattached to the spindle (Fig. 6.1B). After proper spindle attachment to all kinetochores, the checkpoint rapidly inactivates to allow for the destruction of mitotic targets (e.g., cyclin B and securin), which leads to synchronous chromosome separation and segregation. Inappropriate early inactivation of the checkpoint produces lethal chromosomal missegregation (6, 7). However, a functional mitotic checkpoint is required for tumor cell death resulting from treatment with microtubule toxins such as taxol that are widely used in the clinic (8). It is not clear how rapid checkpoint silencing occurs, though several pieces of evidence have emerged (9).



**Figure 6.1** The mitotic checkpoint ensures equal partitioning of chromosomes in anaphase. (A) A human tissue culture cell progressing through mitosis with time indicated in minutes. In the top row, chromosomes (green) are overlaid with a differential interference contrast image of the entire cell. Sister chromatids align at the metaphase plate early in mitosis and wait for ~20 min before chromatid separation in anaphase. Upon final chromosome alignment, the mitotic checkpoint signal decays, allowing the cell to enter anaphase and initiate simultaneous separation of sister chromatids. (B) The mitotic checkpoint signal, comprised in part by a diffusible pool of C-Mad2, emanates from kinetochores that have not yet properly engaged the microtubule-based spindle. A single unattached chromosome is sufficient to generate a checkpoint signal that arrests mitosis before anaphase. (C and D) Interconversion between inactive O-Mad2 (O-Mad2<sup>ANAC</sup>; PDB 1DUJ; (10) and checkpoint-active C-Mad2 (C-Mad2<sup>AN</sup>; PDB 1S2H; (11) involves a major secondary and tertiary structural reorganization of N-terminal (blue) and C-terminal (red) segments. (E) Unattached kinetochores contain the checkpoint protein Mad1, which recruits C-Mad2, providing a catalytic surface for the conversion of the soluble pool of inactive O-Mad2 to active C-Mad2. C-Mad2 is able to bind and inhibit Cdc20 within the MCC, halting progression to anaphase. The Cdc20–C-Mad2 complex may also act to catalyze conversion of the O-Mad2 pool, although this aspect of Mad2 signaling

remains controversial (4, 12). (B and E) Chromosomes are drawn in green with their kinetochores drawn in red.

The Mad2 protein is a centrally important regulator of the mitotic checkpoint machinery. Its activity is controlled by switching between its two different native conformations, open Mad2 (O-Mad2; Fig. 6.1C) and closed Mad2 (C-Mad2; Fig. 6.1D; (11, 13). Before checkpoint activation, freely diffusible monomeric Mad2 is thought to exist largely as O-Mad2, its inactive conformation, as is common for many regulatory proteins. Conformational conversion from inactive free O-Mad2 to active free C-Mad2 is catalyzed by a self–self interaction, namely by binding to the C-Mad2 subunit of a Mad1–C-Mad2 complex (10, 14, 15) anchored at kinetochores that are not yet properly engaged with a spindle (Fig. 6.1E; (16–18). Purified Mad1-bound Mad2 is known to catalyze the O-Mad2 to C-Mad2 transition in the absence of any other effector molecules (19). Recent research has supported this model in vivo (20). Newly converted Mad2 releases from the kinetochore and blocks premature progression to anaphase by binding to and deactivating Cdc20 in conjunction with other essential checkpoint proteins (including BubR1 kinase and Bub3) as part of a mitotic checkpoint complex (MCC; Fig. 6.1E; (21, 22). Although the checkpoint remains active, the inhibition of Cdc20 by C-Mad2 serves to restrain an E3 ubiquitin ligase known as the anaphase-promoting complex/cyclosome (APC; Fig. 6.1E; (4, 5). Once all kinetochores have properly attached to the spindle, Mad2 deactivates and releases Cdc20, allowing it to bind and activate the APC. APC–Cdc20 ubiquitinates several key mitotic substrates, including securin and cyclin B, leading to their removal by the proteasome and initiation of the metaphase to anaphase transition.

Mad2 differs strikingly from most regulatory proteins. Other proteins that change structure drastically, known as metamorphic proteins (23), require the selective stabilization of their intrinsically less stable active form through substrate binding, chemical modification, or environmental change. There is now some evidence that phosphorylation may play a role in Mad2 regulation (24), but this regulation is not strictly required for interconversion to take place. For Mad2, the structural changes from the inactive form to the active form are unusually large (Fig. 6.1C,D) and remarkably slow (11). Furthermore, Mad2 is found initially out of equilibrium in its inactive form (O-Mad2) even though its active form (C-Mad2) is the more stable conformation (11). Thus, checkpoint activation simply requires Mad2 to reach its equilibrium distribution. These properties raise key questions about the mechanism of mitotic checkpoint regulation. How do effector molecules modulate the rate of Mad2 interconversion? Could Mad2 regulation involve kinetic trapping in one of its two conformational states? Do transient conformational intermediates play a functional role?

## 1.2 Mad2 Structural Rearrangement

It has been reported that the spontaneous Mad2 activation reaction, O-Mad2<sup>R133A</sup> to C-Mad2<sup>R133A</sup>, proceeds with a lifetime of 9 hr; the reverse reaction is six fold slower (11)! These unusually slow interconversion rates stem from the magnitude of the structure change, which involves a complete rearrangement of the secondary and tertiary structure of ~60 out of 205 amino acids. In O-Mad2, the N-terminal segment forms a long loop and a short  $\beta$  strand ( $\beta$ 1) that connects to the static core (Fig. 6.1C). In the

transition to C-Mad2, this segment loses its  $\beta$  conformation and reconfigures, adding two more turns to the  $\alpha A$  helix (Fig. 6.1D; (11, 14, 25). The C terminus undergoes an even more dramatic change. In O-Mad2, the C-terminal segment forms strands  $\beta 7$  and  $\beta 8$  (and connecting loops) that dock onto the static core  $\beta 6$  strand. In C-Mad2, the whole segment moves to the opposite side of the major  $\beta$  sheet and forms two new strands,  $\beta 8'$  and  $\beta 8''$ , with a completely different hydrogen-bonding network. Overall, the transition to the C-Mad2 conformer relocates the N-terminal segment to make room for the incoming C-terminal segment, the displacement of which exposes an extended active site that is occluded in O-Mad2.

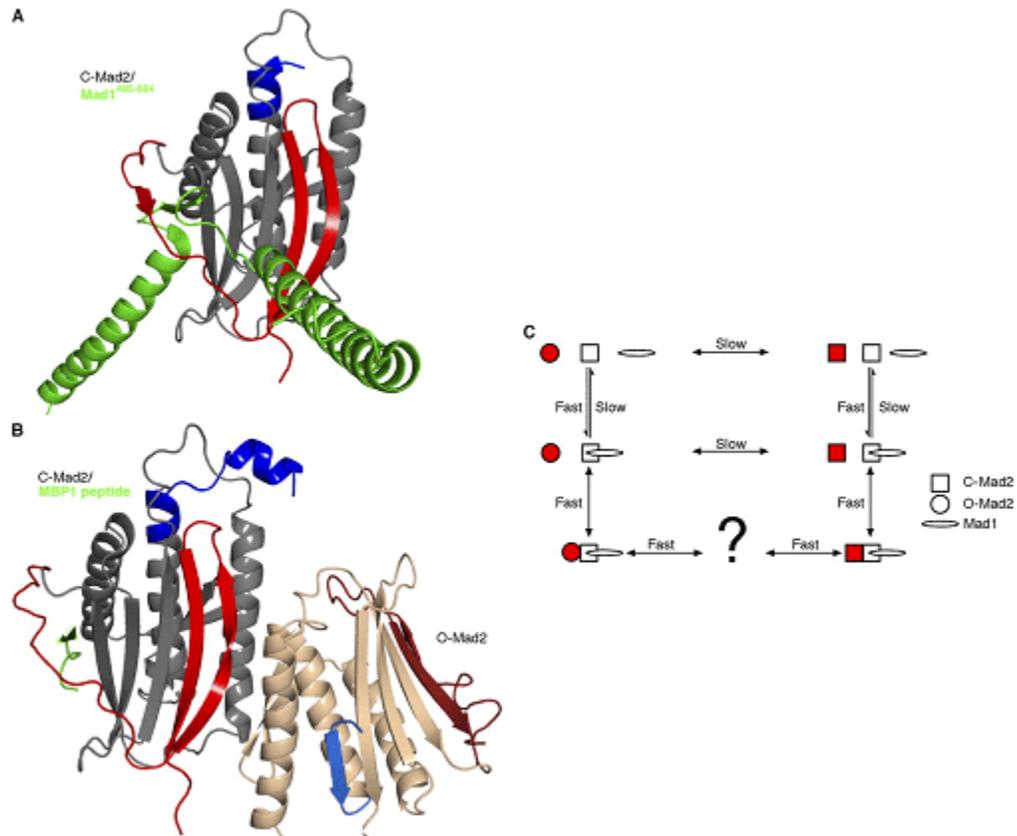
The Mad2 active site is tailored to interact with both its upstream activator Mad1 (Fig. 6.2A) and its downstream target Cdc20 (14, 25). Although Mad1 and Cdc20 appear to be otherwise unrelated, their Mad2-interacting regions are highly homologous and can be mimicked by a synthetic 12-residue consensus sequence peptide (Mad2 Binding Peptide 1 [MBP1]; Fig. 6.2B; (25). These partners bind by incorporating into the major Mad2  $\beta$  sheet as a single  $\beta$  strand, interacting with the  $\beta 6$  strand and a new  $\beta 7'$  strand that forms upon the binding (10, 14, 25). As shown in Figure 6.2A and B, they actually thread through the C-Mad2 sheet like links in a concatenated chain. Once Mad1 binds to Mad2, it forms a very stable complex with no detectable turnover in 4 min, as detected with purified components by FRAP (15), correlating with earlier cell-based FRAP measurements of the hyperstable pool of kinetochore-bound C-Mad2 that is presumably bound to Mad1 (26). Recently, it has been discovered that Mad2 also binds shugoshin-2 (Sgo2) in a similar manner as part of its function in regulating meiosis (27).

For the sake of simplicity, it is often stated that C-Mad2 itself is competent to bind Mad1 or Cdc20. However, in this binding reaction, Mad2 must expose a binding site, load its binding partner, and lock it in place. This implies that binding to either Mad1 or Cdc20 requires a substantial local rearrangement of Mad2 structure (19, 28). Although the possibility exists that Mad1 and Cdc20 may themselves unfold, thread through the Mad2-binding loop, and refold, partial unfolding of Mad2 seems more likely, especially since the O-Mad2 to C-Mad2 conversion appears to require a similar partial unfolding. In this view, a partially unfolded intermediate form of Mad2 would be required for Cdc20 binding and APC inhibition.

### **1.3 Conformational Switching Models**

On time scales relevant to cell biology, the great majority of biomolecules assume their equilibrium distribution among alternative conformations, and their rates of interconversion can be safely ignored. However, the spontaneous O-Mad2 to C-Mad2 conversion rate (many hours) is clearly inadequate for the rapid checkpoint activation required to inhibit anaphase immediately upon mitotic entry. The conversion of freely diffusible O-Mad2 is catalyzed by its self-interaction with the C-Mad2 partner of the kinetochore-bound Mad1–C-Mad2 complex (Fig. 6.2C). How is this catalytic event accomplished? Thermodynamic principles dictate that molecular binding partners promote structure change in allosteric proteins by binding more strongly to the favored form. Two common structure change models exist. Association may promote the structure change by sacrificing some of its binding energy to forcefully distort the protein

conformation (induced fit model), or selection may occur among preexisting dynamically cycling protein conformations by more strongly binding to and thereby trapping the preferred partner (conformational selection model).



**Figure 6.2** Mad2 containing complexes. (A and B) The displacement of the C-terminal segments in C-Mad2 exposes a new  $\beta$  sheet edge that can incorporate Mad1 (A; PDB 1GO4; (14)), Cdc20, or the synthetic peptide MBP1 (B; PDB 2V64; (28)) between newly exposed  $\beta_6$  and newly formed  $\beta_7'$ . In the crystal structure of the MBP1 bound C-Mad2–O-Mad2<sup>LL</sup> (B), asymmetrical dimerization occurs mainly through the unaltered core of Mad2 (gray to tan) but also includes the  $\beta_8'$  strand that is unique to C-Mad2 (coloring as in Fig. 6.1C). O-Mad2, the form undergoing conversion (tan), interacts with C-Mad2 only through its unchanging core. (C) A reaction scheme for Mad2 catalysis. Mad2 (red) represents the molecule undergoing conversion. Uncatalyzed Mad2 interconversion proceeds far more slowly (lifetime >9 h; (11)) than the duration of metaphase (~20 min). The Mad2 structural rearrangement is catalyzed by binding to the Mad1–C-Mad2 complex. In this reaction scheme, catalysis by induced fit would increase the forward O-Mad2→C-Mad2 rate, whereas the conformational selection of C-Mad2 would reduce the reverse O-Mad2←C-Mad2 rate. It is unknown whether Mad2 releases from the Mad1–C-Mad2 dimer as fully folded C-Mad2 or as a partially unfolded intermediate.

If the O-Mad2 to C-Mad2 conversion is catalyzed by induced fit, the structure of Mad2 in the catalytic complex should display the activation mechanism. Mapelli et al. crystallized a valid replica of the Mad1–C-Mad2–O-Mad2 catalytic complex (Fig. 6.2B; (28)). The O-Mad2 subunit was trapped in the open conformation by shortening the loop connecting the  $\beta 5$  strand to the  $\alpha C$  helix. The O-Mad2 loopless mutant (O-Mad2<sup>LL</sup>) was dimerized with a C-Mad2 molecule that was bound in turn to the synthetic activation peptide MBP1 to make a stable MBP1–C-Mad2–O-Mad2<sup>LL</sup> complex. The crystal structure of the complex reveals that the dimerization surface of O-Mad2, the form undergoing conformational change, only involves segments that are not substantially altered upon the O-Mad2 to C-Mad2 switch. Thus, it does not appear that the interaction would serve to forcefully induce the transition, providing evidence against an induced fit mechanism.

In the case of a conformational selection mechanism, one can expect that the catalyzing kinetochore-bound Mad1–C-Mad2 complex would favor the closed form of the substrate Mad2 molecule by binding to sites that are specific for C-Mad2. In fact, the C-Mad2–C-Mad2 complex does involve some of those sites (19). However, conformational selection alone does not increase the rate of conversion to the target structure; rather, it stabilizes the selected form by decreasing the reverse rate. Therefore, conformational selection of C-Mad2 can be effective only if conformer sampling (O-Mad2 to C-Mad2) is appropriately fast. If rapid conformational sampling occurred naturally, catalysis would not be necessary because the target C-Mad2 is actually the more stable form (11), ruling out conformational selection of C-Mad2 (Fig. 6.2C). In

summary, available crystal structures elegantly display the static Mad2 dimerization interface, but they do not suggest a mechanism to explain how the C-Mad2–O-Mad2 interaction catalyzes the O-Mad2 to C-Mad2 transition.

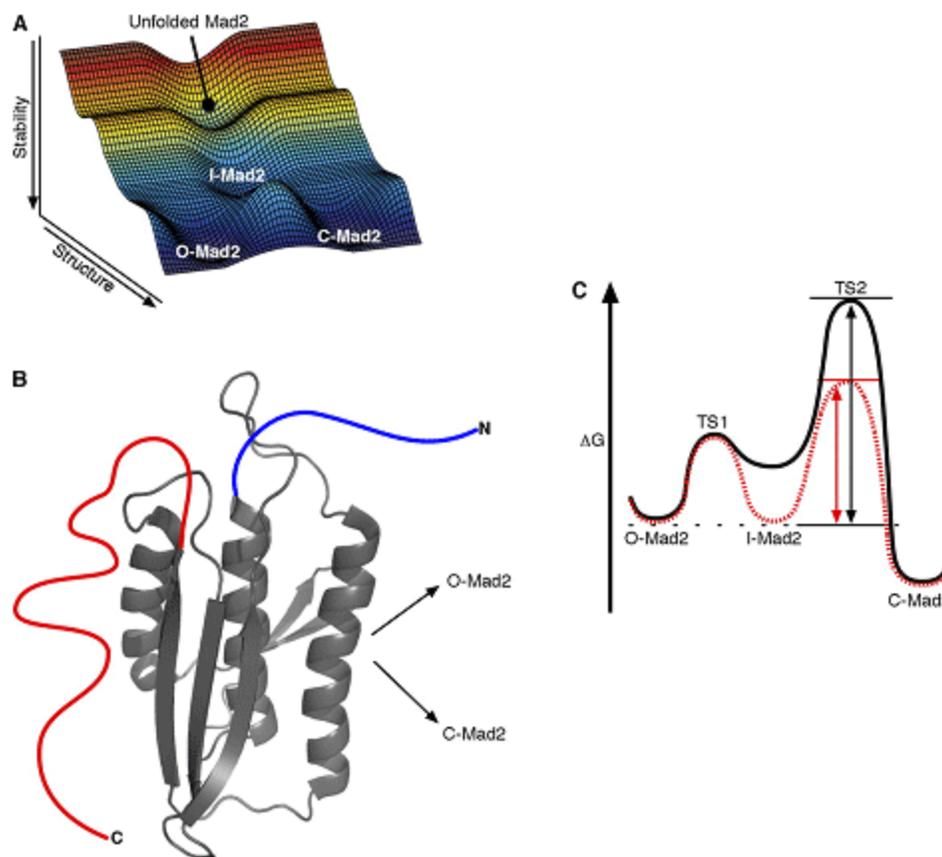
#### **1.4 An Unfolding Model for Mad2 Conformational Change**

It is hard to envision how any kind of straightforward conformational conversion (e.g., by a hinging or rigid body motion) could accomplish the major structural rearrangement between the two natively folded Mad2 forms. Rather, the conformational rearrangement is so extensive that it seems to require a significant unfolding of Mad2 to some transient high energy intermediate followed by kinetic partitioning between the two alternative forms upon refolding. Similarly, the fact that the main chain of Mad1 and Cdc20 actually threads through the major  $\beta$  sheet of C-Mad2 seems to require some transitional partially unfolded intermediate from which the C terminus could refold around the ligand.

A precedent for conformational change through partial unfolding can be found in the much smaller cytochrome *c* alkaline transition. At an elevated pH, the residue ligated to heme is switched from Met80 to the neighboring Lys79. Rather than simply shifting over by one amino acid residue, the transition involves the unfolding and refolding of a 15-residue loop that contains the two critical residues. The loop has been shown to unfold and refold repeatedly under native conditions as a cooperative unit. The stability of the loop determines the equilibrium between the Met80-liganded and Lys79-liganded forms (29), and the unfolding rate limits the kinetics of the transition (30). More generally,

recent work indicates that many proteins act as accretions of units that repeatedly unfold and refold under native conditions. It now appears that cooperative units can account for the steps in protein folding pathways, and, having reached the native state, their continuing dynamic unfolding and refolding behavior can be exploited to control ligand on and off rates (31) and even allosteric communication (32).

Can the emerging folding unit paradigm help to explain the Mad2 conformational switching mechanism? In addition to the aforementioned structural issues (e.g., massive rearrangement and threading), some other Mad2 folding-related observations are suggestive. Chemically denatured Mad2 spontaneously refolds into a nonequilibrium mixture of its two alternative conformations (C-Mad2 and O-Mad2 in a 2:1 ratio; (11). The implication is that the refolding pathway contains some intermediate stage from which Mad2 partitions into its two different stable forms (Fig. 6.3A and B). Spontaneous equilibration from this point is extremely slow. Thus, O-Mad2 is not itself a facile on-pathway precursor for generation of C-Mad2. Rather, O-Mad2 appears to transit to C-Mad2 by backtracking through a partially unfolded intermediate Mad2 (I-Mad2; Fig. 6.3A) and redistributing between O-Mad2 and C-Mad2 over several equilibration cycles.



**Figure 6.3** Mad2 unfolding and refolding considerations. (A) Free energy reaction landscape for Mad2 interconversion through a partially folded intermediate that lies on the folding pathway (created with Matlab version R2007a; The MathWorks, Inc.). When chemically denatured Mad2 is refolded, it initially reaches a nonequilibrium O-Mad2–C-Mad2 mixture, suggesting that the folding pathway to reach either form passes through a common intermediate and kinetically partitions rather than passing through one form on the way to the other. We suggest that the catalyzed interconversion seems likely, on this and other grounds, to pass back through the same partially unfolded intermediate. (B) A notional structure for a Mad2-folding intermediate showing the common (gray) and variable (colored as in Fig. 6.1C) segments. (C) Catalysis through intermediate stabilization. The conversion reaction of Mad2 is drawn with (red dashed line) or without (black solid line) dimerization with the kinetochore-bound C-Mad2–Mad1 complex. The measured C-Mad2/O-Mad2 equilibrium ratio is 8:1 (11), indicating that C-Mad2 is  $\sim 1$  kcal/mol more stable than O-Mad2. Conformational selection of I-Mad2 would equally stabilize I-Mad2 and the TS2 transition barrier relative to O-Mad2, effectively increasing the O-Mad2→C-Mad2 rate even though the energy difference between I-Mad2 and TS2 remains unchanged. The dashed black line indicates the energy state of O-Mad2, the black arrow (left) indicates increasing energy, the double-headed black arrow indicates the energy difference between O-Mad2 and TS2 without dimerization, and the double-headed red arrow indicates the energy difference between O-Mad2 and TS2 with dimerization.

How can an unfolding-dependent binding model promote the rate of the Mad2 conformational transition? As noted before, selective binding to C-Mad2 itself would not be helpful. Instead, the Mad1–C-Mad2 complex needs only to selectively stabilize a partially unfolded intermediate on the O-Mad2 side of the rate-limiting transition barrier, such as the hypothetical I-Mad2 in Figure 6.3C. The stabilization of I-Mad2 would equally stabilize the rate-limiting transition state relative to O-Mad2 and therefore increase the rate of O-Mad2 to C-Mad2 (it should be noted that I-Mad2 may be, but is not necessarily, the same as the intermediate that binds Mad1 and Cdc20 discussed above in section 1.2). Consistent with this view is the observation that Mad1–C-Mad2 catalyzes the O-Mad2 to Cdc20–C-Mad2 transition but not the C-Mad2 to Cdc20–C-Mad2 transition (33).

Unfortunately, a Mad1–C-Mad2–I-Mad2 structure is not likely to be solved by X-ray crystallography because partially unfolded and dynamically interconverting structures are not conducive to crystal formation. Available crystal structures of pertinent dimers used Mad2 variants that would prevent I-Mad2 formation. The Mad2<sup>LL</sup> mutant used to obtain MBP1–C-Mad2–O-Mad2<sup>LL</sup> crystals prevents O-Mad2 from switching into the C-Mad2 conformation by restricting the conformational search space of the N terminus (28). The L13A mutation used to obtain C-Mad2–C-Mad2 crystals stabilizes the native closed conformation so that the alternative O-Mad2 or I-Mad2 forms would not significantly populate (19). The structure of I-Mad2 will have to be studied by methods more applicable to dynamic systems. Unit dependent unfolding behavior in other proteins has so far been studied successfully, not by static crystallography but by dynamic

hydrogen exchange (31). We applied this and other methods to begin unraveling the mysteries underlying Mad2 interconversion.

## 2. Materials and Methods

### 2.1 Mutations

Several mutants were created in order to alter or investigate Mad2 function (Table 6.1). A PreScission protease recognition site was also introduced between a His<sub>6</sub> tag and the Mad2 N-terminus.

**Table 6.1** Mad2 mutants used in this study.

Mutant	Description	Effect	Reference	Notes
R133A		Dimerization reduced	(11)	
RQ	R133E/Q134A	Dimerization deficient	(34)	Fast conversion
RF	R133A/F141A	Dimerization deficient	(34)	
ΔC	Last 10 residues truncated	Cannot form C-Mad2	(10)	
L3P		None	This study	Intended to inhibit oligomerization
L3R		None	This study	Intended to inhibit oligomerization
Q4P		None	This study	Intended to inhibit oligomerization
Loopless	T109 to R117 removed	Locked in O-Mad2	(28)	
C79S		None	This study	Intended to inhibit sample degradation
C106S		None	This study	Intended to inhibit sample degradation

All mutations were introduced using the QuickChange PCR protocol developed by Stratagene. This method was done by building a solution containing 39.5  $\mu\text{L}$  H<sub>2</sub>O, 5  $\mu\text{L}$  10X Pfu Buffer, 2  $\mu\text{L}$  25X dNTPs, 1.25  $\mu\text{L}$  forward primer (100 ng/ $\mu\text{L}$ ), 1.25  $\mu\text{L}$  reverse primer (100 ng/ $\mu\text{L}$ ), 1  $\mu\text{L}$  parent DNA (~35 ng/ $\mu\text{L}$ ) and 1  $\mu\text{L}$  Pfu Turbo polymerase. This solution was incubated at 95 °C for 2 min before beginning 18 cycles of 95 °C for 30 s, 55 °C for 1 min, and 68 °C for 12 min. After amplification finished 1  $\mu\text{L}$  DpnI was added to the solution to digest the methylated parent DNA. The new DNA was then used to transform ultra-competent gold, which further amplified DNA production.

A colony was then chosen for further growth. DNA was purified by midi-prep. A sample was sent for sequencing and the remainder was stored at -20 °C. Glycerol stocks were made from the ultra-competent cells and stored at -80 °C.

## **2.2 Protein Purification**

The existence of two conformers adds an additional complication to the purification of Mad2. It has been reported these conformers can be separated by anion exchange and that interconversion can be effectively halted by storing the protein at 4 °C (11, 35). Therefore it is crucial to keep temperatures as low as possible throughout purification. Also important is the removal of purification tags which can interfere with interconversion.

### 2.2.1 Cell Growth

Many conditions were attempted for both standard and isotopically labeled Mad2. What follows is the optimized conditions for Mad2<sup>RQ</sup> and Mad2<sup>RF</sup> mutants.

2XYT media was used for standard growths and M9 media for <sup>15</sup>N and <sup>13</sup>C enriched growths. M9 media was made by dissolving 6 g/L Na<sub>2</sub>HPO<sub>4</sub>, 3 g/L KH<sub>2</sub>PO<sub>4</sub>, 0.5 g/L NaCl into 1 L H<sub>2</sub>O. This was brought to pH 7.5 and autoclaved. Additional minerals were sterile filtered using a 0.22 µm filter and added to the media, resulting in the following final concentrations: 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 15 mg/L ZnSO<sub>4</sub>, 15 mg/L FeSO<sub>4</sub>, and 1.25 g/L NH<sub>4</sub>SO<sub>4</sub>. Glucose was added to a final concentration of 2.5 g/L for <sup>13</sup>C labeled growths and 4 g/L glucose for unlabeled growths. The antibiotics Amp and Cam were added to prevent bacterial contamination and loss of plasmid.

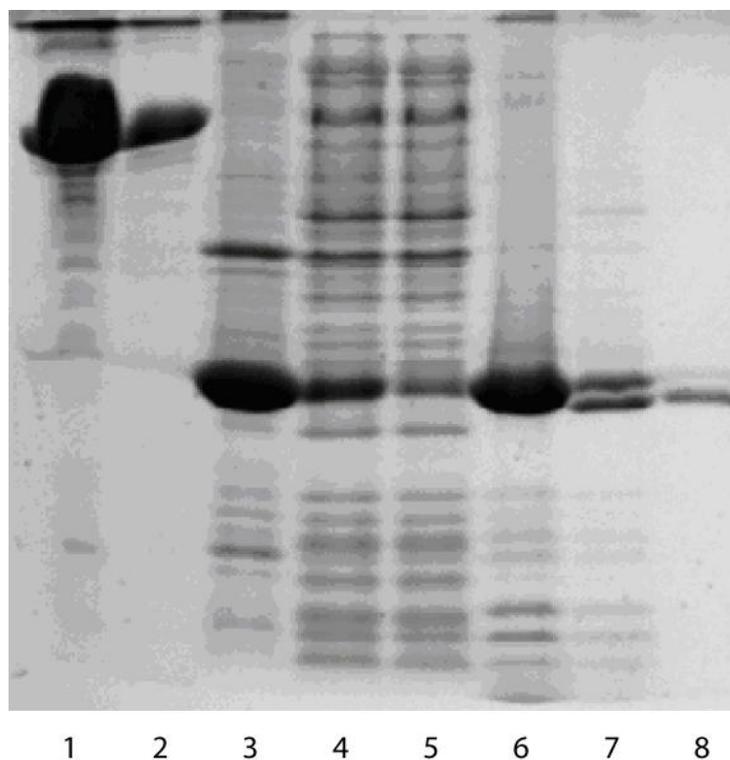
The Mad2 plasmid was transformed into Rosetta pLysS cells for 2XYT growths. The pLysS construct significantly slowed cell growth and greatly reduced yield when minimal media was used. This is likely due to the pLysS construct causing cells to lyse during the centrifuge step that is required when using minimal media. Therefore, Rosetta cells without the pLysS construct were used for M9 growths. Transformed cells were grown overnight at 37 °C in small batches of 2XYT media for seeding 1L growths of 2XYT or in LB for seeding 1L growths of M9. For M9 media growths cells were spun down (10 min, 4000 g) and the media was removed prior to seeding.

Cells were added to the final growth media until they reached an OD<sub>600</sub> of 0.1. The cells were incubated at 37 °C until they reached an OD<sub>600</sub> between 0.8 and 1. The temperature was then reduced to 25 °C and 1 mM IPTG was added. Cells were grown for

4hrs before they were harvested. Harvesting involved centrifuging cells for 10 min at 4000 g and then resuspending each 1 L in 20 mL resuspension buffer (300 nM NaCl, 50 mM Phosphate pH 7.5, 1 mM DTT, 1 µg/mL L/P/A, 1 mM PMSF). Resuspended cells were placed in 50 mL conicals and stored at -80 °C. Cells were broken down by thawing, adding lysozyme, dounce homogenization and sonication. Cell fragments were then removed by centrifugation (30 min, 26,000 g). The remaining cell particles were removed by syringe filtration (0.44 µm).

### **2.2.2 Purification**

The N-terminal His<sub>6</sub> tag was then utilized to bind Mad2 to Talon© cobalt beads. The beads were washed three times with binding buffer (150 mM NaCl, 50 mM Phosphate, 1 mM DTT, 1 µg/mL L/P/A, 1 mM PMSF, 5 mM Imidazol). Binding was done in batch using a ratio of 2 mL packed per 1 L growth media. Beads were incubated with the cell extract on a rotator at 4 °C for 30 min. The beads were then washed several times with binding buffer plus an additional 5 mM Imidazol and then with binding buffer without L/P/A or PMSF. The bound protein was incubated with PreScission protease at 4 °C for 1 hr. The bead slurry was placed in disposable gravity filtration columns and the flow-through was collected. A typical gel result is shown in Figure 6.4. The resulting solution contained only the protease and cleaved Mad2 as the His<sub>6</sub> tag and other contaminants remained behind on the beads.



**Figure 6.4** Mad2 purification gel. Lanes are as follows: 1. 5  $\mu\text{g}/\text{mL}$  BSA, 2. 1  $\mu\text{g}/\text{mL}$  BSA, 3. Post-sonication pellet, 4. Post-sonication supernatant, 5. Talon-bead flowthrough, 6. Protein bound to Talon-beads, 7. Protein remaining on beads after protease incubation. 8. Protein collected after protease incubation.

### 2.2.3 Chromatography

The flow through was then syringe filtered (0.22  $\mu\text{m}$ ) and diluted with ddH<sub>2</sub>O until the final NaCl concentration was <50 mM. The sample was then injected onto a FPLC Source 15Q HR 16/10 column. A NaCl gradient was used to elute Mad2 from the Q-column. The PreScission protease did not bind to the column and thus came out in the flow through. Mad2 eluted as two major peaks, one at 150 nM NaCl and one at 260 nM NaCl (Fig. 6.5A). The peaks correspond to O-Mad2 and C-Mad2 respectively, as determined previous (11, 35) and verified by 1D and 2D NMR (discussed in a later

section). Mad2 can be found predominantly O-Mad2 conformation when purification is done at low temperature (35) (25 °C induction, 4 °C purification). A typical purification from a 6L growth would yield 20-40 mg of protein, all of which could be loaded onto the Q-column in a single batch.

Mad2 oligomers were separated using a HiLoad<sup>TM</sup> 16/60 Superdex<sup>TM</sup> 200 prep grade size exclusion column (SEC). For each mutant it was confirmed that oligomers were only found in the anion exchange fractions associated with C-Mad2. All experiments involving C-Mad2 were done using purified monomer.

### **2.3 Peptide Synthesis**

Mad2 Binding Peptide 1 (MBP1) is a synthetic peptide used as a soluble binding partner for Mad2 (sequence SWYSYPPPQRAV; (25). A fluorescently labeled version of this peptide (RhB-MBP1) was generated by covalently linking this peptide to a Rhodamine B derivative that was kindly provided by Dr. S. Vinogradov. The fluorophore consisted of Rhodamine B attached to the backbone carbonyl of glutamic acid by a piperidine ring. Peptide synthesis was done using a CEM Liberty microwave peptide synthesizer in the Dutton laboratory using standard Fmoc/tBu protection protocols (36). The Rhodamine B derivative was first linked to the resin by the glutamic acid sidechain and the rest of MBP1 was sequentially synthesized in the direction of C-terminus to N-terminus. After synthesis the protein was cleaved from the resin by incubation with a mixture of trifluoroacetic acid (TFA), ethanedithiol, anisole, and thioanisole in a 9:0.2:0.5:0.3 ratio. Rotovap was used to remove excess reagent. The crude product was

then precipitated with methyl,t-butyl ether. The peptide was then purified by reverse phase HPLC using a C18 column and a gradient of acetonitrile with 0.1% (v/v) TFA in water. The final product was verified using an Orbi-trap mass spectrometer.

Concentrations for the synthetic peptide RhB-MBP1 were estimated based on absorbance. Rhodamine B is known to have a maximum absorbance at 542.75 nm with an extinction coefficient of  $106,000 \text{ M}^{-1}\text{cm}^{-1}$ . The absorbance maximum for RhB-MBP1 has been shifted to 568 nm. Concentrations listed here were calculated using the Rhodamine B extinction coefficient applied to the RhB-MBP1 absorbance at 568 nm.

## **2.4 Spectroscopy**

### **2.4.1 Circular Dichroism**

Circular dichroism (CD) measurements were done using an Aviv model 280 circular dichroism spectrometer. All measurements were taken at 4 °C unless otherwise specified. Samples concentrations were such that  $\text{OD}_{220} < 2$  so that sample opacity would not interfere with CD measurements.

### **2.4.2 Fluorescence Correlation Spectroscopy**

Florescence correlation spectroscopy (FCS) experiments were done to determine how much RhB-MBP1 bound to C-Mad2<sup>RQ</sup> based on diffusion times for the bound and unbound species. These experiments were done in the F. Gai lab. A 30-40  $\mu\text{L}$  sample was placed on a cover slip. An Argon laser was used to irradiate a 28 femtoliter region of the

sample. In each experiment 50 nM RhB-MBP1 was mixed with 4  $\mu\text{M}$  protein (either Mad2<sup>RQ</sup> or BSA). FCS results were fit using the auto-correlation function (Eq. 6.1).

$$G(t) = \frac{1}{N} \sum_{i=1}^n f_i \left( \frac{1}{1 + \frac{\tau}{\tau_D^i}} \right) \left( \frac{1}{1 + \frac{\tau}{\omega^2 \tau_D^i}} \right)^{1/2} \quad \text{Eq. 6.1}$$

Where N is the total number of molecules in the confocal volume, n is the number of species in the sample, and  $f_i$  is the fraction of species i in the sample. The ratio of the axial/radial dimensions of the observation volume is denoted as  $\omega$ , with equals 7 in this case. The characteristic diffusion time is  $\tau_D^i$  for species i. The delay time ( $\tau$ ) ranged from 1  $\mu\text{s}$  to 1 s.

Fluorescence intensity experiments were done in the Vanderkooi lab using a Fluorolog-3-21 Jobin-Yvon Spex SA fluorometer equipped with a 450 W xenon lamp for excitation and a cooled R2658P Hamamatsu photomultiplier tube for detection. RhB-MBP1 was excited at 560 nm. Emission was detected at 570-610 nm. NATA was excited at 280 nm and emission was detected at 340-380 nm. Concentrations were 400 nM for RhB-MBP1 and 4  $\mu\text{M}$  Mad2<sup>RQ</sup> unless otherwise noted. Temperature control was maintained using a water bath connected to the cuvette holder in the fluorometer. When multiple samples were used, a second cuvette holder was connected to the water bath in serial. Samples were stored in the second cuvette holder between measurements. Temperatures are reported for the primary cuvette holder in the instrument. The temperature for the secondary cuvette holder was lower than the primary cuvette holder by 0.5  $^{\circ}\text{C}$ .

### 2.4.3 NMR

1D  $^1\text{H}$  NMR experiments for Mad2<sup>RQ</sup> consisted of applying a simple 90 pulse plus a water suppression gradient at regular time increments using a 500 MHz magnet with Varian cold-probe. 2D O-Mad2<sup>RF</sup> to C-Mad2<sup>RF</sup> conversion experiments similarly involved applying a  $^1\text{H}$ - $^{15}\text{N}$  HSQC pulse sequence at regular time intervals using a 500 MHz magnet with a Varian cold-probe. Hydrogen exchange (HX) experiments for O-Mad2<sup>RF</sup> also involved applying a  $^1\text{H}$ - $^{15}\text{N}$  HSQC pulse sequence at regular time intervals following exchange into  $^2\text{H}_2\text{O}$  buffer and were done using a 600 MHz magnet with a Varian cold-probe located at the University of Chicago.

Attempts at peak assignments were done by Dr. Yibing Wu. The initial attempt was done using a 500  $\mu\text{L}$  sample of 0.44 mM  $^{13}\text{C}$ - $^{15}\text{N}$  labeled Mad2<sup>RF</sup> with 25 mM phosphate, 1 mM  $\text{NaN}_3$ , 20 mM deuterated DDT, and 10%  $^2\text{H}_2\text{O}$  kept in a flame sealed tube filled with Argon gas. The sample was checked using a 600 MHz magnet with a Varian room temperature probe at the Fox Chase Cancer Center. The sample was then shipped to the NMR Facility at Madison (NMRFaM) where experiments were run remotely on a 900 MHz magnet with a Varian cold-probe. A second attempt at assigning peaks was done using a 300  $\mu\text{L}$  sample of 0.7 mM  $^{13}\text{C}$ - $^{15}\text{N}$  labeled Mad2<sup>RF</sup> with 25 mM phosphate, 1 mM  $\text{NaN}_3$ , 20 mM deuterated DDT, and 5%  $^2\text{H}_2\text{O}$  in a Shigemi tube. This round of 3D experiments were done using a 500 MHz magnet with a Varian cold probe. All experiments were done at 21 °C unless otherwise noted. Processing was done with nmrPipe (37).

## 2.5 Hydrogen Exchange Mass Spectroscopy

Hydrogen exchange mass spectroscopy for Mad2<sup>RF</sup> were performed and processed by Sandya Ajith. Purified O-Mad2 and purified C-Mad2 were incubated in a solution of 25 mM Tris, 150mM NaCl in a 75% <sup>2</sup>H<sub>2</sub>O/<sup>1</sup>H<sub>2</sub>O mixture at pH<sub>read</sub> 7.0. At each time point the HX reaction was quenched with low pH buffer consisting of 0.8 M GdmCl, 0.8% formic acid and 10% glycerol, which was immediately flash frozen and stored at -80 °C. Samples were later thawed and injected onto a pepsin and fungal protease column for digestion. The resulting peptides were then sprayed into an Orbi-trap mass spectrometer. Peptide identification and analysis were done using DXMS software.

## 2.6 Electrostatic Calculations

Electrostatic calculations were based on the X-ray structures 1DUJ (O-Mad2) and 1S2H (C-Mad2). Hydrogens were added using CHARMM (38). Electrostatic calculations were performed using Qnifft (39, 40). Atom sizes and partial charges were set using the CHARMM22 parameter set (38).

## 3. Results and Discussion

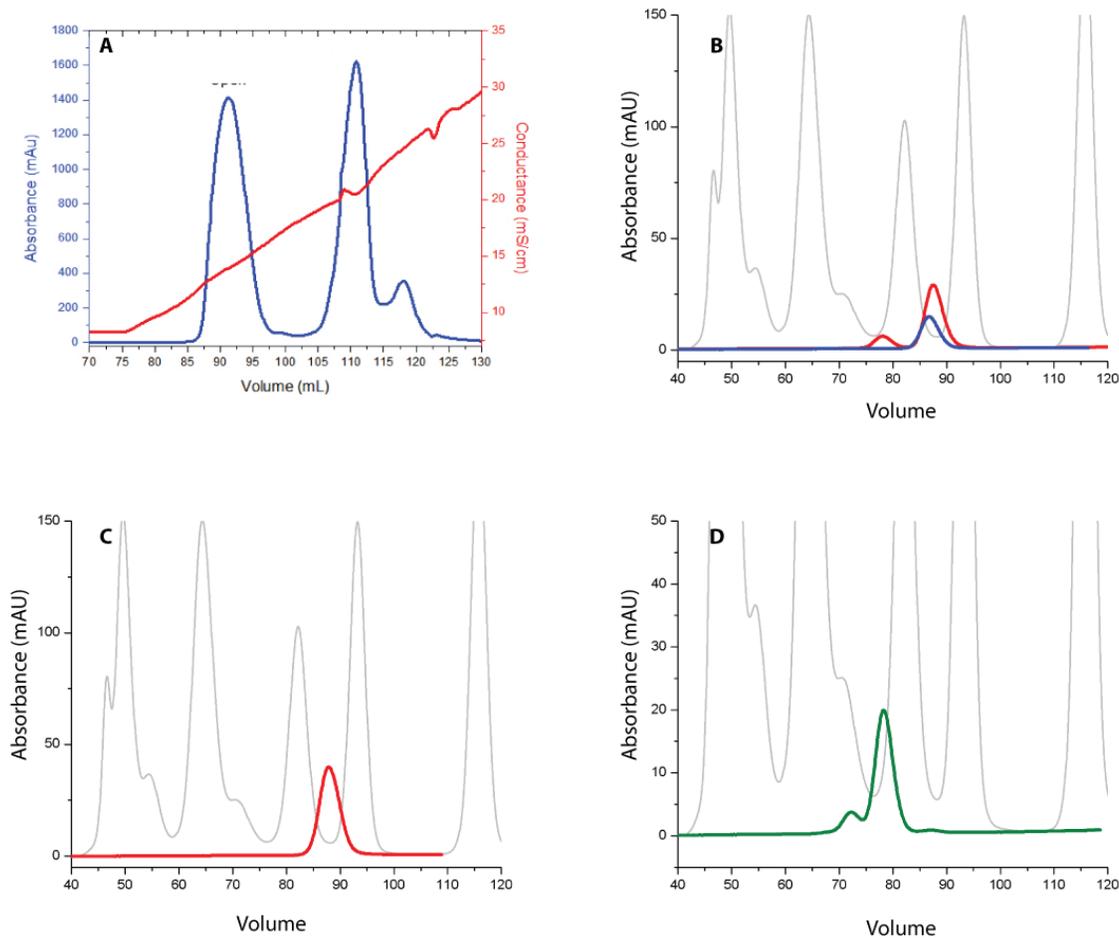
### 3.1 Separation of Mad2 conformers

Mad2 can be separated into its two conformations by anion exchange (11, 35). Figure 6.5A illustrates the separations of Mad2 into the lower salt O-Mad2 and higher salt C-Mad2 conformations. Once separated interconversion could be delayed storing

conformers at 4 °C. Conversion rates as a function of temperature are discussed in section 3.2.

The Mad2<sup>RQ</sup> and Mad2<sup>RF</sup> mutants have been reported to completely abolish Mad2 dimerization (34). We tested this for Mad2<sup>RQ</sup> and Mad2<sup>RF</sup> using size exclusion chromatography (SEC). Purified O-Mad2 did elute as a monomer, but C-Mad2 eluted both as monomer and as dimer (Fig 6.5B). In order to determine if the relative peak volumes represented the monomer-dimer equilibrium, the C-Mad2 monomer fractions were pooled and concentrated to the same concentration as the previous injection sample and incubated for an hour. No dimerization was observed for this injection (Fig 6.5C) despite wild type Mad2 dimerization kinetics being on the order of seconds (15).

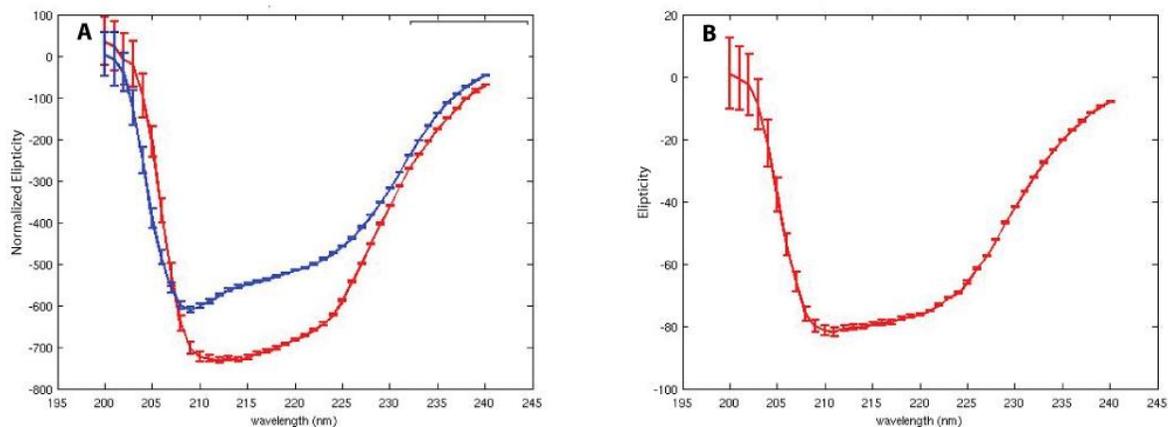
The slow kinetics observed for oligomerization of C-Mad2 dimerization deficient mutants present the possibility that dimerization for these mutants occurs at a different interface or by a different mechanism than the one measured for wild type Mad2. A crystal structure of the C-Mad2—C-Mad2 dimer (2VFX; (19) presents a possible explanation. For this structure the unit cell consists of twelve C-Mad2 molecules. Each molecule interacts with another C-Mad2 molecule at normal dimer interface, as seen for the C-Mad2—O-Mad2 dimer (28), and interacts with another C-Mad2 molecule by incorporating that protein's N-terminus in its  $\beta$ -sheet. This same binding motif is associated with Mad1—C-Mad2 (Fig. 6.2A; (14, 25) and Cdc20—C-Mad2 (10) interactions. Three mutations were designed to abolish this putative interaction. Any one of these mutations (L3P, L3R and Q4P) should disrupt the N-terminal binding seen in the 2VFX structure, but none affect the elution pattern observed by SEC.



**Figure 6.5** Anion exchange and size exclusion chromatograms for Mad2<sup>RQ</sup>. (A) Chromatograph from anion exchange of Mad2<sup>RQ</sup>. O-Mad2 elutes at a lower salt concentration than C-Mad2. (B) Size exclusion chromatograph (SEC) for O-Mad2<sup>RQ</sup> (blue) and C-Mad2<sup>RQ</sup> (red) after anion exchange purification. (C) When the C-Mad2<sup>RQ</sup> monomer was concentrated and incubated for 1 hr it still eluted as a monomer. (D) The small, third peak from A eluted as mostly dimer with some higher order oligomers.

A third peak occasionally observed during anion exchange purification was investigated. This peak elutes at a higher salt concentration and is smaller than O-Mad2<sup>RQ</sup> and C-Mad2<sup>RQ</sup> peaks. The SEC chromatogram for this species (Fig 6.5D) looks very similar to the one observed for C-Mad2<sup>RF</sup> (Fig 6.5B), except the peaks are shifted to the left. The shift indicates that this species is mostly dimer with some tetramer also present.

The CD spectrum for the oligomer species is nearly identical to that of C-Mad2, (Fig. 6.6). This suggests that oligomerization does not significantly distort the helical structure of Mad2. The possibility of a large distortion of the  $\beta$  structure remains.



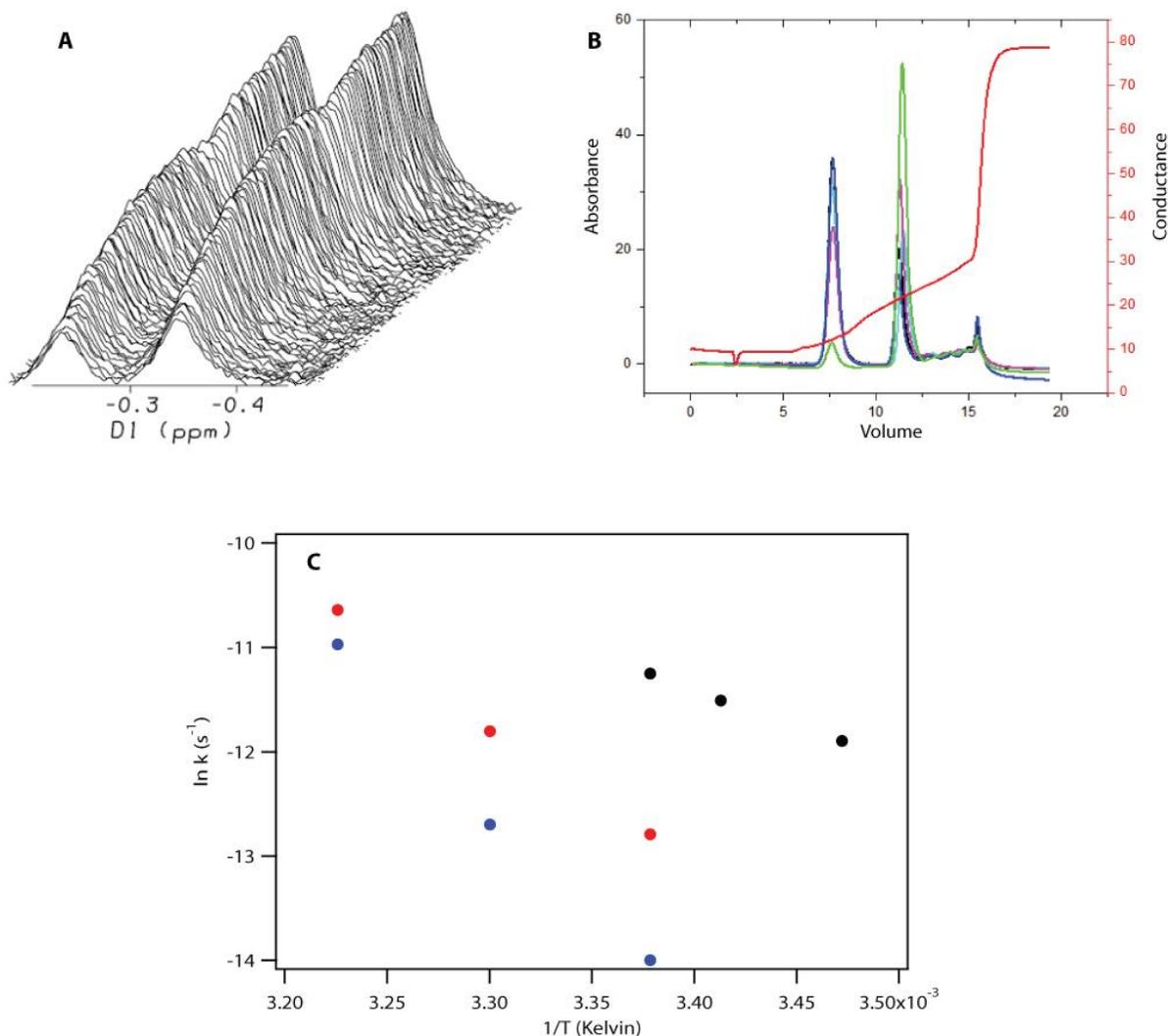
**Figure 6.6** CD spectra for O-Mad2<sup>RF</sup> (A; blue), C-Mad2<sup>RF</sup> (A; red) and Mad2<sup>RF</sup> oligomer (B).

### 3.2 Mad2 Interconversion

#### 3.2.1 Macroscopic Interconversion

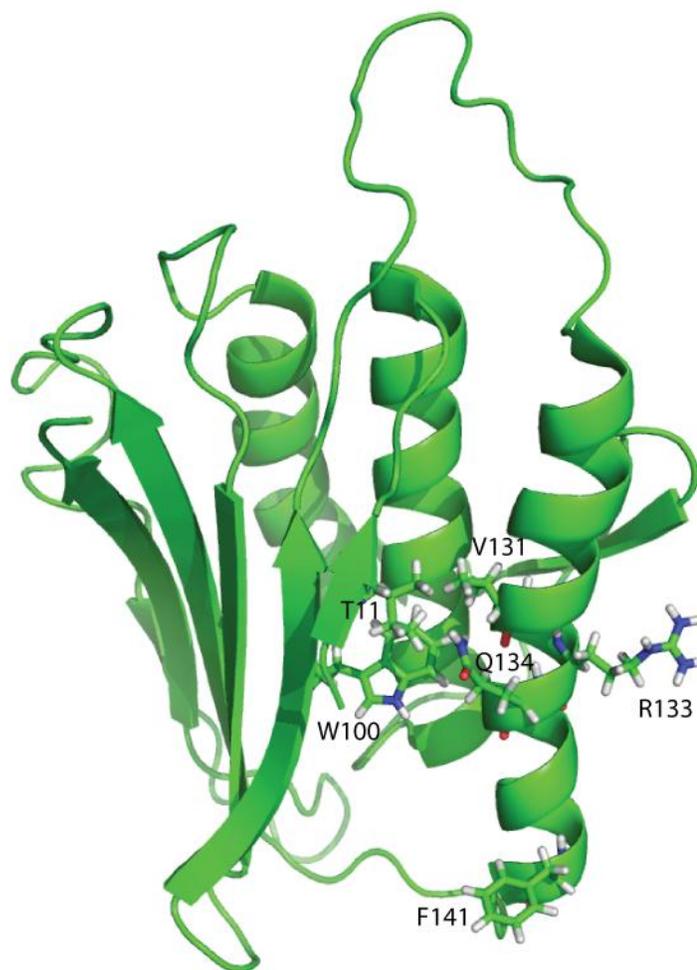
Mad2 interconversion is central to its function as a mitotic regulatory protein (41). The first kinetic measurements of Mad2 interconversion were done by monitoring the 1D Val197 <sup>1</sup>H methyl peak by NMR. Luo and coworkers reported the dimerization reduced Mad2<sup>R133A</sup> mutant to have an O-Mad2 to C-Mad2 rate of  $3.0 \times 10^{-5} \text{ s}^{-1}$  and a reverse rate of  $5.2 \times 10^{-6} \text{ s}^{-1}$  at 30°C (11). Later it was discovered that the double mutant Mad2<sup>RQ</sup> and Mad2<sup>RF</sup> suppressed dimerization better than Mad<sup>R133A</sup> (34). We measured kinetics for the Mad2<sup>RQ</sup> by 1D NMR and found much faster rates (Fig. 6.7A; Table 6.2). Interestingly, we observed two peaks whereas Luo et al reported a single peak, suggesting that the

Val197 sidechain methyl groups experience slightly different environments in Mad2<sup>RQ</sup> relative to Mad2<sup>R133A</sup>. S. Getchel and B. Wexler measured kinetics for Mad2<sup>RF</sup> using anion exchange FPLC (Fig. 6.7B).



**Figure 6.7** Mad2 interconversion. (A) O-Mad2<sup>RQ</sup> to C-Mad2<sup>RQ</sup> conversion was measured by monitoring the C-Mad2 Val197 <sup>1</sup>H methyl peaks by 1D NMR. (B) Overlay of chromatograms from several Mad2<sup>RF</sup> interconversion time points which were analyzed by anion exchange FPLC. Chromatograms for various time points are overlaid. Anion exchange measurements were done by S. Getchel and B. Wexler. (C) Arrhenius plot interconversion rates for O-Mad2<sup>RQ</sup> to C-Mad2<sup>RQ</sup> (black), O-Mad2<sup>RF</sup> to C-Mad2<sup>RF</sup> (red), and C-Mad2<sup>RF</sup> to O-Mad2<sup>RF</sup> (blue).

The kinetics of open to closed conversion for the Mad2<sup>RQ</sup> and Mad2<sup>RF</sup> mutants differ in an absolute sense and in terms of temperature dependence (Fig. 6.7C; Table 6.2). This temperature dependence corresponds to a  $\Delta H^\ddagger$  of 14 kcal/mol for the Mad2<sup>RQ</sup> open to closed conversion and 28 kcal/mol for Mad2<sup>RF</sup>. The NMR structure of O-Mad2<sup>ANAC</sup> may provide some insight into the cause of this difference (Fig. 6.8). Notably, the polar Gln134 sidechain is inserted into a hydrophobic pocket. This interaction should be enthalpically disfavored and therefore promote open to closed conversion, yet measurements indicate the opposite. Gln134 is absent in the rapidly converting Mad2<sup>RQ</sup> mutant. The suggestion is that the Q134A mutation is even more enthalpically favored in the transition state. Also notable is that the difference in rates for the two mutants corresponds to a  $\Delta\Delta G^\ddagger$  of only 0.9 kcal/mol, indicating that the large enthalpic discrepancy is largely offset by an entropic difference that is nearly as large ( $T\Delta\Delta S^\ddagger = 13$  kcal/mol).



**Figure 6.8** O-Mad2<sup>ΔNΔC</sup> structure (pdb 1DUJ) highlighting interactions involving the residues that have been mutated in Mad2<sup>RQ</sup> and Mad2<sup>RF</sup>. Mutations involving Arg133 and Phe141 presumably disrupt Mad2 dimerization without disturbing internal interactions. Conversely, the Gln134 sidechain is packed among hydrophobic residues Thr11, Trp100, and Val131.

Kinetic differences between Mad2<sup>R133A</sup> and Mad2<sup>RF</sup> provide insight into the role of Phe141. The O-Mad2 to C-Mad2 conversion rates for these mutants differ by a factor of 4 while their O-Mad2 and C-Mad2 equilibrium differs by only a factor of 2 (Table 6.2). This means that the F141A mutation destabilizes C-Mad2 relative O-Mad2, but destabilizes the transition state even more. The difference in equilibrium may be due to Phe141 being exposed to solvent in O-Mad2 (Fig. 6.8) and buried among hydrophobes in

C-Mad2. The C-terminus of  $\alpha$ C rearranges from an  $\alpha$  helical conformation in O-Mad2 to a 3-10 helical conformation in C-Mad2. The fact that the F141A mutation affects conversion rates more than equilibrium populations suggests that the rearrangement of residue 141 is integral to the rate-limiting step in conformational conversion.

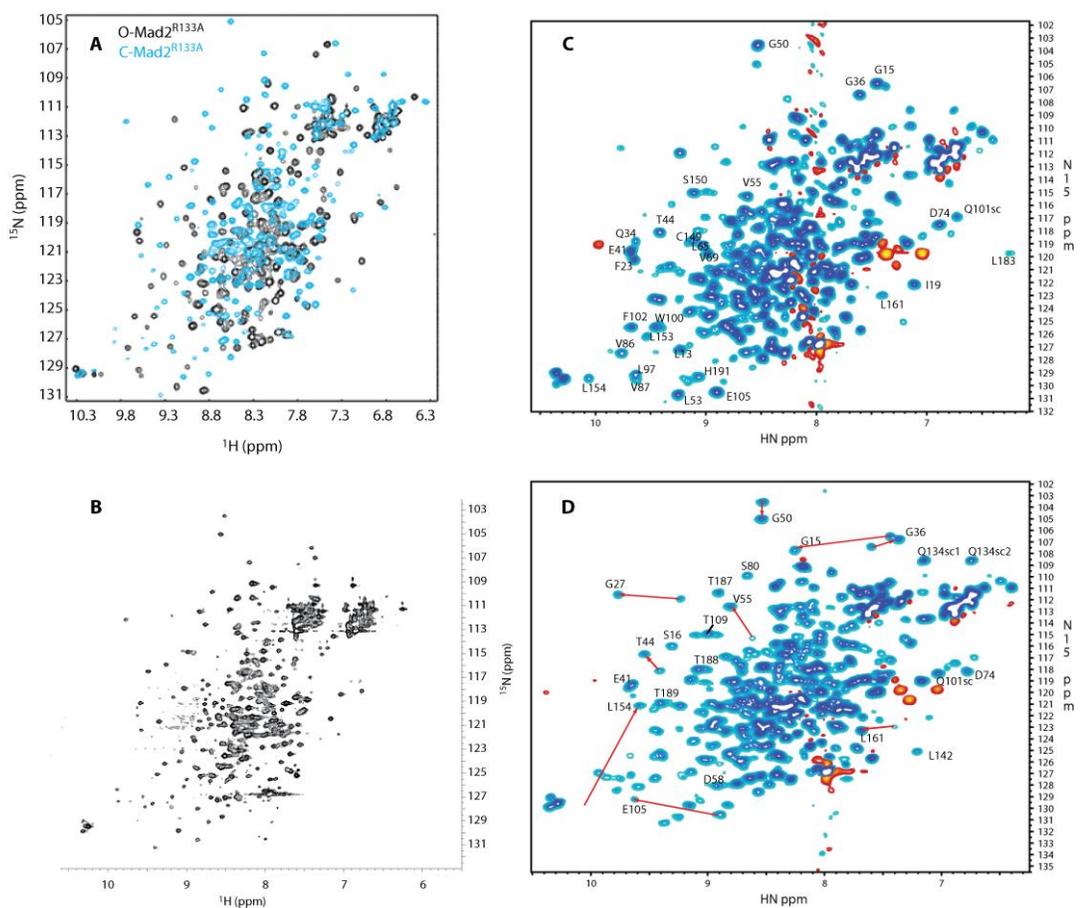
**Table 6.2** Temperature dependent interconversion rates.

Conversion	Mutant	Temperature (K)	Rate (s <sup>-1</sup> )
O to C	RQ	288	6.8E-06
O to C	RQ	293	1.0E-05
O to C	RQ	296	1.3E-05
O to C	RF	296	2.8E-06
O to C	RF	303	7.5E-06
O to C	RF	310	2.4E-05
O to C	R133A	303	3.0e-05
C to O	RF	296	8.3E-07
C to O	RF	303	3.1E-06
C to O	RF	310	1.7E-05
C to O	R133A	303	5.2E-06

### 3.2.2 Residue Specific Interconversion

<sup>1</sup>H-<sup>15</sup>N HSQC NMR experiments were used to monitor the open to closed conversion at single residue resolution for Mad2<sup>RF</sup>. Initial assignment of amide peaks was done by direct comparison to published peak tables. Previous workers reported NMR assignments for O-Mad2<sup>ΔNΔC</sup> (BMRB 4775; (10) and C-Mad2<sup>ΔN</sup> bound to MBP1 (BMRB 5299; (25). These spectra have fewer peaks than the Mad2<sup>RF</sup> spectra due to truncations, which makes peak assignment by direct comparison difficult. This problem was alleviated somewhat by the publication of O-Mad2<sup>R133A</sup> and C-Mad2<sup>R133A</sup> spectra (Fig. 6.9A; (11). Though peak tables were not published for these mutants, direct comparison of previously published peak tables to Mad2<sup>R133A</sup> spectra allowed for the assignment of

some well dispersed peaks for Mad2<sup>RF</sup>. One complexity that had to be accounted for was the presence of both O-Mad2 and C-Mad2 peaks in the same spectra (Fig. 6.9B). Peaks were assigned to conformation by purifying O-Mad2<sup>RF</sup> and watching peaks decay or rise as the population converted to C-Mad2<sup>RF</sup> (Fig. 6.9C-D).



**Figure 6.9** <sup>1</sup>H-<sup>15</sup>N HSQC spectra for Mad2. Yu and coworkers published <sup>1</sup>H-<sup>15</sup>N HSQC spectra for O-Mad2<sup>R133A</sup> and C-Mad2<sup>R133A</sup> (A; adapted from (11)). Both conformations can be seen for Mad2<sup>RF</sup> at equilibrium (B; 900 MHz). Spectra collected during a time series monitoring O-Mad2<sup>RF</sup> to equilibrium conversion are displayed for the initial (C) and ~4hr (D) time points. Peaks that could be assigned are shown for O-Mad2 (C) and C-Mad2 (D). An arrow points from the O-Mad2 peak to the C-Mad2 peak for peaks that were assigned for both conformations. The peak corresponding to W100 in O-Mad2 does not decay over the course of this experiment.

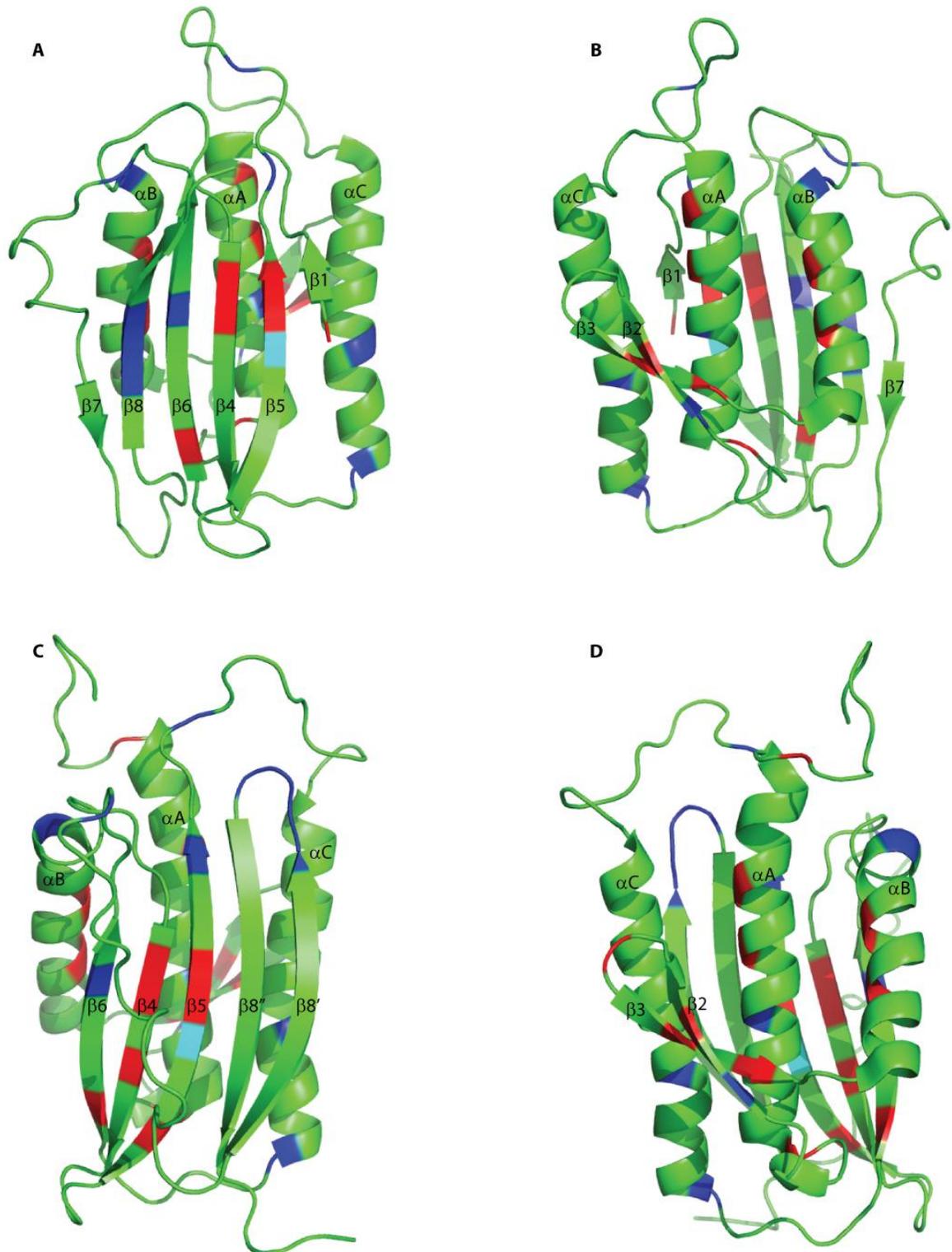
Oddly, the peak transitions observed in Mad2<sup>RF</sup> spectra at 21 °C occur at a rate of  $\sim 1.5 \times 10^{-4} \text{ s}^{-1}$  (Table 6.3) much faster than the open to closed conversion measured by anion exchange (Table 6.2). The implication is that Mad2<sup>RF</sup> forms a stable intermediate between the closed and open transition. We have suggested this possibility previously (41), but no experimental evidence has been reported until now. The shifting peaks correspond to residues found throughout Mad2 as can be seen by mapping those that could be assigned onto the structure (Fig. 6.10). Surprisingly, many of these peaks correspond to residues found in  $\beta 2$ ,  $\beta 3$  and  $\alpha A$ , which, other than the N-terminus of  $\alpha A$ , are very similar in the open and closed structures (Fig. 6.10B, D).  $\beta 2$  and  $\beta 3$  serve to connect  $\alpha A$  to  $\alpha B$ , even though the C-terminus of  $\alpha A$  is near the N-terminus of  $\alpha B$ . The fact that these strands are not necessary to connect  $\alpha A$  to  $\alpha B$  causes one to wonder why they are present in Mad2. The fact that open to closed conversion causes residues in and around these strands to experience a change that alters their  $^1\text{H}$ - $^{15}\text{N}$  chemical shifts suggests that  $\beta 2$  and  $\beta 3$  play some subtle, yet to be determined role in conformational conversion.

**Table 6.3** Rates of decay for O-Mad2 and rise of C-Mad2 peaks that could be clearly distinguished

O-Mad2	log k (s <sup>-1</sup> )	C-Mad2	log k (s <sup>-1</sup> )
G9	-4.4	G27	-3.7
I19	-3.9	D74	-3.8
F23	-4.2	T109	-3.5
G27	-3.9	Q134sc	-3.5
G36	-4.2	Q134sc	-3.6
T44	-4.1	L154	-3.9
V46	-4.2	T187	-3.7
G50	-4.0	T188	-3.5
L53	-4.0		
V55	-4.2		
L65	-4.1		

O-Mad2	log k (s <sup>-1</sup> )	C-Mad2	log k (s <sup>-1</sup> )
V69	-4.1		
L84	-4.4		
V85	-4.1		
V87	-3.9		
L97	-3.9		
Q101sc	-3.6		
F102	-4.1		
S150	-4.4		
L154	-3.8		

Also notable is that the O-Mad2 peak for W100 does not decay over the time course, indicating that it still experiences an O-Mad2-like environment. W100 interacts with  $\beta 8''$  in C-Mad2 (Fig. 6.10C). This is surprising given that some peaks corresponding to residues found in C-Mad2  $\beta 8'$  and  $\beta 8''$  rise in this experiment. One possibility is that these strands rearrange into a C-Mad2-like arrangement before moving toward the opposite side of the major  $\beta$  sheet.

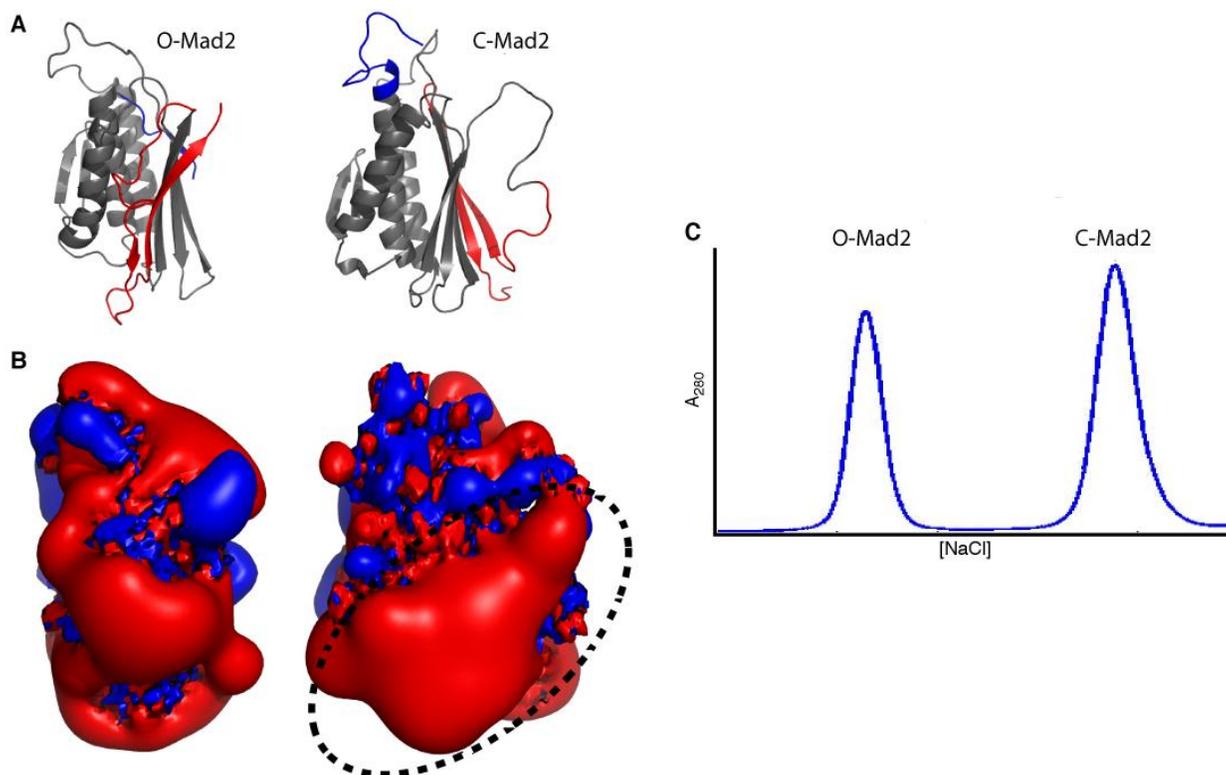


**Figure 6.10**  $^1\text{H}$ - $^{15}\text{N}$  peaks that shift from open to closed conformations mapped onto O-Mad<sup>ANAC</sup> (A, B) and C-Mad2<sup>R133A</sup> (C, D) structures. Colors indicate residues corresponding to C-Mad2 peaks that increased

with time (blue) and to O-Mad2 peaks that clearly decayed, but for which no rising C-Mad2 peak could be found (red). The peak corresponding to O-Mad2 W100 (cyan) maintained constant volume throughout the experiment. Residues that could not be assigned or for the corresponding peak has identical chemical shifts in both forms are shown in green.

### 3.2.3 Electrostatic Considerations

The differences in conversion rates measured by  $^1\text{H}$ - $^{15}\text{N}$  NMR compared and anion exchange for the same mutant (Tables 6.2, 6.3) may be due to these experiments responding to different regions of Mad2. The signals measured by  $^1\text{H}$ - $^{15}\text{N}$  NMR arise from the environment of individual residues as discussed in the previous section. What is being monitored by anion exchange is not as clear. At first glance it is surprising that the two Mad2 conformations have different affinities for anion exchange given that their primary sequence is identical. A protein's affinity for an ion exchange column can be estimated by its pI, which can be calculated from its sequence. Clearly the difference in affinity arises from conformational rearrangement, which suggests that binding affinity is influenced by local structure. We used Qniff (39) to calculate the electrostatic fields for each conformation (Fig 6.11). Calculations indicate that C-Mad2 has a high density of negative charges at a single region of its surface while the surface charges for O-Mad2 are more evenly distributed. This high charge density is associated with  $\beta_6$ , which is exposed when  $\beta_7$  and  $\beta_8$  rearrange to form  $\beta_8'$  and  $\beta_8''$ . Therefore, it is possible that the O-Mad2 to C-Mad2 transition measured by  $^1\text{H}$ - $^{15}\text{N}$  NMR represents an intermediate that has undergone some rearrangement, but for which  $\beta_6$  is not exposed.



**Figure 6.11** Electrostatic analysis of differences in binding affinity for O-Mad2 and C-Mad2. Crystallographic structures are shown (A) in the same orientation as electrostatic maps (B). Negative (red) and positive (blue) surfaces extend to a distance equivalent to a potential energy of 1 kT. A large negative patch on the surface of C-Mad2 is circled. A typical anion exchange elution profile is shown in C.

### 3.2.4 Summary of the Mad2 Intermediate

Experiments designed to measure the rate of O-Mad2 to C-Mad2 conversion and performed near room temperature resulted in rate measurements on the order of  $\sim 3 \times 10^{-6} \text{ s}^{-1}$  for anion exchange and 1D NMR and  $\sim 1.5 \times 10^{-4} \text{ s}^{-1}$  (Table 6.2) for 2D NMR (Table 6.3). This 200-fold difference can be explained by the presence of an intermediate I-Mad2 conformation that is not observable by the first two methods. Examination of what is being detected by each method provides some insight into what I-Mad2 may look like.

Several pieces of evidence point to the possibility that O-Mad2  $\beta 7$  and  $\beta 8$  do not rearrange into their C-Mad2 positions in the intermediate I-Mad2 conformation. The rearrangement of these strands exposes a large negative patch at the protein surface (Fig. 6.11), which causes C-Mad2 to have a higher affinity to an anion exchange column. Similarly, strand rearrangement causes the sidechain of V197 to interact with aromatic groups which allows V197 methyl groups to be monitored by 1D NMR (Fig. 6.7A). Also notable is that 2D NMR experiments show that W100 does not switch to its C-Mad2 environment with the same rapid rate as other residues shown in Figure 6.10. This is significant because W100 interacts with  $\beta 8''$  in C-Mad2 and its sidechain interacts with the V197 sidechain. Taken together, these observations clearly indicate that these strands are not in their C-Mad2 position in I-Mad2. However, it should be noted that several residues located in these strands do experience a change in chemical environment. This may indicate that the strands undergo some rearrangement prior to switching to the opposite side of the sheet.

Many of the residues experiencing a C-Mad2-like environment in I-Mad2 are found in  $\beta 2$  and  $\beta 3$  and the closely linked helix  $\alpha A$  (Fig. 6.10). The fact that these residues experience different chemical environments in the two known Mad2 conformers is somewhat surprising given their structural similarities. It is possible that these residues play an important role in interconversion, which may explain why it is partially involved in the catalyzing Mad2 dimerization reaction (Fig. 6.2B). It is quite possible that this interaction is essential to catalyzing the Mad2 conversion by stabilizing I-Mad (Fig. 6.3C).

### 3.3 Hydrogen Exchange

#### 3.3.1 NMR

Hydrogen exchange (HX) monitored by NMR has proven to be an effective tool for studying protein high energy states (42). This technique was applied to O-Mad2<sup>RF</sup> in the hope of identifying the residues that are involved in the O-Mad2 to C-Mad2 conversion. Several problems arose. The most significant issue was that most NMR peaks could not be assigned. Even when peaks could be assigned, it is difficult to distinguish peak volume decay due to HX from decay due to interconversion. If these rates are significantly different from one another then peak decay will fit to a double exponential, but if these rates are similar then the data will fit well to a single exponential with a rate somewhere between the two actual rates.

Rates are given for peaks that could be assigned and whose rates were not compromised by peak overlap (Table 6.4). It is not clear why some peaks would decay more slowly during HX experiments than they would in the absence of <sup>2</sup>H<sub>2</sub>O (i.e. G27, V55, L65, V69, L84, V85 and L97). It should be noted that a change in chemical shift does not necessarily correspond to a structural change that would expose an amide to HX catalyst. For example, Gly27 is found in the middle of  $\alpha$ A in both O-Mad2 and C-Mad2, but nearby changes involving  $\beta$ 1 and  $\beta$ 2 cause the peak to shift. In this case one expects the O-Mad2 Gly27 peak to decay as the population reaches its O-Mad2 to C-Mad2 equilibrium and then to decay further as the amide is exposed to HX catalyst. Strangely, this initial decay was not observed in these experiments.

Also puzzling is that HX rates do not exhibit expected pH dependence. Under most experimental conditions one expects HX rates to increase by a factor of 10 for each pH unit (see Chapter 1 for an explanation of HX rates). Instead, HX rates increased by a factor of ~5 per pH unit. One possibility is that these residues have extremely fast reclosing rates so that HX is in the EX1 regime. Another possibility is that the stability of O-Mad2 increases with pH. At the pHs measured here one expects only Histidine to titrate in the absence of a significant pKa shift. Mad2 has only one histidine, His191, which is exposed to solvent in O-Mad2, but is buried among hydrophobes in C-Mad2. Unfortunately, attempt to determine protein stability as a function of pH proved fruitless as the unfolding and refolding rates were too slow to make equilibrium measurements possible. However, future workers may consider conformational interconversion as a function of pH to be an interesting avenue of investigation.

**Table 6.4** Peak decay rates for O-Mad2<sup>RF</sup> to C-Mad2<sup>RF</sup> conversion and O-Mad2<sup>RF</sup> HX. Rates in log (s<sup>-1</sup>).

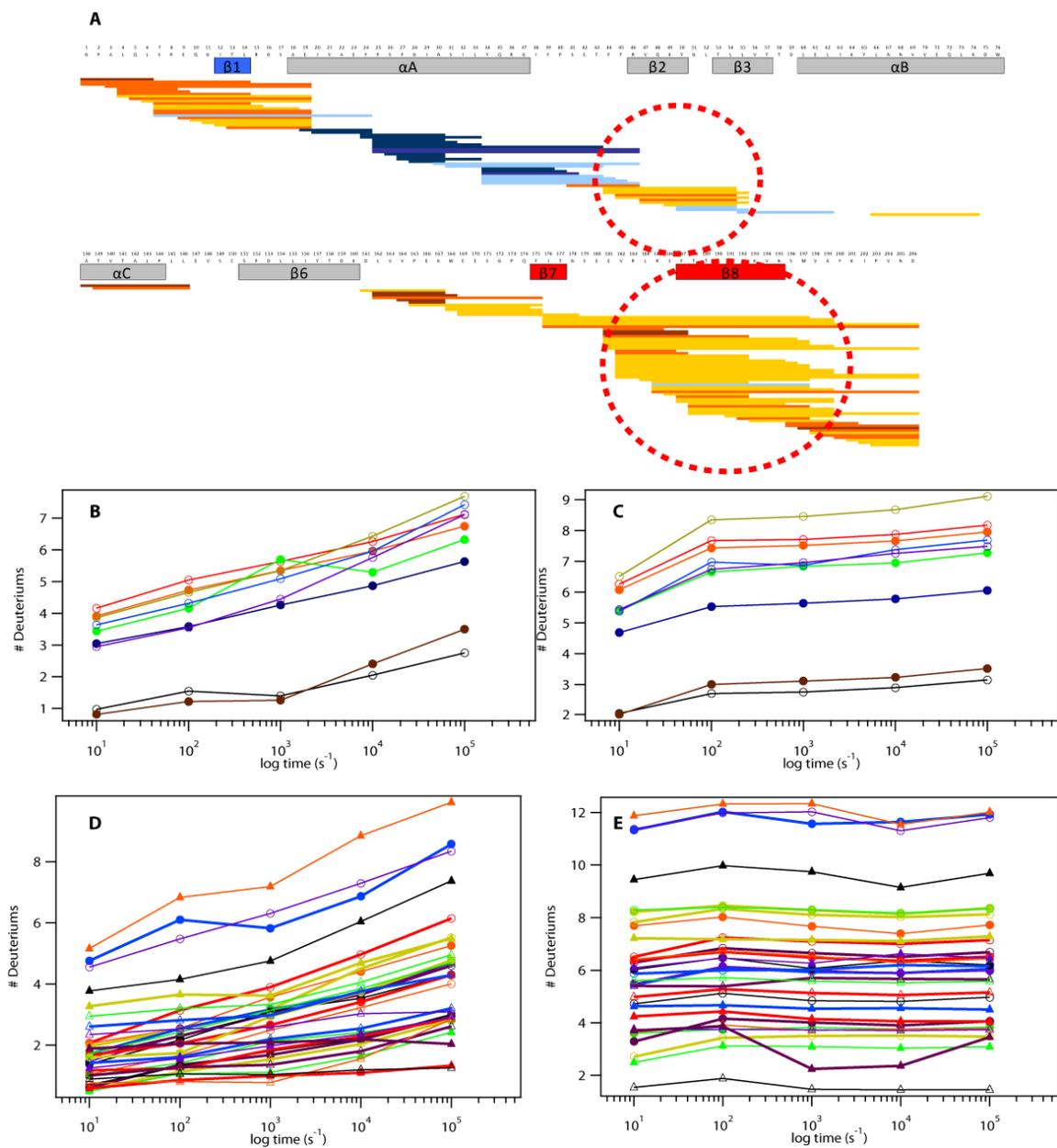
Residue	O to C	pD 6.56	pD 7.26	pD 8.15	pD 8.61	pD 9.91
I19	-3.9	-3.7	-3.5			
F23	-4.2	-4.2			-5.1	-3.6
G27	-3.9	-4.7	-5.1	-4.6	-4.6	
Q34		-4.5	-4.5			
G36	-4.2	-3.3	-3.0			
Y38		-4.7	-4.4			
T44	-4.1	-4.0	-3.7			
V46	-4.2	-4.5	-4.1			
G50	-4.0					
L53	-4.0	-4.1	-4.0			
V55	-4.2	-4.9	-4.3			
L65	-4.1	-5.3				-4.1
V69	-4.1	-4.7				-4.3
L84	-4.4	-5.0	-4.5			-3.3
V85	-4.1	-4.9	-4.0			
V86		-4.7				
V87	-3.9		-4.5			
L97	-3.9	-4.9	-4.7	-4.6		
W100		-4.9				
C149						-3.7

Residue	O to C	pD 6.56	pD 7.26	pD 8.15	pD 8.61	pD 9.91
S150	-4.4	-4.0	-4.1			
V181			-3.4			

### 3.3.2 Hydrogen Exchange Mass Spectroscopy

Hydrogen exchange mass spectroscopy (HXMS) measurements for O-Mad2<sup>RF</sup> and C-Mad2<sup>RF</sup> were done by Sandya Ajith. A typical peptide map for a single time point is displayed in Figure 6.12A. Peptides spanning the  $\beta 2$   $\beta 3$  region exchanged differently for each conformation (Fig. 6.12B, C), as did peptides spanning  $\beta 8$  (Fig. 6.12D, E). For C-Mad2<sup>RF</sup>, these strands completely exchange within 100 s, while the O-Mad2<sup>RF</sup> peptides continue to exchange over 24 hr later. Interestingly, these peptides in O-Mad2<sup>RF</sup> continue to increase at each measured time point, indicating that HX is not occurring in a cooperative manner.

The fact that one can easily distinguish O-Mad2 from C-Mad2 in these regions bodes well for identifying I-Mad2 by HXMS. As mentioned in Section 3.2.4, one expects the  $\beta 2$   $\beta 3$  region of I-Mad2 to be C-Mad2-like and the C-terminal strands to be more O-Mad2-like. Our NMR experiments indicate that I-Mad2 is stably populated during the O-Mad2 to C-Mad2 transition. Thus a simple way to test our hypothesis about this transition is to apply an HX pulse of fixed duration at various time points throughout an O-Mad2 to C-Mad2 conversion experiment. If our hypothesis is correct then the  $\beta 2$   $\beta 3$  region will exhibit a C-Mad2 profile prior to the C-terminal region.



**Figure 6.12** HXMS data for O-Mad2<sup>RF</sup> (A,B,D) and C-Mad2<sup>RF</sup> (C,E). (A) A peptide map for a single O-Mad2<sup>RF</sup> time point. Color scale from blue to red indicates percent exchange for each peptide. The regions implicated in conversion in Figure 6.10 are circled. (B, C) HXMS time course for peptides spanning strands  $\beta$ 8 ( $\beta$ 8' and  $\beta$ 8'' in C-Mad2; spanning residues 44 to 55) for O-Mad2<sup>RF</sup> (B) and C-Mad2<sup>RF</sup> (C). (D, E) HXMS time course for peptides spanning strands  $\beta$ 2 and  $\beta$ 3 (spanning residues 181 to 206) for O-Mad2<sup>RF</sup> (D) and C-Mad2<sup>RF</sup> (E). Line and symbol colors in B and D correspond to the identical peptide in C and E respectively.

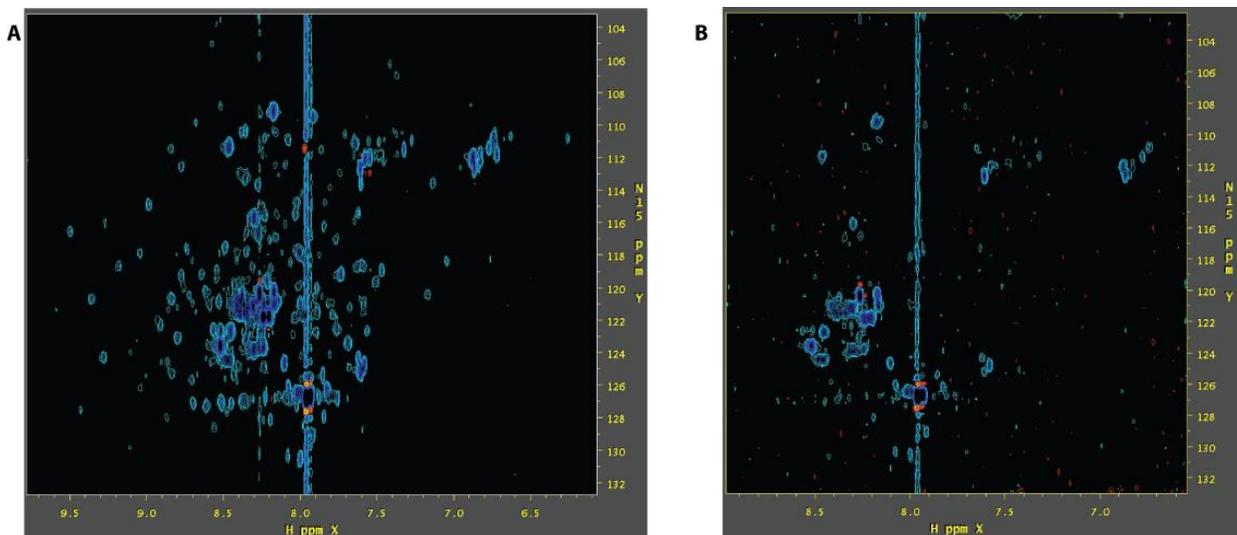
### 3.4 NMR Peak Assignments

The standard procedure for obtaining NMR peak assignments is to apply a series of 3D experiments that effectively trace the backbone. These experiments require high protein concentrations and long collection times, which are not ideal for the Mad2 system. The high protein concentration proved to be a challenge. Firstly, the use of minimal media to grow  $^{13}\text{C}/^{15}\text{N}$  labeled protein greatly reduces yield. Secondly, Mad2 tends to aggregate at concentrations approaching 1 mM. The concentration problem was partially mitigated by using a 900 MHz magnet with a cold probe.

The long times needed for 3D experiments were also challenging due to sample degradation and conformational interconversion. In order to minimize protein degradation from oxidation, the sample was kept under Argon gas in a flame sealed tube in the presence 20 mM deuterated DDT. Even under these conditions samples continued to aggregate over time. The problem of interconversion can be avoided by reducing temperature, but lowering temperature also increases molecular tumbling time which causes a significant reduction in signal-to-noise due to line broadening. We attempted to circumvent the interconversion problem by allowing the sample to reach equilibrium so that the total populations of O-Mad2 and C-Mad2 would remain constant during measurements.

In an attempt to assign peaks for Mad2<sup>RF</sup>, Dr. Yibing Wu planned and implemented a series of 3D experiments. Ultimately, this goal proved intractable due to low sample concentration, the presence of multiple conformations, and dynamics occurring on the time scale of the experiments. 2D projections from HNCA and

HN(CO)CA experiments clearly indicate that the signal is too weak for the purposes of backbone tracing (Fig. 6.13). Additional attempts using longer collection times and increased temperatures also failed to yield desired results.



**Figure 6.13** 2D projections of HNCA (A) and HN(CO)CA (B) spectra collected using a 900MHz magnet with cold probe.

Why were earlier workers able to obtain NMR assignments while we were not?

The answer most likely comes from the differences in mutants that were used. Published O-Mad2 assignments came from the double truncation mutant O-Mad2<sup>ΔNΔC</sup> (10). This mutant is incapable of forming C-Mad2, which eliminates complications arising from populating multiple conformations. Assignments published for C-Mad2 came from the C-Mad2<sup>ΔN</sup> mutant bound to ligand (25). Including the ligand greatly increases stability of C-Mad2, thereby eliminating multiple conformations. In both cases the unstructured N-terminus has been truncated, perhaps reducing aggregation. However, an apo-C-Mad2

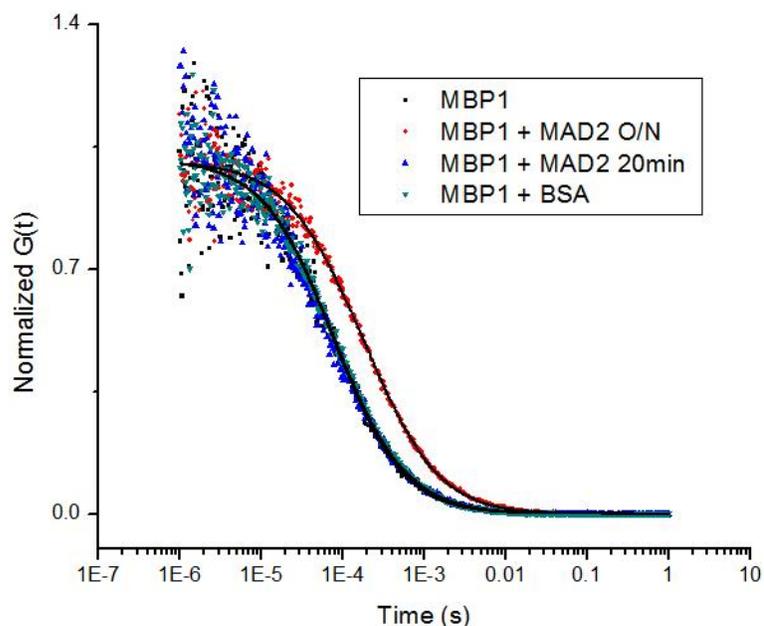
NMR structure was published for full length C-Mad2 (11), suggesting that the N-terminal truncation is not strictly necessary for obtaining NMR assignments. The essential difference between the protein used in that study and the one used here is that our version has the additional mutation F141A. In addition to further inhibiting dimerization, the F141A mutation stabilizes O-Mad2 relative to C-Mad2 as indicated by the nearly even distribution between the two conformations at equilibrium compared to the 1:6 ratio of O-Mad2 to C-Mad2 at equilibrium reported for Mad2<sup>R133A</sup> (11). This even distribution makes assignments more difficult. Also complicating assignments are the dynamic reactions that occur on the time scale of NMR measurements. It is possible that the F141A mutation has stabilized the I-Mad2 conformation or transition states, causing more rapid rearrangements that earlier workers did not have to deal with.

### **3.5 Peptide Binding**

Mad2 utilizes the same binding site to interact with both Mad1 and Cdc20 (25). The synthetic Mad2 binding peptide 1 (MBP1) was designed to mimic this interaction (25). We synthesized MBP1 covalently linked to a Rhodamine B derivative (RhB-MBP1) in order to monitor Mad2 binding by fluorescence.

Fluorescence correlation spectroscopy (FCS) can be used to calculate the diffusion constants for fluorescently labeled molecules. The diffusion time for RhB-MBP1 changes drastically upon binding to Mad2 due to the relative sizes of the molecules. Our results show that there is no discernable difference in RhB-MBP1 diffusion time for RhB-MBP1 alone, mixed with BSA, or mixed with C-Mad2 for 20 min

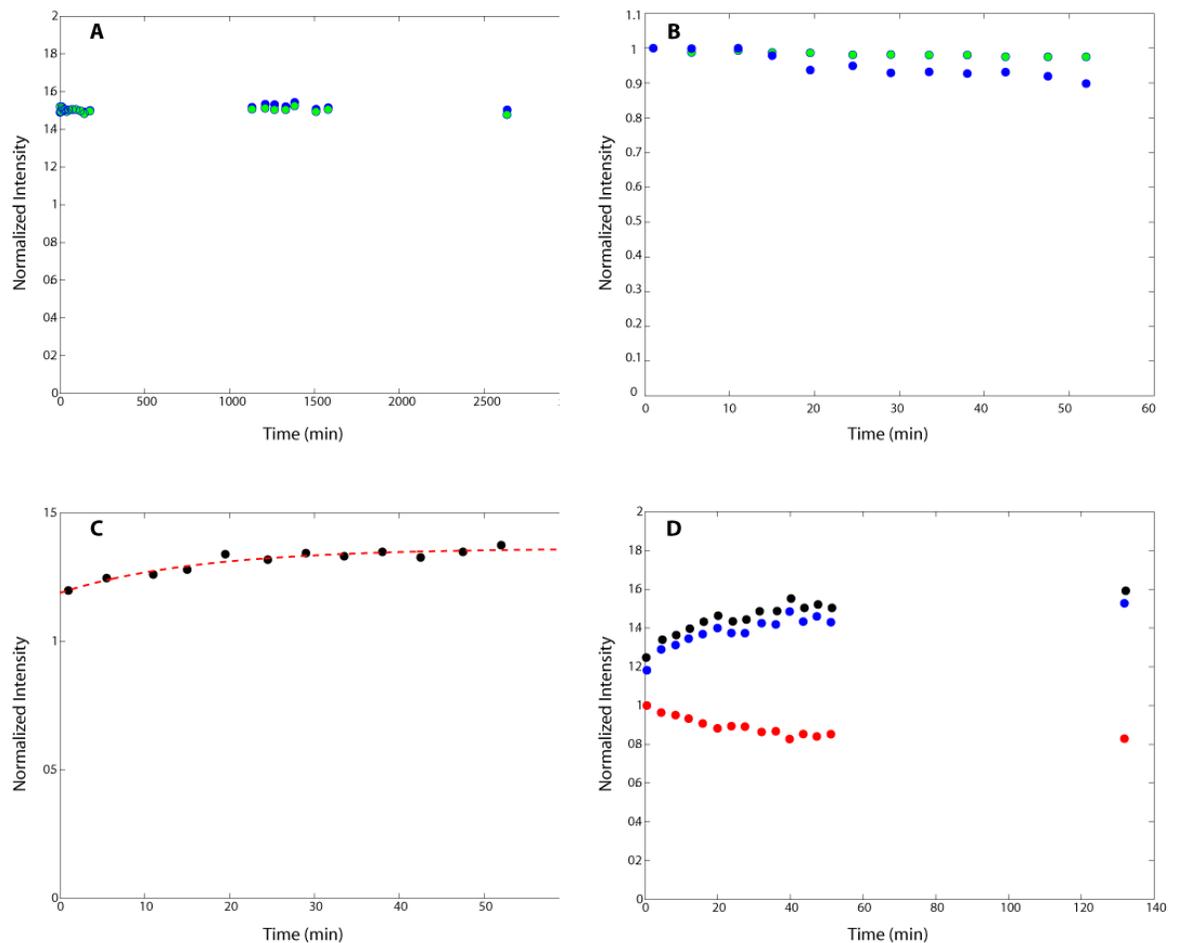
(Fig. 6.14). Incubating RhB-MBP1 with C-Mad2 at room temperature overnight resulted in 60% binding. This corresponds to a  $K_D$  of 3  $\mu\text{M}$ , which is much higher than previously published (190 nM for MBP1—Mad2<sup>RQ</sup>; (13). These results suggest that Mad2 binding has not reached equilibrium and is therefore incredibly slow, perhaps because the C-Mad2 conformation must partially unfold prior to binding (41).



**Figure 6.14** FCS results for RhB-MBP1 alone, mixed with C-Mad2 for 20 min, mixed with C-Mad2 overnight, and mixed with BSA. The data are fit to the auto-correlation function (Eq. 6.1).

One observation from the FCS experiments was that the fluorescence intensity increased when RhB-MBP1 was bound to Mad2, suggesting that binding could be monitored by simply measuring fluorescence intensity. A fluorescence intensity time course at 25°C revealed that RhB-MBP1 fluorescence intensity was essentially unaffected by doubling the C-Mad2<sup>RQ</sup> concentration, indicating that RhB-MBP1 has been

saturated. What was more surprising was that fluorescence intensity increased by a factor of 1.5 prior to the first time point and then remained constant (Fig. 6.15A). This result is consistent with binding reaching a steady state in under a minute, in sharp contrast to FCS results. One possible explanation is that fluorescence intensity measurements require a higher concentration of RhB-MBP1 (400 nM) than FCS (50 nM), which should promote binding.



**Figure 6.15** RhB-MBP1 binding monitored by fluorescence intensity monitored as a function of time. (A) 400 nM RhB-MBP1 mixed with 4  $\mu\text{M}$  (green) and 8  $\mu\text{M}$  (blue) C-Mad2<sup>RQ</sup> at 25°C. Binding appears to reach equilibrium by the first time point. (B) Signal drift for NATA (green) and RhB-MBP1 (blue) at 7.5 °C. (C) C-Mad2<sup>RQ</sup> mixed with RhB-MBP1 at 7.5 °C. After normalization the data fits with a rate of 0.61  $\text{min}^{-1}$  with a burst to 1.18. (D) RhB-MBP1 mixed with C-Mad2<sup>RQ</sup> (black) and O-Mad2<sup>RQ</sup> (blue) at 6.9 °C. The apparent time-dependent binding actually arises due to normalization to RhB-MBP1 signal (red).

Signals have been normalized to an RhB-MBP1 control time course except for panel B and for RhB-MBP1 in panel D for which signals have been normalized to the initial time point.

Binding kinetics needed to be slowed in order to move into the experimentally observable range. This could not be accomplished by reducing RhB-MBP1 concentrations because using a monochromator requires much higher fluorescence intensities than a broadband filter. Structural data suggests that MBP1 binding requires a rearrangement of secondary structure by Mad2 (10, 25). Since this rearrangement disrupts secondary structure it is reasonable to assume that there is a significant enthalpic barrier between the bound and unbound states. This should correspond to sharp temperature dependence. We therefore repeated RhB-MBP1 binding experiments at lower temperatures.

The first low temperature fluorescence intensity experiments were done to check drift in fluorescence intensity of RhB-MBP1 and the control fluorophore NATA (Fig. 6.15B). Fluorescence intensity decreased for both fluorophores, but much more so for RhB-MBP1. RhB-MBP1 loses 10% of its fluorescence intensity in less than 1 hr.

Reducing the temperature to 7.5 °C resulted in a measurable binding rate of 0.6 min<sup>-1</sup> (Fig. 6.15C). In order to verify the specificity of RhB-MBP1 binding, this experiment was repeated using O-Mad2<sup>RQ</sup> as a negative control. Surprisingly, O-Mad2<sup>RQ</sup> displayed a binding rate identical to that of C-Mad2<sup>RQ</sup> (Fig. 6.15D). Analysis of the control peptide revealed that the apparent rate arises from normalizing to RhB-MBP1, which loses signal presumably due to photobleaching or aggregation. The initial burst signal likely does result from binding, but we can conclude that this interaction is non-specific since binding is similar for O-Mad2<sup>RQ</sup> and C-Mad2<sup>RQ</sup>.

The contrasting observations from FCS and fluorescence intensity experiments are puzzling. One possible explanation is that there are two different binding reactions taking place: one that is kinetically slow, but highly stable and a second that is kinetically fast, and marginally stable. FCS results are consistent with a slow, highly stable binding event like the one that would be expected for MBP1 incorporation into the Mad2 structure. Fluorescence intensity results are more consistent with a fast, marginally stable binding event such as one in which the fluorophore interacts with a hydrophobic pocket. Since the two experiments were done with RhB-MBP1 concentrations that differed by nearly an order of magnitude, it is conceivable that FCS experiments were done at concentrations below the  $K_D$  of the non-specific interaction.

Lad et al had success detecting Mad2 binding to a Cdc20 fragment by attaching a Texas Red fluorophore to the fragment and measuring fluorescence polarization (33). Based on isothermal calorimetry (ITC) experiments they were able to determine that the fluorophore did not significantly alter the  $K_D$  (160 nM). They measured binding constants for the Cdc20<sup>111-138</sup> fragment to O-Mad2<sup>RQ</sup> and C-Mad2<sup>RQ</sup> at room temperature ( $5.3 \times 10^{-5} \mu\text{M}^{-1}\text{s}^{-1}$  and  $3.7 \times 10^{-2} \mu\text{M}^{-1}\text{s}^{-1}$  respectively). If we assume the same kinetics for RhB-MBP1 then this corresponds to a binding rate of  $0.15 \text{ s}^{-1}$  for the Mad2 concentration we used for fluorescence intensity measurements ( $4 \mu\text{M}$ ). In agreement with our measurements, this rate would be too fast to measure at room temperature by our technique. However, the non-specific binding we observe at lower temperatures makes this point moot. Future efforts to study Mad2 binding will need to take into consideration both the potential non-specific nature of certain fluorophores and the Mad2 concentrations used given the rates reported by Lad et al (33).

#### 4. Conclusions

Ever since Anfinsen demonstrated that purified proteins could refold into their native structure a central dogma of biochemistry has been that proteins have a single native state that is determined by primary sequence (43). The mitotic checkpoint protein Mad2 challenges this view by forming two “native” conformations with similar free energy (11). Interestingly, the differences in these conformations and the slow interconversion between them are integral to its function (4, 41). Despite the importance of Mad2 interconversion, little is known about the transition from the inactive O-Mad2 conformation to the inhibitory C-Mad2 conformation. The investigation described here begins to illuminate this important subject.

Mad2 is capable of forming dimers which catalyze O-Mad2 to C-Mad2 conversion under certain conditions (34). Mutations that abolish dimerization reduce the complexity of interconversion and are therefore ideal for studying this reaction. We find these mutations do not completely abolish dimerization (Fig. 6.5B), but instead limit it to an extremely slow binding reaction involving C-Mad2. While it is not clear what regions of Mad2 are involved in this binding we have demonstrated that this interaction does not involve the N-terminus being inserted into the Mad2 binding pocket.

The O-Mad2 to C-Mad2 conversion was studied at the whole molecule and single residue level, which yielded drastically different rates. This may be explained by the presence of an intermediate that can be observed by  $^1\text{H}$ - $^{15}\text{N}$  NMR (Figs. 6.9, 6.10) but not by anion exchange (Fig. 6.7B). Though NMR peaks could not be assigned by standard 3D experiments, due to low concentrations and conformational heterogeneity (Fig. 6.13),

several peaks could be assigned by comparison with published spectra and peak tables (Fig. 6.9). We were thereby able to determine that many of the peaks shifting from their O-Mad2 to C-Mad2 positions were found in  $\beta 2$  and  $\beta 3$  and the adjoining helix  $\alpha A$  despite these regions appearing very similar in both conformations (Fig. 6.10). HX results are consistent with this observation (Fig. 6.12). These results suggest that these strands are important for interconversion, apparently adopting a C-Mad2 arrangement in an intermediate that appears to be on the O-Mad2 side of the transition state. Also central to conformational change is residue Phe141. Mutating Phe141 to Ala alters the O-Mad2/C-Mad2 equilibrium, but has an even greater effect on the rate of conversion. This suggests that the rearrangement of residue 141 is part of the rate limiting step in the O-Mad2 to C-Mad2 transition.

We have previously proposed that catalysis of the O-Mad2 to C-Mad2 transition by dimerization occurs by stabilization of an intermediate Mad2 conformer (I-Mad2) found on the O-Mad2 side of the rate limiting transition state (Fig. 6.3C; (41)). Observations can easily be interpreted in terms of this model. The shift in NMR peaks associated with  $\beta 2$  and  $\beta 3$  likely corresponds to a transition from O-Mad2 to I-Mad2. Only later do the C-terminal strands and Phe141 move into position to form C-Mad2. Fully consistent with this model is the X-ray structure of the O-Mad2/C-Mad2 which shows that O-Mad2  $\beta 3$  interacts with catalyzing C-Mad2 (28). Notably, the observation by Lad et al (33) that C-Mad2 can bind to ligand more rapidly than O-Mad2 can suggests the presence of a second intermediate on the C-Mad2 side of the transition state which allows for the C-Mad2 structure to rearrange into a binding competent state.

## 5. Future Directions

Plenty of opportunity remains for investigating the I-Mad2 conformation and other questions related to Mad2 conformational interconversion. As mentioned in section 3.3.2, HXMS pulse labeling experiments look to be a promising avenue for determining the rate of conformational change for interesting structural regions. Of all potential future work stemming from this study, this method appears to have the highest likelihood of yielding interesting results. This method could be applied to studying Mad2 catalysis as well. Vital to studying catalysis is being able to distinguish the Mad2 population undergoing conversion from the population acting as catalyst. This can be done by growing one population under perdeuterating conditions so that this species will be easily distinguishable by mass spectroscopy. These non-exchanging deuteriums would greatly increase the mass of peptide fragments from the catalyst population so that they would not overlap with peptides from the population undergoing interconversion. We have reported differences in the macroscopic conversion rates for Mad2<sup>R133A</sup>, Mad2<sup>RQ</sup> and Mad2<sup>RF</sup> mutants, but have not reported residue specific rates. HX pulse labeling experiments may be applied to these mutants to see if they vary in how rapidly they attain the I-Mad2 conformation.

More ambitious experiments may involve mutations affecting the  $\beta$ 2 and  $\beta$ 3 region. Both strands can be removed or more subtle mutations can be introduced that will disrupt the aromatic interactions between linking these strands to  $\alpha$ A (Tyr 33 to Phe43 or Tyr49 to Phe23 and Phe26). Combining these mutations with whole molecule conversion experiments, such as those monitored by 1D NMR or anion exchange, would allow for the role of this region in Open to Closed conversion to be investigated without the need

for HX experiments. However, such experiments run the risk of creating aggregation prone mutants, particularly if these strands were simply removed and the underlying hydrophobic core were exposed to solvent. Another possibility is to investigate the influence of pH on conversion rates and to determine if His191 plays a role in this behavior. Finding a pH dependence within the physiological range may suggest a yet to be described mechanism for spindle checkpoint control.

If NMR experiments are to be pursued then it may be advisable to truncate the first ten residues of Mad2. Published peak assignment tables come from a truncated version of Mad2 (10, 25), so this truncation may be required for adequately assigning peaks. However, truncation of the N-terminus likely affects kinetics and perhaps even the presence or absence of I-Mad2.

Investigations of the Mad2 ligand binding reaction will clearly need to utilize a different fluorophore than the one described here. Recently, other workers have had success monitoring Mad2 binding by measuring fluorescence polarization of Texas Red attached to a ligand (33). This technique combined with HXMS can be used to identify the conformational change that allows C-Mad2 to bind ligand. Based on the structure of Mad2 we can assume that this binding reaction minimally requires the rearrangement of the C-terminal strands  $\beta 8'$  and  $\beta 8''$  (Figs. 6.1D, 6.2A), which should limit the binding rate to the rate of rearrangement. By increasing ligand concentration one should reach a concentration independent rate that corresponds to the opening rate (identical to the HX EX1 case; see Chapter 1). Similarly, applying HXMS to C-Mad2 at high pH should reveal segments of the protein that exchange with a pH-independent rate identical to the rate identified by ligand binding, identifying these regions as being involved in the

rearrangement that is necessary for ligand binding. A potential pitfall to this method would be that these rearrangements may be so rapid that determining the rearrangement rate may require ligand concentrations higher than are suitable for experiments. In this case it may be necessary to reduce the binding rate. The easiest way to do so will be to reduce the experiment temperature, which should greatly reduce the binding rate due to the enthalpic barrier associated with breaking the hydrogen bonds that keep  $\beta'$  and  $\beta''$  in place. Alternatively, one can create mutants that stabilize  $\beta'$  and  $\beta''$ , though this may require extensive trial and error.

## 6. Bibliography

1. Cleveland, D.W., Y. Mao, and K.F. Sullivan. 2003. Centromeres and Kinetochores: From Epigenetics to Mitotic Checkpoint Signaling. *Cell*. 112: 407-421.
2. Walczak, C.E., and R. Heald. 2008. Mechanisms of Mitotic Spindle Assembly and Function. In: *A Survey of Cell Biology*. Academic Press. pp. 111-158.
3. Black, B.E., and E.A. Bassett. 2008. The histone variant CENP-A and centromere specification. *Current Opinion in Cell Biology*. 20: 91-100.
4. Musacchio, A., and E.D. Salmon. 2007. The spindle-assembly checkpoint in space and time. *Nat Rev Mol Cell Biol*. 8: 379-393.
5. Yu, H. 2007. Cdc20: A WD40 Activator for a Cell Cycle Degradation Machine. *Molecular Cell*. 27: 3-16.

6. Kops, G.J.P.L., D.R. Foltz, and D.W. Cleveland. 2004. Lethality to human cancer cells through massive chromosome loss by inhibition of the mitotic checkpoint. *Proceedings of the National Academy of Sciences of the United States of America*. 101: 8699 -8704.
7. Michel, L., E. Diaz-Rodriguez, G. Narayan, E. Hernando, V.V.V.S. Murty, and R. Benezra. 2004. Complete loss of the tumor suppressor MAD2 causes premature cyclin B degradation and mitotic failure in human somatic cells. *Proceedings of the National Academy of Sciences of the United States of America*. 101: 4459 -4464.
8. Gascoigne, K.E., and S.S. Taylor. 2008. Cancer Cells Display Profound Intra- and Interline Variation following Prolonged Exposure to Antimitotic Drugs. *Cancer Cell*. 14: 111-122.
9. van Mourik, T.R.M. 2011. Inhibiting the inhibitor: how the spindle assembly checkpoint gets silenced. .
10. Luo, X., G. Fang, M. Coldiron, Y. Lin, H. Yu, M.W. Kirschner, and G. Wagner. 2000. Structure of the Mad2 spindle assembly checkpoint protein and its interaction with Cdc20. *Nature Structural Biology*. 7: 224.
11. Luo, X., Z. Tang, G. Xia, K. Wassmann, T. Matsumoto, J. Rizo, and H. Yu. 2004. The Mad2 spindle checkpoint protein has two distinct natively folded states. *Nat Struct Mol Biol*. 11: 338-345.

12. Yu, H. 2006. Structural activation of Mad2 in the mitotic spindle checkpoint: the two-state Mad2 model versus the Mad2 template model. *The Journal of Cell Biology*. 173: 153 -157.
13. De Antoni, A., C.G. Pearson, D. Cimini, J.C. Canman, V. Sala, L. Nezi, M. Mapelli, L. Sironi, M. Faretta, E.D. Salmon, and A. Musacchio. 2005. The Mad1/Mad2 Complex as a Template for Mad2 Activation in the Spindle Assembly Checkpoint. *Current Biology*. 15: 214-225.
14. Sironi, L., M. Mapelli, S. Knapp, A.D. Antoni, K.-T. Jeang, and A. Musacchio. 2002. Crystal structure of the tetrameric Mad1-Mad2 core complex: implications of a “safety belt” binding mechanism for the spindle checkpoint. *EMBO J*. 21: 2496-2506.
15. Vink, M., M. Simonetta, P. Transidico, K. Ferrari, M. Mapelli, A. De Antoni, L. Massimiliano, A. Ciliberto, M. Faretta, E.D. Salmon, and A. Musacchio. 2006. In Vitro FRAP Identifies the Minimal Requirements for Mad2 Kinetochores Dynamics. *Current Biology*. 16: 755-766.
16. Chen, R.-H., A. Shevchenko, M. Mann, and A.W. Murray. 1998. Spindle Checkpoint Protein Xmad1 Recruits Xmad2 to Unattached Kinetochores. *The Journal of Cell Biology*. 143: 283 -295.
17. Chen, R.-H., D.M. Brady, D. Smith, A.W. Murray, and K.G. Hardwick. 1999. The Spindle Checkpoint of Budding Yeast Depends on a Tight Complex between the Mad1 and Mad2 Proteins. *Molecular Biology of the Cell*. 10: 2607 -2618.

18. Waters, J.C., R.-H. Chen, A.W. Murray, and E.D. Salmon. 1998. Localization of Mad2 to Kinetochores Depends on Microtubule Attachment, Not Tension. *The Journal of Cell Biology*. 141: 1181 -1191.
19. Yang, M., B. Li, C.-J. Liu, D.R. Tomchick, M. Machius, J. Rizo, H. Yu, and X. Luo. 2008. Insights into Mad2 Regulation in the Spindle Checkpoint Revealed by the Crystal Structure of the Symmetric Mad2 Dimer. *PLoS Biol*. 6: e50.
20. Fava, L.L., M. Kaulich, E.A. Nigg, and A. Santamaria. 2011. Probing the in vivo function of Mad1:C-Mad2 in the spindle assembly checkpoint. *EMBO J*. 30: 3322-3336.
21. Hardwick, K.G., R.C. Johnston, D.L. Smith, and A.W. Murray. 2000. MAD3 Encodes a Novel Component of the Spindle Checkpoint Which Interacts with Bub3p, Cdc20p, and Mad2p. *The Journal of Cell Biology*. 148: 871 -882.
22. Sudakin, V., G.K.T. Chan, and T.J. Yen. 2001. Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *The Journal of Cell Biology*. 154: 925 -936.
23. Murzin, A.G. 2008. Biochemistry. Metamorphic proteins. *Science*. 320: 1725-1726.
24. Kim, S., H. Sun, H.L. Ball, K. Wassmann, X. Luo, and H. Yu. 2010. Phosphorylation of the spindle checkpoint protein Mad2 regulates its conformational transition. *Proceedings of the National Academy of Sciences*. 107: 19772 -19777.

25. Luo, X., Z. Tang, J. Rizo, and H. Yu. 2002. The Mad2 Spindle Checkpoint Protein Undergoes Similar Major Conformational Changes Upon Binding to Either Mad1 or Cdc20. *Molecular Cell*. 9: 59-71.
26. Shah, J.V., E. Botvinick, Z. Bonday, F. Furnari, M. Berns, and D.W. Cleveland. 2004. Dynamics of Centromere and Kinetochores Proteins: Implications for Checkpoint Signaling and Silencing. *Current Biology*. 14: 942-952.
27. Orth, M., B. Mayer, K. Rehm, U. Rothweiler, D. Heidmann, T.A. Holak, and O. Stemmann. 2011. Shugoshin is a Mad1/Cdc20-like interactor of Mad2. *EMBO J*. 30: 2868-2880.
28. Mapelli, M., L. Massimiliano, S. Santaguida, and A. Musacchio. 2007. The Mad2 Conformational Dimer: Structure and Implications for the Spindle Assembly Checkpoint. *Cell*. 131: 730-743.
29. Maity, H., J.N. Rumbley, and S.W. Englander. 2006. Functional role of a protein foldon—An  $\Omega$ - loop foldon controls the alkaline transition in ferricytochrome c. *Proteins: Structure, Function, and Bioinformatics*. 63: 349-355.
30. Hoang, L., H. Maity, M.M.G. Krishna, Y. Lin, and S.W. Englander. 2003. Folding Units Govern the Cytochrome c Alkaline Transition. *Journal of Molecular Biology*. 331: 37-43.
31. Englander, S.W., L. Mayne, and M.M.G. Krishna. 2007. Protein Folding and Misfolding: Mechanism and Principles. *Quarterly Reviews of Biophysics*. 40: 287-326.

32. Hilser, V.J., and E.B. Thompson. 2007. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proceedings of the National Academy of Sciences*. 104: 8311 -8315.
33. Lad, L., S. Lichtsteiner, J.J. Hartman, K.W. Wood, and R. Sakowicz. 2009. Kinetic Analysis of Mad2–Cdc20 Formation: Conformational Changes in Mad2 Are Catalyzed by a C-Mad2–Ligand Complex. *Biochemistry*. 48: 9503-9515.
34. Mapelli, M., F.V. Filipp, G. Rancati, L. Massimiliano, L. Nezi, G. Stier, R.S. Hagan, S. Confalonieri, S. Piatti, M. Sattler, and A. Musacchio. 2006. Determinants of conformational dimerization of Mad2 and its inhibition by p31comet. *EMBO J*. 25: 1273-1284.
35. Sudakin, V., and T.J. Yen. 2004. Purification of the Mitotic Checkpoint Complex, an Inhibitor of the APC/C From HeLa Cells. In: *Checkpoint Controls and Cancer*. New Jersey: Humana Press. pp. 199-212.
36. Bodansky, M. 1993. *Peptide chemistry: a practical approach*. 2nd ed. New York: Springer-Verlag.
37. Delaglio, F., S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*. 6.
38. MacKerell, A., B. Brooks, C.L. Brooks, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998. CHARMM: the energy function and its parameterization with an

- overview of the program. In: Schleyer P, editor. The encyclopedia of computational chemistry. Chichester: Wiley. pp. 271-277.
39. Gilson, M.K., K.A. Sharp, and B.H. Honig. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry*. 9: 327-335.
  40. Sharp, K.A., R.A. Friedman, V. Misra, J. Hecht, and B. Honig. 1995. Salt effects on polyelectrolyte–ligand binding: Comparison of Poisson–Boltzmann, and limiting law/counterion binding models. *Biopolymers*. 36: 245-262.
  41. Skinner, J.J., S. Wood, J. Shorter, S.W. Englander, and B.E. Black. 2008. The Mad2 partial unfolding model: regulating mitosis through Mad2 conformational switching. *The Journal of Cell Biology*. 183: 761 -768.
  42. Chamberlain, A.K., and S. Marqusee. 2000. Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. In: *Protein folding mechanisms*. Academic Press. pp. 283-328.
  43. Anfinsen, C.B., E. Haber, M. Sela, and F.H. White. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*. 47: 1309-1314.